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

In Vitro Mutagenesis Protocols

Third Edition

Edited by

Jeff Braman

Stratagene, An Agilent Technologies Division, La Jolla, CA, USA

Humana Press

Editor
Jeff Braman, Ph.D.
Stratagene
An Agilent Technologies Division
La Jolla, CA
USA
jeff.braman@agilent.com

ISSN 1064-3745 e-ISSN 1940-6029
ISBN 978-1-60761-651-1 e-ISBN 978-1-60761-652-8
DOI 10.1007/978-1-60761-652-8
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010925108

Cover illustration: Based on Figure 2A of Chapter 16

Printed on acid-free paper

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

Preface

In the preface to the second edition of this volume, I claimed that the book represented a toolbox containing protocols to advance the understanding of the connection between nucleotide sequence and sequence function. The same holds true in this third edition, with a notable exception; the third edition contains a variety of specialty tools successfully employed by scientists just like you to unravel the intricacies of protein–protein interaction, protein structure–function, protein regulation of biological processes, and protein activity. A novel section is included containing mutagenesis methods for unique microbes as a guide to the generalization of mutagenesis strategies for a host of microbial systems.

Each chapter was expanded from the "Methods" section of a paper published in a reputable peer-reviewed journal for the purpose of solving one or more of the problems described above. Chapter "Notes" are included to highlight critical experimental details. Many of the authors describe the utility of their protocol to answer a difficult experimental question. All the authors, including myself, desire that you be successful in your research efforts.

I want to thank my parents who always taught me the value of hard work.

La Jolla, CA *Jeff Braman, Ph.D.*

Contents

Contributors

Sophie H. Adjalley • *Department of Microbiology, College of Physicians and Surgeons, Columbia University, New York, NY, USA*
Miguel Alcalde • *Instituto de Catálisis y Petroleoquímica, CSIC, Cantoblanco, Madrid, Spain*
Mehdi Arbabi-Ghahroudi • *Institute for Biological Sciences, National Research Council of Canada, Ottawa, ON, Canada*
Khalil Bouayadi • *MilleGen SA, Labège, France*
Samya Chakravorty • *Department of Biology, University of Vermont, Burlington, VT, USA*
John C. Chaput • *The Biodesign Institute, and Department of Chemistry and Biochemistry, Center for BioOptical Nanotechnology, Arizona State University, Tempe, AZ, USA*
Zong-Ming Cheng • *Department of Plant Sciences, University of Tennessee, Knoxville, TN, USA*
Vincent Dammai • *Hollings Cancer Center and Pathology and Laboratory Medicine, Medical University of South Carolina, Charleston, SC, USA*
Feng Ding • *Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*
Nikolay V. Dokholyan • *Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*
Qihan Dong • *Department of Medicine, University of Sydney, Sydney, NSW, Australia*
Thorsten Eggert • *Evocatal GmbH, Düsseldorf, Germany*
Stéphane Emond • *MilleGen SA, Labège, France*
Oran Erster • *Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel*
Thomas A. Ficht • *Department of Veterinary Pathobiology, College of Veterinary Medicine, Texas A&M University, College Station, TX, USA*
David A. Fidock • *Department of Microbiology, College of Physicians and Surgeons, Columbia University, New York, NY, USA*
Josef D. Franke • *Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, USA*
Duncan J.H. Gaskin • *Institute of Food Research, Norwich, UK*
Dimitris K. Grammatopoulos • *Warwick Medical School, University of Warwick, Coventry, UK*
David Grand • *MilleGen SA, Labège, France*

Soon-Duck Ha • *Infectious Diseases Research Group, Department of Microbiology and Immunology, Siebens-Drake Research Institute, University of Western Ontario, London, ON, Canada*
Christian Heinis • *Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*
Ichiro Hirao • *Systems and Structural Biology Center, RIKEN, and TagCyx Biotechnologies, Yokohama, Kanagawa, Japan*
Holly H. Hogrefe • *Stratagene Products Division, Agilent Technologies, Inc., La Jolla, CA, USA*
Ying Huang • *Department of Microbiology and Molecular Genetics, University of Medicine and Dentistry of New Jersey – New Jersey Medical School, Newark, NJ, USA*
Karl-Erich Jaeger • *Institute of Molecular Enzyme Technology, Research Centre Jülich, Heinrich Heine University Düsseldorf, Jülich, Germany*
Kai Johnsson • *Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*
Melissa Kahl-McDonagh • *Department of Veterinary Pathobiology, College of Veterinary Medicine, Texas A&M University, College Station, TX, USA*
Hakim Kharrat • *MilleGen SA, Labège, France*
Seahee Kim • *Department of Biochemistry, Yonsei University, Seoul, Korea*
Sung Ouk Kim • *Infectious Diseases Research Group, Department of Microbiology and Immunology, Siebens-Drake Research Institute, University of Western Ontario, London, ON, Canada*
Michiko Kimoto • *Systems and Structural Biology Center, RIKEN, and TagCyx Biotechnologies, Yokohama, Kanagawa, Japan*
Ulrich Krauss • *Institute of Molecular Enzyme Technology, Research Centre Jülich, Heinrich Heine University Düsseldorf, Jülich, Germany*
Jehan Lee • *Department of Biochemistry, Yonsei University, Seoul, Korea*
Marcus C.S. Lee • *Department of Microbiology, College of Physicians and Surgeons, Columbia University, New York, NY, USA*
Moti Liscovitch • *Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel*
Roger MacKenzie • *Institute for Biological Sciences, National Research Council of Canada, Ottawa, ON, Canada; Department of Environmental Biology, Ontario Agricultural College, University of Guelph, Guelph, ON, Canada*
Danijela Markovic • *Department of Cell Physiology and Pharmacology, University of Leicester, Leicester, UK*
Tomoaki Matsuura • *Department of Bioinformatics Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka, Japan*
Elizabeth O. McCullum • *The Biodesign Institute, and Department of Chemistry and Biochemistry, Center for BioOptical Nanotechnology, Arizona State University, Tempe, AZ, USA*
Philippe Mondon • *Director of Antibody Engineering and Molecular Evolution Department, MilleGen SA, Labège, France*

Ghazala Muteeb • *Laboratory of Transcription Biology, Center for DNA Fingerprinting and Diagnostics, Hyderabad, India*
Venugopal Nair • *Viral Oncogenesis Group, Division of Microbiology, Institute for Animal Health, Compton, Berkshire, UK*
Hiroshi Nikaido • *Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA*
Nikolaus Osterrieder • *Institut für Virologie, Freie Universität Berlin, Berlin, Germany; Department of Microbiology and Immunology, Cornell University, Ithaca, NY, USA*
Jianwu Pei • *Department of Veterinary Pathobiology, College of Veterinary Medicine, Texas A&M University, College Station, TX, USA*
Ri-He Peng • *Biotechnology Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai, China*
Dong-Kyun Ryu • *Department of Biochemistry, Yonsei University, Seoul, Korea*
Wang-Shick Ryu • *Department of Biochemistry, Yonsei University, Seoul, Korea*
Ranjan Sen • *Laboratory of Transcription Biology, Center for DNA Fingerprinting and Diagnostics, Hyderabad, India*
Myeong-Kyun Shin • *Department of Biochemistry, Yonsei University, Seoul, Korea*
Gregory A. Smith • *Department of Microbiology-Immunology, Northwestern University, Chicago, IL, USA*
Robert A. Smith • *Department of Pathology, University of Washington, Seattle, WA, USA*
Nathalie Souyris • *MilleGen SA, Labège, France*
Julien Sylvestre • *PhotoFuel SAS, Paris, France*
Yumiko Takatsuka • *Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA*
Jamshid Tanha • *Institute for Biological Sciences, National Research Council of Canada, Ottawa, ON, Canada; Department of Environmental Biology, Ontario Agricultural College, University of Guelph, Guelph, ON, Canada; Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON, Canada*
B. Karsten Tischer • *Institut für Virologie, Freie Universität Berlin, Berlin, Germany*
Jim O. Vigoreaux • *Department of Biology, University of Vermont, Burlington, VT, USA*
Arnoud H.M. van Vliet • *Institute of Food Research, Norwich, UK*
Berea A.R. Williams • *Center for BioOptical Nanotechnology, The Biodesign Institute, and Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ, USA*
Ai-Sheng Xiong • *Biotechnology Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai, China*
Hayato Yanagida • *Graduate School of Frontier Biosciences, Osaka University, Osaka, Japan*
Quan-Hong Yao • *Biotechnology Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai, China*

Shuangye Yin • *Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*
Tetsuya Yomo • *Department of Bioinformatics Engineering, Graduate School of Information Science and Technology, Graduate School of Frontier Biosciences, Osaka University, Osaka, Japan; Exploratory Research for Advanced Technology (ERATO), Japan Science and Technology Agency (JST), Osaka, Japan*
Lei Young • *Department of Synthetic Biology, J. Craig Venter Institute, Rockville, MD, USA*
Jinglei Zhang • *Center for BioOptical Nanotechnology, The Biodesign Institute, and Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ, USA*
Vivian Q. Zhang • *Stratagene Products Division, Agilent Technologies, Inc., La Jolla, CA, USA*
Zhen Zhang • *Department of Microbiology and Molecular Genetics, University of Medicine and Dentistry of New Jersey – New Jersey Medical School, Newark, NJ, USA*
Yuguang Zhao • *Viral Oncogenesis Group, Division of Microbiology, Institute for Animal Health, Compton, Berkshire, UK*
Hua Zhu • *Department of Microbiology and Molecular Genetics, University of Medicine and Dentistry of New Jersey – New Jersey Medical School, Newark, NJ, USA*

Part I

Mutagenesis in Various Microbial Backgrounds

Chapter 1

Mutagenesis Protocols in *Saccharomyces cerevisiae* by In Vivo Overlap Extension

Miguel Alcalde

Abstract

A high recombination frequency and its ease of manipulation has made *Saccharomyces cerevisiae* a unique model eukaryotic organism to study homologous recombination. Indeed, the well-developed recombination machinery in *S. cerevisiae* facilitates the construction of mutant libraries for directed evolution experiments. In this context, in vivo overlap extension (IVOE) is a particularly attractive protocol that takes advantage of the eukaryotic apparatus to carry out combinatorial saturation mutagenesis, site-directed recombination or site-directed mutagenesis, avoiding ligation steps and additional PCR reactions that are common to standard in vitro protocols.

Key words: IVOE, *Saccharomyces cerevisiae*, Combinatorial saturation mutagenesis, In vivo recombination, Directed evolution

1. Introduction

Directed Molecular Evolution is a powerful protein engineering tool to improve the known features of enzymes or to generate novel activities that are not required in natural environments (1, 2). Through this methodology, the scientist recreates the key events of natural evolution in a laboratory environment (mutation, DNA-recombination and selection), thereby making it possible to design interesting and technologically useful enzymes. In the framework of protein engineering, saturation mutagenesis has long been used to carry out semirational studies (3) since this approach involves the mutation of any single amino acid codon to all the other codons that will generate the 20 naturally occurring amino acids. This technique is commonly employed to improve the characteristics of enzymes at "*hot-spot*" residues already identified by conventional random mutagenesis (see Note 1).

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_1, © Springer Science+Business Media, LLC 2010

In addition, it can be employed to simultaneously mutate several codons (combinatorial saturation mutagenesis, CSM), which will enable all possible combinations of interesting residues to be evaluated in order to identify their optimal interactions and synergies (4). CSM is typically carried out by laborious in vitro protocols that are based on several consecutive PCR reactions and an additional ligation step with the vector in order to clone the whole mutagenized fragment (see Note 2, Fig. 1) (5, 6).

The exchange of genetic material by recombination occurs in all living organisms and it is the main process that generates diversity in the evolution of species. The eukaryotic machinery of *Saccharomyces cerevisiae* offers an array of possibilities to construct mutant libraries or to recombine ("shuffle") DNA fragments. Unlike other heterologous hosts used for directed evolution, the high frequency of homologous recombination in

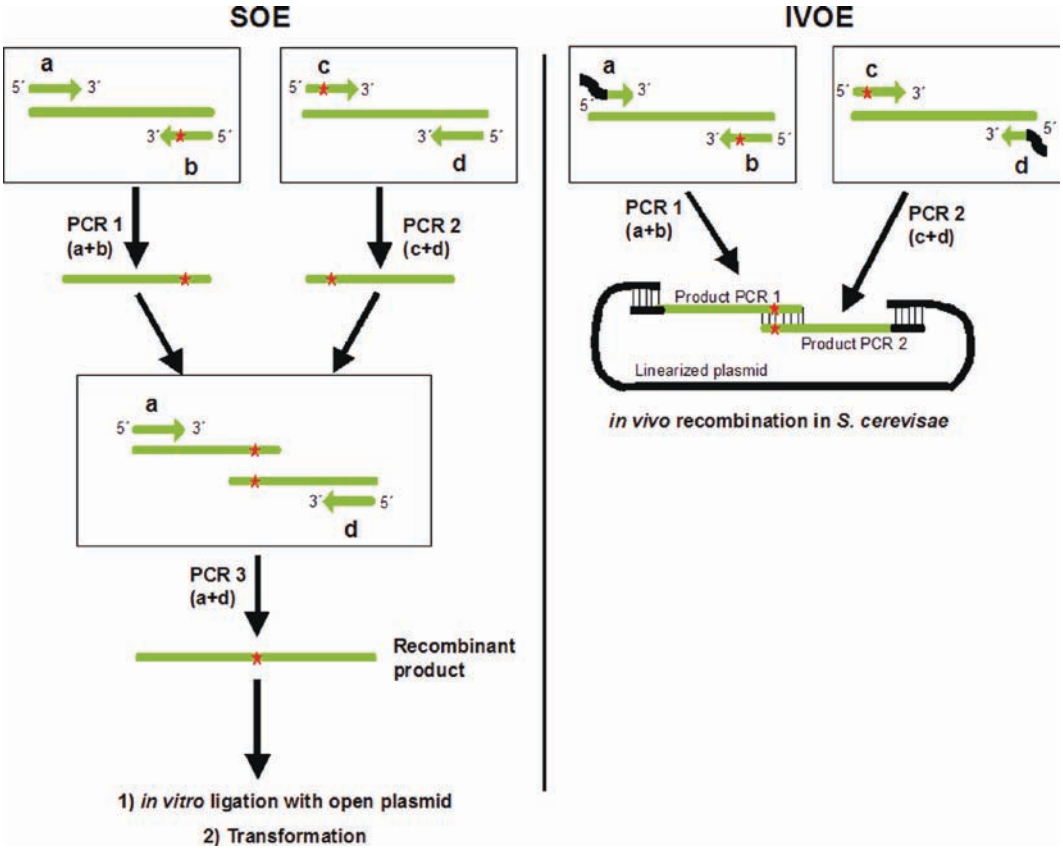


Fig. 1. Sequence splicing by SOE is a mutagenic method that recombines DNA sequences containing mutations through several consecutive PCR reactions. This method requires an additional in vitro ligation step in order to clone the whole fragment within the vector. As an alternative, IVOE eliminates one PCR step and the ligation in vitro with the linearized plasmid. Accordingly, it takes advantage of the eukaryotic apparatus of *S. cerevisiae* and thus, it is necessary to design mutagenized primers with suitable overhangs.

S. cerevisiae favors its use to clone eukaryotic proteins and in new in vivo protocols aimed at generating diversity (7, 8) (see Note 3).

In this chapter, we describe how to perform CSM by in vivo overlap extension (IVOE), taking advantage of the eukaryotic machinery of *S. cerevisiae* (6). In this protocol mutant libraries are recombined, spliced and at the same time repaired in a circular autonomously replicating vector, a process that offers many attractive advantages when compared to classical in vitro protocols (see Note 4). Furthermore, we describe how IVOE can be employed for site-directed recombination and site-directed mutagenesis. These methods have been used to generate mutant genes during the artificial evolution of a fungal laccase in *S. cerevisiae* in order to enhance its activity and stability in organic cosolvents. The evolved laccase variant engineered in this project was capable of resisting a wide array of cosolvents at high concentrations (9, 10). More recently, the protocol has been applied to carry out semirational studies in order to understand the role of different regions in the laccase structure that are involved in the traffic of oxygen towards its trinuclear copper cluster (4, 6).

2. Materials

2.1. Reagents

All chemicals used were of reagent grade purity.

2.1.1. Chemicals

1. dNTPs (Sigma, Spain).
2. Appropriate PCR primers.
3. Low melting point agarose (Bio-rad, Spain).
4. DNA extraction from agarose gels: Zymoclean gel DNA recovery kit (Zymo Research, USA).
5. Yeast transformation: yeast transformation kit (Sigma, Spain).
6. Zymoprep kit (Zymo Research, USA).

2.1.2. Biological Materials

1. *E. coli* XL2-blue competent cells (Stratagene, USA).
2. *S. cerevisiae* (e.g., protease deficient strain BJ 5465, ATCC 208289).
3. Expression shuttle vector containing the gene of interest under the appropriate promoter, a signal sequence for secretion (e.g., the native sequence or the alpha factor preproleader), and selection markers for *S. cerevisiae* and *E. coli*. For example: pJRoC30, Gal10 promoter, *Myceliophthora thermophila* laccase gene and its T2 mutant (9, 10) with the native signal sequence, and the uracil and ampicillin selection markers.

4. Gene variants from different hosts or those created by random mutagenesis.
5. Restriction endonucleases (9, 10).
6. Proofreading polymerase, e.g., Pfu (Stratagene, USA).

2.1.3. Buffers and Solutions

1. Sterile chloramphenicol stock solution: 25 mg chloramphenicol in 1 mL of ethanol.
2. Minimal medium*: 100 mL of 6.7% sterile yeast nitrogen base, 100 mL of 19.2 g/L sterile yeast synthetic drop-out supplemented medium without uracil, 100 mL of 20% sterile raffinose, 700 mL of double distilled H_2O, 1 mL of 25 g/L chloramphenicol.
3. YP medium**: 10 g of yeast extract, 20 g of peptone and double distilled H_2O to a final volume of 650 mL.
4. Expression medium: 720 mL of YP, 67 mL of 1 M sterile KH_2PO_4 pH 6.0 buffer, 10 µL of 1 M sterile $CuSO_4$, 111 mL of 20% sterile galactose, 1 mL of 25 g/L chloramphenicol and double distilled H_2O to a final volume of 1,000 mL.
5. YPAD solution**: 10 g of yeast extract, 20 g of peptone, 100 mL of 20% sterile glucose***, 100 mg of adenine hemisulphate, 1 mL of 25 g/L chloramphenicol*** and double distilled H_2O to a final volume of 1,000 mL.
6. SC drop-out plates*: 6.7 g of sterile yeast nitrogen base, 100 mL of 19.2 g/L sterile yeast synthetic drop-out medium supplement without uracil, 20 g of bacto agar**, 100 mL of 20% sterile glucose, 1 mL of 25 g/L chloramphenicol, and double distilled H_2O to a final volume of 1,000 mL.
7. TAE-buffer (50×): 121 g of Tris-base, 28.05 mL of glacial acetic acid, 50 mL of 0.5 M ethylenediaminetetraacetic acid (EDTA) pH 8.0 and double distilled H_2O to a final volume of 500 mL.

   *Store in darkness (light sensitive).

   **Autoclave for 15 min at 121°C.

   ***Added after autoclaving.

2.2. Equipment

1. Thermocycler Mycycler (Biorad, USA).
2. Agarose gel electrophoresis system (Biorad, USA).
3. Gel Doc TM XR (Biorad, USA).
4. Spectrophotometer Uvikon 930 (Kontron Instruments, Italy).
5. Humidity shaker Minitron-Infors (Biogen, Spain).
6. Plate centrifuge Eppendorf 5810R (Eppendorf, Germany).
7. Liquid Handler Quadra 96-320 (Tomtec, USA).
8. Plate reader Versa Max (Molecular Devices, USA).

3. Methods

The IVOE methodology takes advantage of the high frequency of homologous recombination displayed by eukaryotic machinery to splice mutagenized DNA fragments, and of the yeast gap repair mechanisms to substitute in vitro ligation (4, 6) (see Note 5). In our example (Fig. 1), two PCR reactions are carried out using mutagenized primers (see Note 6) in order to produce two PCR fragments that share homologous sequences at the 3′and 5′ ends. These products already contain the mutagenized codons and they are then directly shuffled by *S. cerevisiae* in vivo through their sites of recombination to give rise to a whole gene. Likewise, recombination not only splices the two fragments in a complete gene but it also shuffles the mutagenized codons. The whole mutagenized gene possesses large overhangs that recombine with the ends of the linearized vector, thereby forming an autonomously replicating plasmid. It is not straightforward to determine which event takes place first (the splicing of the PCR fragments between themselves or their linkage to the linearized plasmid) and in fact, it is even likely that both phenomena happen simultaneously.

3.1. IVOE Method

1. Digest the plasmid for recombination – in vivo cloning – with appropriate restriction endonucleases (see Note 7).
2. Purify the opened plasmid by agarose gel extraction using a low melting point agarose at 4°C and with an applied voltage of less than 5 V/cm (distance between the electrodes of the unit, see Note 8). Measure the absorption of the preparation at 260 nm to determine its concentration (see Note 9).
3. Choose the residue(s) in your gene to be submitted to IVOE. For several residues, the distance between the mutations that will be recombined must be either smaller than 15 bp, so that they can be recombined within one primer, or greater than 120 bp, so that PCR products of at least 140 bp can be generated (see Note 10).
4. Synthesize a pair of sense and antisense primers for each mutation site (see Note 11), and prepare two external non-mutagenic primers (20–30 bp) that bind within the plasmid at a distance of at least 140 bp from the first mutagenic primer.
5. Carry out PCR reactions (see Note 12). The N-terminal non-mutagenic primer is paired with the most N-terminal mutagenic antisense primer in one PCR, while the corresponding sense primer is paired with the next antisense primer downstream in another reaction, and so on (Fig. 1).

6. Purify the PCR products following the conditions indicated in step 2.
7. Prepare an equimolar mixture of the PCR fragments to be recombined and add this equimolar mixture to the preparation of the open vector at a ratio of 4:1, with no less than 100 ng open plasmid per 100 µL of cell suspension (i.e., 400 ng of PCR products + 100 ng of linear plasmid; see Note 13).
8. Transform the mixture into fresh, competent *S. cerevisiae* cells using the Sigma transformation kit.
9. Plate the appropriate amount of the transformation mix on SC-drop-out plates and incubate at 30°C for 3 days.

3.2. Library Construction and Screening

10. Fill an appropriate number of 96-well plates with 50 µL minimal medium per well using an 8-channel pipette. Pick individual clones from the SC-drop out plates and transfer them into the 96-well plates. Column 6 of each plate, should be inoculated with the standard (parental) and one well (H1) should not be inoculated (control).
11. Wrap the plates in Parafilm and incubate them for 48 h at 30°C and 220 rpm in a shaker at 80–85% humidity (see Note 14).
12. Remove the Parafilm and add 160 µL of expression medium to each well. Reseal the plates with Parafilm and incubate them for 24 h under the conditions specified in step 11 (see Note 15).
13. Remove the Parafilm from the culture plates and centrifuge these plates (master plates) for 5 min at 2,000 × *g* at 4°C.
14. Transfer 20 µL of the supernatants (see Note 16) onto activity plates using a liquid handler (see Note 17).
15. Add 180 µL of activity assay solution to each well of the activity plate using the liquid handler. Mix and measure the activity in the assay with the plate reader (see Note 18).
16. With the data from the experiment, construct the library landscape (Fig. 2; see Note 19).
17. Sequence the best variants selected in the screen to define the new mutations (see Note 20).

4. Notes

1. Typically, error-prone PCR methods employed for in vitro evolution are limited to single-point mutations and they have a specific bias; therefore, a large fraction of the protein sequence space remains unexplored. Indeed, on average only

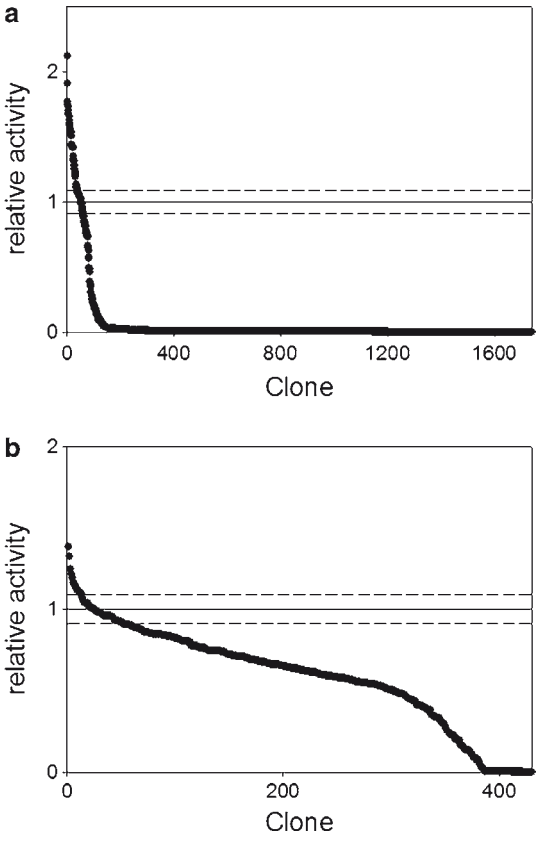


Fig. 2. Activity landscapes of libraries constructed by IVOE. (**a**) Landscape of a CSM library for two positions essential for enzyme activity (6). (**b**) Landscape of a site-directed recombination library for the evaluation of four positions (10). For each case, the library size was calculated by a binomial probability approximation (5, 14). The activity of the clones is plotted in a descending order. The *solid horizontal line* shows the activity of the parental type during the assay and the *dashed lines* indicate the coefficient of variation for the assay.

~5.7 amino acid substitutions are accessible to single-base mutations for any given amino acid residue. Thus, the optimization of beneficial mutations can be accomplished by further exploring those positions using saturation mutagenesis.

2. The sequence overlap extension (SOE) technique has been widely used to construct libraries by CSM (5, 11). SOE is a PCR-based method to recombine DNA sequences without relying on restriction sites, and to directly generate mutated DNA fragments in vitro. Based on gene splicing, degenerate oligonucleotides are designed so that the ends of the resultant PCR products contain complementary sequences. Each primer pair is synthesized with a mismatched random codon in the middle (such as –NNN–, where N can be A/T/C/G), flanked on both sides by nucleotides that specifically anneal to

the target region. Therefore, the DNA fragments must first be amplified by two separate PCR reactions giving rise to two DNA fragments that overlap at a specific region (Fig. 1). Subsequently, a third PCR reaction is performed in which the two PCR products are mixed, and the complementary sequences at their 3′ ends anneal and act as primers for one another. This step allows a new version of the original full-length sequence to be reassembled, where the target codons are effectively randomized. Finally, the entire amplified fragment, specifically/randomly mutagenized at one/several codon(s), must be ligated into a linearized vector in vitro to guarantee protein expression.

3. One advantage of in vivo recombination of *S. cerevisiae* is that during recombination, the proofreading apparatus of the yeast cell prevents the appearance of additional mutations that are common to in vitro methods (12, 13).
4. Apart from the indispensable in vitro ligation, the main bottlenecks in CSM based on SOE stem from the consecutive PCR reactions, and they are associated with poor reaction yields and the formation of by-products. Unlike SOE, IVOE proceeds by recombining PCR fragments, shuffling the mutagenized codons, and repairing the linearized vector with the help of specifically engineered overhangs (Fig. 1). In SOE, an additional PCR reaction and in vitro ligation are fundamental requisites. In contrast, these steps are avoided in IVOE by simply taking advantage of the eukaryotic machinery, and in particular, the frequency of homologous recombination displayed by *S. cerevisiae*.
5. Ligation of the mutant genes into expression vectors is in many cases a tedious and nonrobust step that needs fine-tuning for new plasmid–gene combinations. Yeast gap repair can substitute for ligation to give more reliable high transformation frequencies and to shorten the protocol for library expression. In terms of gap repair, the mutant gene inserts are cotransformed with the open plasmid that contains sequences homologous to the ends for the inserts at both ends. Homologous recombination combines these to form a complete plasmid (12).
6. IVOE can be used for CSM, site-directed recombination and site-directed mutagenesis. The size of the library generated is strictly dependent on the genetic code, the type of mutagenic codon, and the number of sites chosen for mutagenesis. It is advisable to use an NNG/C randomization strategy instead of NNN randomization for CSM libraries. NNG/C reduces the total number of variants, while all amino acids remain accessible and the complexity of the

library can still be maintained. Site directed recombination is useful to discard deleterious mutations that are close to beneficial ones and that escape elimination by homologous recombination. In such cases, site-directed recombination can be performed using primers synthesized for the mutation sites with 50% wild-type sequence. For site-directed mutagenesis by IVOE, the specific codon is introduced in the corresponding mutagenized primer.

7. The choice of the restriction endonucleases dictates the parts of the gene and the plasmid that will participate in the recombination event. The restriction sites and the positions of the primers used for the amplification of the genes regulate the length of the homologous sequences in the open plasmid and genes. This overhang length influences the recombination frequency between the gene and the open plasmid and, therefore, the transformation efficiency. Transformation efficiency does not change much if the homologous sequences are longer than 50 bp (good results can be also obtained with overhangs over 160 bp), although the efficiency is compromised if the overhangs are smaller than 50 bp. Generally, 20–50 bp homology is good for making libraries of ~15,000 clones per transformation (Sigma Yeast Transformation Kit, 100 μL of cells, 500 ng of DNA). It is worth noting that the primers must provide overhangs that specifically recombine without altering the open reading frame in order to guarantee the quality of the library (see Note 10).
8. It is very important to pay attention to the purification of the linearized plasmid. Gel extraction under mild conditions helps to prevent the degeneration of the linearized vector and is the most suitable option to avoid contamination with the circular plasmid.
9. The concentrations of linearized plasmid and PCR products can also be estimated from the gel using MW-ladders, but more accurate measurement enables the relative amounts of the ingredients of the DNA mix to be adjusted more precisely and, therefore, helps to prevent low efficiencies or mistakes during in vivo recombination.
10. Primers should be ~50 bp to provide an acceptable recombination area. Indeed, the homology region for in vivo recombination comes from the length of the primers and mismatches should preferably lie in the middle of the primer. Between the 3′ end of the primer and any mismatch there should be at least 10 bp of matching nucleotides to achieve proper annealing, while the 5′ end is less critical. From our experience, the best results are obtained with mismatches flanked by ~20 bp at both the 5′ and 3′ ends.

11. If the site is going to be submitted to saturation mutagenesis, degenerate primers must be used with NNG/C at the selected codon (see Note 6). If the mutation is going to be evaluated by site-directed recombination, include 50% wild type and 50% mutated sequence so that it will be reverted if deleterious.
12. Use proofreading polymerase and nonmutagenic conditions. As general rule, carry out always one more PCR reaction than the number of mutagenic primers synthesized.
13. To guarantee the recombination/transformation efficiency, it is advisable to explore different ratios of equimolar library/linearized vector. From our experience, it is useful to evaluate the mutant library with the preparation of the open vector in molar ratios ranging from 5:1 to 20:1, with no less than 100 ng of open plasmid per transformation reaction. Example: 400 ng mutant library (2 kb); 200 ng/kb + 100 ng open plasmid (10 kb); 10 ng/kb. Ratio mutant library: open plasmid = 20:1.
14. Parafilm is used to seal the gap between the plate and the lid. This prevents excessive evaporation of the medium, which would increase the variability of the screen. A humidity shaker is used for the same reason.
15. It is important to synchronize cell growth in all the wells. The expression levels may vary from one gene to another and, therefore, the incubation times must be studied for each specific case.
16. The expression plasmid used in this example includes the native signal sequence from the gene that targets its expression into the secretory pathway of *S. cerevisiae*. Therefore, additional cell lysis steps are not required and all the laccase expressed will be found in the supernatant after centrifugation. In cases where the protein is not secreted, lysis protocols must be incorporated prior to performing the activity assay.
17. It is highly advisable to use the liquid handler to achieve reproducible results and little variability when screening the library.
18. The activity assay may differ for each application and each particular protein. The reliability of this assay is reflected by a coefficient of variance below 10–15%, an indispensable prerequisite to identify beneficial mutations during the directed evolution experiment (14).
19. A statistical analysis of the CSM-libraries generated by NNG/C randomization implies that when two codons are randomized, 400 variants must be screened at the amino acid level rather than >3,000 variants at the DNA level. Multiple

saturation mutagenesis (three or more) generates libraries with large numbers of variants that cannot really be explored by conventional high-throughput methods (i.e., liquid microcultures in 96 well plates). In such cases, it is advisable to incorporate a solid format prescreening in order to discriminate the clones that exhibit weaker activity than the parental type.

20. Before sequencing selected variants, a secondary screen must be included to rule out the presence of false positives (15). In particular, rescreening with fresh transformant cells should be carried out in order to correctly compare the clones. This rescreening also synchronizes cell growth. First, selected variants are submitted to plasmid purification (Zymoprep yeast plasmid miniprep kit), and they are then overproduced and purified by transforming them into competent *E. coli* cells. Finally, the plasmids are then transformed into *S. cerevisae* again (five wells per variant) along with the corresponding parental type to estimate the improvement.

Acknowledgments

This work was supported by the Spanish Ministry of Science and Innovation(projectsCCG08-CSIC/PPQ-3706;PIE200880I033) and EU project FP7-NMP4-SL-2009-229255.

References

1. Bloom JD, Meyer MM, Meinhold P, Otey CR, MacMillan D, Arnold FH (2005) Evolving strategies for enzyme engineering. Curr Opin Struct Biol 15:447–452
2. Tao H, Cornish VW (2002) Milestones in directed enzyme evolution. Curr Opin Chem Biol 6:858–864
3. Chica RA, Doucet N, Pelletier JN (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Curr Opin Biotechnol 16:378–384
4. Zumárraga M, Domínguez CV, Camarero S, Shleev S, Polaina J, Martínez-Arias A, Ferrer M, de Lacey AL, Fernández V, Ballesteros A, Plou FJ, Alcalde M (2008) Combinatorial saturation mutagenesis of the *Myceliophthora thermophila* laccase T2 mutant: the connection between the C-terminal plug and the conserved 509VSG511 tripeptide. Comb Chem High Throughput Screen 11:807–816
5. Arnold FH, Georgiou G (eds) (2003) Directed evolution: library creation, methods and protocols, vol 231. Humana Press, Totowa, NJ
6. Alcalde M, Zumárraga M, Polaina J, Ballesteros A, Plou FJ (2006) Combinatorial saturation mutagenesis by in vivo overlap extension for the engineering of fungal laccases. Comb Chem High Throughput Screen 9:719–727
7. Zumárraga M, Camarero S, Shleev S, Martinez-Arias A, Ballesteros A, Plou FJ, Alcalde M (2008) Altering the laccase functionality by in vivo assembly of mutant libraries with different mutational spectra. Proteins 71:250–260
8. Abecassis V, Pompon D, Truan G (2000) High efficiency family shuffling based on multi-step PCR and in vivo DNA recombination in yeast: statistical and functional analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. Nucleic Acids Res 28:1–10

9. Zumárraga M, Bulter T, Shleev S, Polaina J, Martínez-Arias A, Plou FJ, Ballesteros A, Alcalde M (2007) *In vitro* evolution of a fungal laccase in high concentrations of organic cosolvents. Chem Biol 14:1052–1064
10. Bulter T, Alcalde M, Sieber V, Meinhold P, Schlachtbauer C, Arnold FH (2003) Functional expression of a fungal laccase in *Saccharomyces cerevisiae* by directed evolution. Appl Environ Microbiol 69:987–995
11. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51–59
12. Bulter T, Alcalde M (2003) Preparing libraries in *S. cerevisiae*. In: Arnold FH, Gergiou G (eds) Directed evolution library creation. Methods and protocols. Humana Press, Totowa, NJ, pp 17–22
13. Okkels JS (2004) *In vivo* gene shuffling in yeast: a fast and easy method for directed evolution of enzymes. In: Svendsen A (ed) Enzyme functionality: design, engineering, and screening. Marcel Dekker, New York, pp 413–424
14. Arnold FH, Georgiou G (eds) (2003) Directed enzyme evolution: screening and selection methods, vol 230. Humana Press, Totowa, NJ
15. Bulter T, Sieber V, Alcalde M (2003) Screening mutant libraries in *Saccharomyces cerevisiae*. In: Arnold FH, Gergiou G (eds) Directed enzyme evolution. Screening and selection methods, vol 230. Humana Press, Totowa, NJ, pp 99–108

Chapter 2

In Vitro Mutagenesis of *Brucella* Species

Thomas A. Ficht, Jianwu Pei, and Melissa Kahl-McDonagh

Abstract

Three major techniques have been employed for broad-range in vitro mutagenesis of *Brucella* species. Shotgun approaches capable of generating large libraries of randomly inserted transposon mutants include Tn5, mariner (Himar1), and mini-Tn5 signature-tagged mutagenesis. Allelic exchange has also been extensively employed for targeted gene deletion. In general, plasmid and transposon delivery into *Brucella* has relied upon electroporation; however, conjugation has also been successfully employed. Both approaches have been used to identify critical virulence determinants necessary for disease and intracellular survival of the organism. Perhaps more importantly these approaches have provided an opportunity to develop attenuated vaccine candidates of improved safety and efficacy. Future experiments are designed to identify individual functions that govern the interaction between host and agent and control intracellular trafficking and survival. Toward this goal, this chapter describes current approaches used to mutagenize *Brucella* spp.

Key words: *Brucella*, Conjugation, Transposon mutagenesis, Electroporation, Allelic exchange

1. Introduction

Brucella species are a group of Gram-negative, facultative intracellular bacteria that cause brucellosis, a worldwide zoonosis. There are at least six recognized species characterized biochemically, and serologically, but primarily on host preference. Recent identification from marine mammals suggests at least three additional species based upon isolation from porpoises, dolphins, and pinnipeds (1–3). Most work with *Brucella* has been restricted to the three classical species that affect agricultural animals and are readily transmissible to humans: *Brucella melitensis*, *Brucella abortus*, and *Brucella suis*. As a result, prevention of animal disease has been used as the primary approach to reduce human disease.

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_2,

These organisms do not express classic virulence factors such as toxins, hemolysins, etc., and express a lipopolysacchraride (LPS) component that is also greatly reduced in toxicity (4–6). The search for genetic factors important for virulence has been explored in many labs worldwide using a variety of mutagenic approaches.

The protocols outlined below describe techniques employed for broad range in vitro mutagenesis of *Brucella* species (7–11), as well as targeted gene deletion (12–14) and delivery methods (8, 14, 15). The absence of naturally occurring plasmids has led to the use of broad-range, low copy number plasmid, RK2-derived delivery vehicles, such as pBBR1mcs, pMR10, and pGL10, capable of replicating in the *Brucella* and providing experimental approaches to restore gene function (complementation) in the second step of the molecular version of Koch's postulates (16–19). Plasmids such as pSUP202-1/Tn5 (ColE1), pUT-mini-Tn5 (oriRK6), pSC189 (oriRK6), and pEX18Ap (oriT) do not replicate in *Brucella* and are used as delivery vehicles for random and site-specific mutagenesis (8, 9, 20–22). Allelic exchange is performed using pUC and pBluescript (ColE1) based replicons due to their inability to replicate in *Brucella* (13, 14). Curiously, despite the use of naturally occurring mechanisms, the *Brucella* spp. lack indigenous plasmids. One explanation for this is that *Brucella* occupies environments that are free of organisms capable of transferring plasmid via conjugation. Although acquired via ingestion, *Brucella* are rapidly taken up and transported to the lymphoreticular system, presumably limiting interaction with other organisms. Finally, experimental evidence indicates that in the absence of selection pressure none of the plasmids described persists in *Brucella*.

It is important to note that the introduction of any antibiotic resistance into class 3 agents such as the *Brucella* species requires the approval of regulatory agencies. Furthermore, antibiotics representing primary treatment regimens should never be considered for introduction. Before considering such experimentation, it is recommended to consult the Johns Hopkins ABX guide (http://prod.hopkins-abxguide.org/) listing of therapies. Select biological agents (SBAT), such as *B. melitensis*, *B. abortus* and *B. suis*, are under the oversight of the Centers for Disease Control or US Department of Agriculture (USDA). Introduction of antibiotic resistance requires the approval by the ISATTAC (Intergovernmental Select Agents and Toxins Technical Advisory Committee). Introduction of recombinant *Brucella* species that are not listed as SBAT requires the approval by the NIH/RAC through the local institutional biosafety committee (IBCs) (Table 1).

Table 1
Antibiotic resistance (Kirby-Bauer technique) expressed by Himar1 transposon mutants of *Brucella melitensis*

	Antibiotic[a]							
Gene	**Am(10)**	**Cm(30)**	**Do(30)**	**Gm(10)**	**Km(30)**	**Nm(30)**	**Rf(5)**	**St(50)**
16M	20	37	37	20	25	25	30	26
bacA	17	36	42	18	0	0	32	20
bacA	18	35	39	19	0	0	31	23
grsT	20	32	39	22	0	0	32	30
grsT	20	34	40	26	0	0	32	30
nifB/elp	20	33	38	20	0	0	30	26
nifB/elp	18	34	44	20	0	0	30	24
hlyD	23	36	45	22	0	0	27	26
aidA-hyp	22	37	43	21	0	0	28	24
dppB	20	35	35	14	0	0	30	26
mbl	19	41	40	22	0	0	30	25
mtrC	19	36	42	26	0	0	30	23
btuB	18	37	40	20	0	0	24	22
uspA	18	34	42	24	0	0	28	25
colV	14	38	40	20	0	0	30	20

Am Ampicillin, *Cm* Chloramphenicol, *Do* Deoxycyclin, *Gm* Gentamycin, *Km* Kanamycin, *Nm* Neomycin, *Rf* Rifampin, *St* Streptomycin

[a]All antibiotic concentrations in (μg/ml); the numbers in the table are the size of the zone (mm) surrounding the antibiotic disks. Mutants are resistant to Km/Nm.

2. Materials

2.1. Isolation of Bacterial Genomic DNA

1. Phenol saline: 0.5% (v/v) phenol, 0.15% (w/v) NaCl.
2. Tris–NaCl–EDTA Buffer (TNE): 10 mM Tris–HCl, pH 8.0, 10 mM NaCl, and 10 mM EDTA.
3. Triton X-100.
4. Lysozyme: 5 mg/ml in water.
5. Proteinase K: 20 mg/ml in water.
6. RNase: 20 mg/ml in water.

2.2. Mariner (Himar1) Transposon Mutagenesis

1. *B. melitensis* 16M American Type Culture Collection (ATCC) 23444. *B. abortus* S2308 (NADC) or *B. suis* 1330T ATCC 23444 as recipient (see Note 1).
2. *E. coli* β2155 [*thr*B1004 *pro thi str*A *hsd*S *lac*ZDM15 (F9 *lac*ZDM15 *lac*I^q *tra*D36 *pro*A1 *pro*B1) Δ*dap*A::*erm* (Erm^r)] *pir*::RP4 [::Km (Km^r) from SM10] as donor control.
3. *E. coli* β2155 bearing plasmid pSC189 as donor strain.
4. Tryptic soy broth (TSB) from Difco™.
5. Tryptic soy agar (TSA), TSB containing 1.5% (w/v) Bacto-Agar (Difco™).
6. Gentamicin (20 mg/ml) in water.
7. Kanamycin (Km) (100 mg/ml) in water.
8. Diaminopimelic acid (DAP) (50 mg/ml) in water.
9. Petri plates for bacterial growth on solid media.
10. TSA-Km (100 μg/ml).
11. TSA-DAP (50 μg/ml).
12. TSA-Km-DAP (100 μg/ml Km, 50 μg/ml DAP).
13. TSB-Km (100 μg/ml).
14. TSB-gentamicin (100 μg/ml).
15. Peptone saline: 1% (w/v) Bacto-peptone™ (Difco™) and 0.5% (w/v) NaCl.
16. 50% (v/v) glycerol in TSB.
17. J774.A1 macrophage (ATCC TIB-67).
18. Dulbecco's modified Eagle's medium (DMEM) with 10% (v/v) fetal bovine serum, 1 mM L-glutamine, and 1 mM non-essential amino acids.
19. 3.7% (w/v) formaldehyde.
20. Goat anti-*Brucella* serum.
21. Donkey anti-goat IgG Alexa Fluor 488 (Molecular Probes).
22. Phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM sodium phosphate dibasic, 2 mM potassium phosphate monobasic pH 7.4) (PBS).
23. 0.5% Tween-20 in distilled water, filter sterilized.
24. Triton X-100.
25. Restriction enzyme: *Hae*III, *Rsa*I.
26. T4 DNA ligase.
27. Wizard Genomic DNA Purification Kit (Promega®).
28. Inverse PCR primers: forward primer 5′-CAACACTCAACCC TATCTCG-3′; reverse primer 5′-CACTCAACCCTATCTCG GTC-3′ to amplify the region containing the interrupted loci.

29. QIAquick Gel Extraction Kit (Qiagen®).
30. PRISM™ Cycle Sequencing Kit (Applied Biosystems Inc, ABI).

2.3. Signature-Tagged Mutagenesis (STM)

1. Recipients as described in item 1 of Subheading 2.1.
2. Tryptic soy agar (TSA).
3. Tryptic soy broth (TSB).
4. Kanamycin (100 mg/ml) in water.
5. TSA-Km (100 µg/ml).
6. TSB-Km (100 µg/ml).
7. Suicide plasmid pool, pUT carrying signature-tagged mini-Tn5Km2 was obtained from Dr. D.W. Holden (Imperial College, London) and is described in detail in Subheading 3.5 below (10, 23).
8. Primer P2: 5′-TACCTACAACCTCAAGCT-3′.
9. Primer P3: 5′-CATGGTACCCATTCTAAC-3′.
10. Primer P4: 5′-TACCCATTCTAACCAAGC-3′.
11. Primer P5: 5′-CTAGGTACCTACAACCTC-3′.
12. QIAprep Spin Miniprep Kit (Qiagen®).
13. Balb/c mice from commercial vendor.
14. Nitrocellulose membrane circles.
15. 20× SSC: 3 M NaCl, 0.3 M sodium citrate, pH 7.0.
16. Prewash solution: 50 mM Tris–HCl (pH 8.0), 1 M NaCl, 1 mM EDTA, 0.1% (w/v) SDS.
17. Prehybridization/hybridization solution: 5× SSC, 0.5% (w/v) nonfat dried milk, 2.5% (w/v) denatured salmon sperm DNA, 1% (w/v) SDS.
18. ^{32}P-labeled STM probe prepared by PCR amplification of genomic DNA extracted from input and output pools (described in Subheading 3.8 below) using ^{32}P-dATP in the PCR.
19. QIAquick PCR Purification Kit (Qiagen).

2.4. Targeted Gene Deletion

1. Recipients as described in item 1 of Subheading 2.2.
2. *E. coli* Top10 [F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*Δ139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G] from Invitrogen.
3. pBluescriptKSII$^+$ from Stratagene (f1+ origin, Ap^R, β-galactosidase α-fragment, ColE1 origin, *lac* promoter).
4. pKD4 (FLP/FRT, Km^R) from Dr. H.P. Schweizer (24).
5. pEX18Ap (*sacB*, Ap^R) from Dr. H.P. Schweizer (25).

6. Primers:
   (a) Forward primer (F_{Km}) to amplify Km cassette from pKD4: 5′-CGGGATCCCGCACGTCTTGAGCGATTGTGTAGG-3′ (with BamHI linker)
   (b) Reverse primer (R_{Km}) to amplify Km cassette from pKD4: 5′-CGGGATCCCGGGACAACAAGCCAGGGATGTAAC-3′ (with BamHI linker)
   (c) Forward ($F_{5'}$ and $F_{3'}$) and reverse ($R_{5'}$ and $R_{3'}$) primers engineered to amplify flanking regions of the gene(s) to be deleted (sequences are specific for the gene(s) to be deleted) and to contain restriction sites for cloning $F_{5'}$ (site 1), $R_{3'}$ (site 2), $R_{5'}$, and $F_{3'}$ (site 3).
7. QIAquick Gel Extraction Kit (Qiagen®).
8. SOC: 6% trypticase soy broth (w/v), 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, and 20 mM glucose.
9. SOC-B: 6% trypticase soy broth (w/v), 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose.
10. Sucrose broth: TSB supplemented with 6% (w/v) sucrose, lacking salt and antibiotics.
11. Sucrose agar: TSA supplemented with 6% (w/v) sucrose, lacking salt and antibiotic.
12. Luria–Bertani broth (LB).
13. LB agar [LB broth containing 1.5% (w/v) Bacto-Agar (Difco™)].
14. TSB.
15. TSA [TSB containing 1.5% (w/v) Bacto-Agar].
16. Ampicillin (Am, 100 mg/ml) in water.
17. Kanamycin (Km, 100 mg/ml) in water.
18. Carbenicillin (Cb, 100 mg/ml) in water.
19. 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal, 20 mg/ml) in water.
20. TSA-Am (100 μg/ml).
21. TSA-Km (100 μg/ml).
22. TSA-Cb (100 μg/ml).
23. Sigma Miniprep Kit.

2.5. Intracellular Survival Assay

1. J774.A1 macrophage (ATCC TIB-67).
2. DMEM with 10% (v/v) fetal bovine serum, 1 mM L-glutamine, and 1 mM nonessential amino acids.
3. Gentamicin (Gm, 20 mg/ml) in water.
4. 0.5% (v/v) Tween-20 in distilled water, filter sterilized.

5. Peptone saline.
6. TSA.
7. TSA-Km (100 μg/ml).

3. Methods

3.1. Isolation of Genomic DNA

There are many approaches to isolate genomic DNA from *Brucella*. The method described provides sufficient amounts of genomic DNA for Southern blotting and PCR amplification and requires little manipulation of the bacteria.

1. Bacteria are grown overnight with agitation at 37°C in TSB and pelleted by centrifugation (10,000 × *g* for 5 min).
2. Pelleted bacteria are resuspended in the same volume of phenol saline and heated for 1 h at 60°C to kill the organism.
3. DNA sufficient for Southern blotting or PCR amplification may be obtained from a cell culture volume as small as 0.3 ml, as follows.
4. The bacteria are pelleted as described above and the supernatant removed.
5. The cell pellet is washed in 1.0 ml TNE by resuspension using a pipette.
6. Pelleting and washing is repeated at least twice.
7. The cells are then resuspended in 135 μl of TNE.
8. The cell suspension is diluted with 135 μl TNE containing 2% (v/v) Triton X-100.
9. Thirty microliters of freshly prepared lysozyme solution (5 mg/ml) is added and mixed by tapping the tube.
10. The suspension is incubated at 37°C in a water bath for 30 min.
11. Fifteen microliters of proteinase K solution (20 mg/ml) is added and the suspension mixed by inversion.
12. The mixture is incubated at 65°C in a water bath for at least 2 h.
13. Heat-treated RNase is added to a final concentration of 10 μg/ml.
14. These DNA preparations are best stored at –20°C until used.

3.2. Mariner Transposon Mutagenesis

Transposons of the Mariner family integrate nonspecifically at T/A base pairs, and do not require species-specific host factors for efficient transposition. In this protocol, plasmid vector pSC189, containing both the hyperactive transposase C9 and

transposon terminal inverted repeats flanking a kanamycin resistance gene, is used to deliver Himar1 transposable element into the *B. melitensis* 16M genome by conjugation. Conjugation is performed efficiently and rapidly in less than one generation in order to minimize the formation of siblings while assuring the highest level of genome coverage [8].

1. *Brucella* are removed from frozen stock and streaked for growth on TSA plates for 48–72 h (see Note 2).
2. Stock cultures should be checked to make sure that kanamycin resistance is either undetectable or significantly below the frequency observed for transposition. Briefly, frozen stocks are removed from the freezer and evaluated for growth on TSA and TSA-Km plates. Following incubation, the appearance of spontaneous kanamycin resistant (Km^R) colonies is assessed.
3. *E. coli* β2155 with pSC189 is grown for 24 h on TSA supplemented with DAP (50 μg/ml) and Km (100 μg/ml).
4. The bacteria are harvested from plates prepared in steps 1 and 3 above into 5 ml of peptone saline supplemented with DAP (50 μg/ml).
5. Equal volumes of *Brucella* and *E. coli* β2155 bearing pSC189 are mixed together to provide a donor to recipient ratio of approximately 1:100.
6. Nitrocellulose filters (25 mm diameter) are placed on the surface of TSA-DAP plates that have been dried by incubation overnight at 37°C.
7. Two hundred microliters of bacterial conjugation mixtures are pipetted onto individual nitrocellulose filters.
8. The liquid is rapidly absorbed by the dried plates that are then incubated at 37°C for 2 h.
9. The bacterial conjugation mixtures on the nitrocellulose filters are removed by resuspension into 5 ml of peptone saline.
10. Serial tenfold dilutions of the conjugation mixtures are prepared in peptone saline.
11. One hundred microliters of each dilution are plated on TSA-Km and incubated at 37°C for 3 days.
12. The remaining bacterial conjugation mixture is stored at –80°C in peptone saline supplemented with 15% (v/v) glycerol.
13. Colony forming units (CFU) are determined and used to calculate conjugation efficiency and the number of plates necessary to amplify the mutant bank.
14. There should be no Km-resistant growth from control conjugation mixtures, which include *E. coli* β2155 alone or *B. melitensis* 16M alone.

15. Conjugation mixtures are diluted with peptone saline based on the conjugation efficiency to give a concentration of conjugant about 100–300 CFU/100 µl.
16. One hundred microliters of diluted conjugation mixture is spread on the surface of TSA-Km.
17. The plates are incubated at 37°C for 3 days.
18. Well-isolated single colonies are selected using sterilized toothpicks and used to inoculate 100 µl of TSB-Km in 96-well microtiter dishes.
19. The dishes are incubated at 37°C for 2 days.
20. Duplicate plates are prepared by inoculating 10 µl from each well of the microtiter dishes into new dishes containing 90 µl TSB-Km.
21. These dishes are incubated at 37°C for 48 h.
22. Fifty microliters of 50% (v/v) glycerol is added to each well and mixed.
23. The plates are sealed with parafilm and stored at −80°C (see Fig. 1).

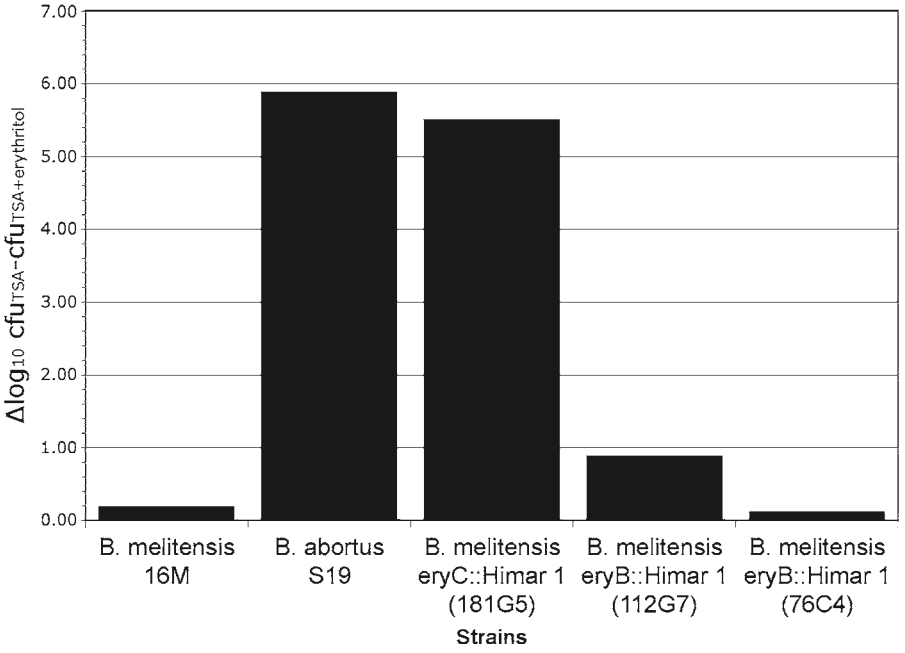


Fig. 1. Erythritol sensitivity was first described in *B. abortus* S19. The cause of the defect is believed to be the buildup of a toxic intermediate (D-erythrulose-1-phosphate, a product of the reaction catalyzed by eryB). In S19, the locus is truncated by a deletion that removes portions of the genes encoding EryC and the repressor protein EryD. The result is uncontrolled expression of the locus and buildup of the toxic product produced in a reaction catalyzed by EryB and the failure to reduce its concentration due to the absence of EryC. Interruption of eryC by Himar1 has an identical effect to the observed deletion. In contrast, Himar1 interruption of eryB has little effect on survival in the presence of erythritol. This may be explained by the failure to produce the toxic product.

3.3. Identification of Attenuated Mutants Using Immunofluorescence

1. J774.A1 macrophage form monolayers on the flat bottom of 96-well dishes when seeded at a density of 5×10^4 cells/well in 0.1 ml DMEM 1 day prior to infection.
2. Bacterial cultures are removed from frozen stock and grown on TSA-Km plates (mutants) or TSA (16M) for 72 h.
3. J774.A1 macrophage monolayers are infected with *B. melitensis* at a multiplicity of infection (MOI) of 100:1 using 10 µl of bacterial culture.
4. Uninfected cells are used as negative control.
5. Cell culture plates are centrifuged for 5 min. at $200 \times g$ to initiate the infection.
6. The macrophages are incubated at 37°C in an atmosphere containing 5% CO_2 for 20 min.
7. The culture supernatant in each well is removed using a sterile pipette and replaced with 100 µl of TSB supplemented with 100 µg/ml of gentamicin to kill extracellular bacteria.
8. The macrophages are incubated at 37°C in atmosphere containing 5% CO_2 for 48 h.
9. The culture media and the monolayers are washed twice with an equal volume of PBS as described in step 7 above.
10. Three hundred microliters of 3.7% (v/v) formaldehyde in PBS is added to each well, and the plates are incubated at room temperature for 1 h to fix the cells and any intracellular bacteria.
11. Each well is washed with 300 µl of PBS three times as described in step 9 above.
12. Fifty microliters of goat anti-*Brucella* serum (1:1000) diluted in PBS-TT is added to each well.
13. The plates are incubated at room temperature for 1 h.
14. Each well is washed three times with 300 µl of PBS-T as described in step 11 of Subheading 3.2.
15. PBS-T is removed and replaced with 50 µl of donkey anti-goat IgG Alexa Flour 488 (1:1000) diluted in PBS-TT.
16. Mutant virulence is determined microscopically by evaluating fluorescence intensity compared with positive and negative controls. Mutants that are unable to replicate within the macrophage are present in reduced number or in fewer cells compared to the control *B. melitensis* wildtype. As such, wildtype fluorescence is stronger than mutant fluorescence, and uninfected macrophages are expected to display no fluorescence.
17. A second round of fluorescence screening is used to provide statistically valid results and to confirm the attenuated phenotype.

3.4. Identification of Interrupted Loci

1. Genomic DNA isolated from attenuated mutants is digested with restriction enzymes *Hae*III or *Rsa*I.
2. The digested DNA fragments are self-ligated and amplified by inverse PCR (see Subheading 2.2, step 28 above for primers).
3. Inverse PCR is performed by heating to 95°C 4 min., followed by 30 cycles of (95°C 30 s, 57°C 30 s, 72°C 90 s), and 72°C for 7 min.
4. Agarose gel electrophoresis is performed to ensure the production of a unique PCR product, reflecting a single transposon insertion.
5. PCR products are purified from gels using QIAquick Gel Purification Kit.
6. The purified products are sequenced using reverse primer (see Subheading 2.2, step 28 above) with ABI PRISM™ Cycle sequencing kits.
7. The sequence obtained is compared to the *B. melitensis* genome sequence available in GenBank using any of the commercially available software packages.

3.5. Signature-Tagged Mutagenesis (STM)

Signature-tagged mutagenesis was developed for in vivo selection of Tn5 transposon mutants that are defective in colonization of specific tissue in the host. The advantage of this approach is that the signature tags can be used to distinguish individual mutants within a pool of mutants permitting distinction of survival characteristics of multiple mutants in a single host. In practice, recovery of the organism limits the diversity of the input pools to 50–100 mutants per host and multiple hosts are used for statistical evaluation. Tags are arrayed on nitrocellulose for comparison of recovery based on hybridization of the tags amplified from input and output pools of mutants. In this protocol, tagged transposons were obtained from Dr. D.W. Holden (see Subheading 2.3, step 7 above) and individual tags identified that readily amplify and provide a strong hybridization signal without cross hybridization between tags [26]. Plasmids containing these tags are used to generate *B. melitensis* mutants by performing separate conjugations for each tag and selecting 80 mutants or more per tag. Pools of mutants are arrayed in groups of 48 reflecting the 48 unique tags in 96-well microtiter dishes for replica plating. Pools of 48 mutants were washed off the plates and used to infect the host (input pool) and at desired time points after infection, bacteria are recovered from selected tissues (output pools). For *Brucella* infection, the spleen is the preferred tissue, although the liver also has elevated colonization, and the lymph nodes may hold special interest. The signature tags present in the output pool are PCR-amplified and labeled as are the signature tags present

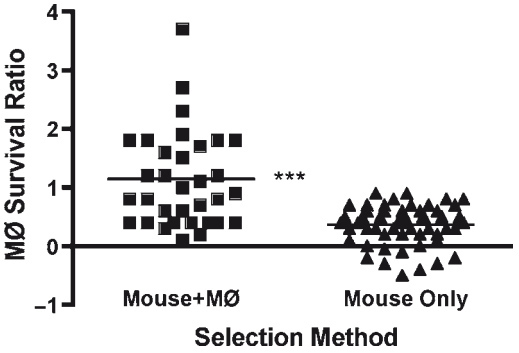


Fig. 2. Comparison of macrophage survival of attenuated *Brucella melitensis* mutants. Mutants were obtained by screening for survival in mice and macrophage (Mϕ) in culture and divided into two groups based on identification in the macrophage screen. The replication ratio (CFU48h/CFU0h) for each mutant was determined relative to the parental strain and presented as the $\log_{10}$ of wild type to mutant. Mutants that were only identified in the mouse model exhibited an average survival ratio that was significantly lower than those mutants identified using the macrophage and confirmed in the mouse model. The enhanced sensitivity of the mouse model may be explained in part by the contribution of extracellular killing present in the mouse model, but missing from macrophage screening. The horizontal line represents the average mutant survival ration from the group.

in the input pool. Following the removal of the flanking arms by *Hind*III digestion, the probes are hybridized to replica arrays of the original signature tags corresponding to the pool. Attenuated mutants are identified based upon differential hybridization signals for input and output pools, tags present in attenuated mutants will not be amplified [10, 26] (see Fig. 2).

3.6. Identification of Useful Signature-Tagged Transposons

1. Signature-tagged miniTn5Km2 transposons or others are prepared as described elsewhere [26]. Briefly, DNA tags are prepared from the variable oligonucleotide pool, RT1 (5′-CTAGGTACCTACAACCTCAAGCTT-$[NK]_{20}$-AAGCTTGGTTAGAATGGGTACCATG-3′) in a 100 µl volume PCR containing 1.5 mM $MgCl_2$, 50 mM KCl, and 10 mM Tris–Cl (pH 8.0) with 200 pg of RT1 as target; 250 µM each of dATP, dCTP, dGTP, dTTP; 100 pM of primers P3 and P5; and 2.5 U Amplitaq (Perkin-Elmer). Cycling conditions are 30 cycles of 95°C for 30 s, 50°C for 45 s, and 72°C for 10 s. The PCR product is gel purified (see Subheading 3.4, step 4 above) and digested with restriction enzyme KpnI prior to ligation into pUT-mini-Tn5Km2.
2. *E. coli* bearing plasmid are grown on TSA-Km plates and individual colonies selected using toothpicks to inoculate fresh TSB-Km in the wells of microtiter dishes.

3. The bacteria are replica-plated onto the surface of TSA-Km plates using a 48 prong replica plater.
4. The plates are incubated at 37°C for 16 h.
5. The plates are used for plasmid isolation and tag amplification, as well as "colony lifts".
6. Each well contains a unique signature-tagged transposon.
7. *E. coli* are washed off the surface of the plates by adding 5–10 ml sterile LB and gently scraping with a sterile plate spreader. The liquid is then removed with a sterile pipette, and pooled plasmids are purified using commercial kits, such as the QIAprep Miniprep kit.
8. Signature tags are labeled by incorporation of P^{32}-dATP during PCR amplification with primers P2: 5′-TACCTACAACCT-CAAGCT-3′ and P4: 5′-TACCCATTCTAACCAAGC-3′ using conditions described in step 1 of Subheading 3.6.
9. The radioactive tags are used as the probe during hybridization with the colony lifts.
10. Colony lifts are obtained by overlaying the plates with 100 mm nitrocellulose circles that are peeled back in order to transfer the colonies to the nitrocellulose.
11. The nitrocellulose circles are laid colony-side up on a stack of filter paper soaked with 0.4 N NaOH/1.5 M NaCl for 5 min to lyse the cells and denature the genomic DNA, and then neutralized with 0.5 M Tris–HCl, pH 7.4/1.5 M NaCl using the same method.
12. The nitrocellulose filters are baked at 80°C for 2 h under vacuum to fix the DNA to the membranes.
13. The filters are wetted with 2× SSC, and then transferred to glass tubes or seal-a-meal bags and prehybridized in excess solution at 68°C for at least 2 h.
14. Hybridization is performed for 16 h at 68°C in a minimal volume of solution containing ^{32}P-labeled probe (100,000 dpm/cm^2 nitrocellulose).
15. The filters are washed in 2× SSC, 0.1% (w/v) SDS twice for 15 min at room temperature, and then in 0.2× SSC, 0.1% (w/v) SDS twice for 15 min at 68°C.
16. The membranes are air-dried on Whatman 3MM paper at room temperature and sealed in plastic bags.
17. The membranes are exposed to X-ray film overnight.
18. Tags that are useful for screening are identified as those producing strong signals on the colony lifts due to stable hybridization, and tag amplification without cross hybridization.

3.7. Transposon Mutagenesis and Mutant Bank Assembly

1. Plasmid pUT containing transposon miniTn5Km2 with signature tags are introduced into *B. melitensis* by conjugation as described in Subheading 3.2 above.
2. Following conjugation, serial dilutions are prepared and the transformation efficiency is determined by plating portions of the serial dilutions on TSA-Km plates.
3. The plates are incubated at 37°C for 3 days. Depending on the number of signature tags employed ($n=48$) and the complexity of the genome, between 80 and 400 mutants are picked from 48 conjugations with plasmid having different tags (see Note 3).

3.8. Mutant Screen and Identification of Attenuated Mutants

1. Pools are assembled from 48 mutants grown in the wells of microtiter dishes and replica-plated as described above (Subheading 3.6, step 3 above).
2. Forty-eight mutants from each plate are pooled by washing the cells from the surface of the replica plate.
3. A 1-ml portion of the mutant pool is removed for genomic DNA extraction using lysozyme and proteinase K treatment (input pool) (see Subheading 3.1 above).
4. The concentration of the bacterial pool is adjusted to approximately 1×10^7 CFU/ml with PBS.
5. Six Balb/c mice are inoculated (i.p.) with 0.1 ml of the pooled bacteria.
6. Three mice are euthanized at 2- and 8-week post-infection.
7. Spleens are removed and homogenized in PBS.
8. Serial tenfold dilutions are prepared in peptone-saline.
9. Homogenates and dilutions (0.1 ml) are plated on TSA-Km.
10. Plates are incubated at 37°C for 4 days.
11. Bacteria are collected from plates containing 1,000–5,000 colonies (output pool) (see Note 4).
12. Genomic DNA is isolated from input pools and output pools as described in Subheading 3.1 above.
13. PCR amplification of signature tags using input pool and output pool genomic DNA as template and primers P2 and P4 is performed as described in step 8 of Subheading 3.6.
14. Signature tags are labeled by incorporation of ^{32}P-dATP during PCR amplification.
15. The PCR tags are digested with *Hind*III (1 U enzyme/μg DNA) to release the shared flanking regions.
16. Labeled tags are hybridized to the corresponding colony blots prepared from 96-well plates generated in Subheading 3.6 above.

17. Mutants that hybridize to the probe from the input pool but weakly or not to the probe from the output pool are attenuated. Failure to amplify signature tags is due to their absence from the output pool.
18. Mutant attenuation is confirmed using intracellular survival assay described in Subheading 3.16 below.

3.9. Identification of Interrupted Loci

1. Genomic DNA isolated from attenuated mutants is digested with *Rsa*I, self-ligated and used as template for inverse PCR.
2. Inverse PCR conditions include an initial 4 min. at 95°C, 30 cycles (95°C 30 s, 57°C 30 s, 72°C 90 s), and a final elongation at 72°C for 7 min with forward 5′-GCCGAACTTGT GTATAAGAGTCAG-3′ and reverse 5′-AAAGGTAGCGTT GCCAATG-3′ primers.
3. PCR products are gel purified using QIAquick PCR Purification Kit and the products sequenced using the reverse primer.
4. DNA sequences are compared to the *B. melitensis* sequence in GenBank to identify the disrupted genes.

3.10. Targeted Gene Deletion

In order to eliminate specific genes of interest, primers are designed to amplify sequences flanking the segment or gene to be deleted. The flanking regions referred to are located 5′ and the 3′ to the gene of interest, and are joined to each other using overlap extension PCR, i.e., the reverse primer of the 5′ fragment and the forward primer of the 3′ fragment contain complimentary sequences, as well as unique restriction sites [27]. Deletion is typically constructed to avoid downstream polar effects, where genes downstream in an operon would be affected during transcription. Genes in an operon are typically deleted so as to severely truncate the gene product, but avoid having the ribosome disrupted. The 5′ and 3′ fragments are amplified in separate reactions, gel-purified, and PCR-amplified in the same reaction to produce a joined product. The final products are digested with the restriction enzymes engineered into the primers and the final fragment is gel purified for cloning into pBluescript KSII$^+$. Antibiotic resistance cassettes are inserted between the 5′ and 3′ fragments and the construct is used to generate marked deletion mutants. To create unmarked deletion mutants (free of foreign DNA or selectable markers), the joined PCR product (without the kanamycin cassette) is cloned into the plasmid pEX18Ap, which contains *sacB*, encoding levansucrase. This *sacB* gene product is lethal to the cell, and bacteria possessing the plasmid are eliminated in the presence of sucrose. Thus, the presence of sucrose selects for the loss of the plasmid, and growth of the unmarked knockout.

3.11. Recombinant Plasmid Construction

1. Sequences flanking the gene of interest are amplified via PCR using conditions determined empirically for the gene and primers employed (see Note 5).
2. Amplify the 5′ fragment using *Brucella* genomic DNA as template and upstream primers $F_{5'}$ and $R_{5'}$ (see Subheading 2.4, step 6c above). PCR conditions depend upon the gene size and GC%.
3. Amplify the 3′ fragment using *Brucella* genomic DNA as template and downstream primers $F_{3'}$ and $R_{3'}$ (see Subheading 2.4, step 6c above).
4. The 5′- and 3′-fragments are joined during a second amplification in which only the primers $F_{5'}$ and $R_{3'}$ are used.
5. The amplification product is isolated by gel electrophoresis and purified using the QIAquick Gel Extraction Kit.
6. The PCR product and plasmids are digested with appropriate restriction enzymes to clone the insert into either pBluescript KSII$^+$ (the first step in plasmid construction to develop marked *Brucella* mutants) or pEX18Ap (to develop unmarked *Brucella* mutants).
7. Ligation is performed overnight at 15°C using a 3:1 molar ratio of insert to plasmid DNA and T4 DNA ligase.
8. Antibiotic resistance cassettes such as kanamycin (nptII) are amplified via PCR from plasmid template pKD4 with specific primers: F_{Km} 5′-CGGGATCCCGCACGTCTTGAGCGATTG TGTAGG-3′ (with BamHI linker) and R_{Km}5′-CGGGATCC CGGGACAACAAGCCAGGGATGTAAC-3′ (with BamHI linker) (see Subheading 2.4, step 6).
9. The primers F_{Km} and R_{Km} have been constructed to contain the same unique restriction enzyme sites (in this case $BamH_I$) as the junction of the 5′- and 3′-fragments (5′-Km^R-3′) (see Subheading 2.4, step 6c above).
10. The amplified resistant cassette is isolated by gel electrophoresis and purified using Qiagen's QIAquick Gel Extraction Kit.
11. Ligation of the antibiotic resistance cassette between the upstream and downstream regions is performed as described in step 6 of Subheading 3.11.

3.12. Transformation and Selection of Recombinant Plasmids

1. Competent *E. coli* are transformed with plasmid DNA as described by the manufacturer, and the culture is plated onto solid media supplemented with appropriate antibiotic (depending upon the plasmid backbone) and X-gal (20 μg/ml).
2. Colonies are selected by blue–white screening after overnight growth at 37°C and individual white colonies are used to prepare fresh cultures in (LB) broth supplemented with appropriate antibiotic (depending upon plasmid encoded resistance).

3. Recombinant plasmids are purified (Sigma Miniprep Kit) and verified using restriction enzyme digestion.
4. Bacterial frozen stocks are prepared in LB broth supplemented with 50% glycerol (v/v) and stored at −80°C.

3.13. Creation of Marked Mutants by Electroporation

1. *Brucella* are harvested from the surface of confluent plates after 3–4 days of growth at 37°C.
2. The bacteria are pelleted by centrifugation at 5,000 × *g* for 15 min at 4°C.
3. The cell pellet is washed three times with sterile, ice-cold water by repeating the previous step and is then resuspended in 1-ml ice-cold water.
4. Seventy microliters of the cell suspension is placed into a pre-chilled 1-mm gap electroporation cuvette along with 1 μg plasmid in 1–5 μl water (see Note 6).
5. The mixture is electroporated using a BT× 3000 apparatus set at 2.2–2.5 kV and 246 Ω.
6. The bacterial suspension is immediately diluted with 1 ml of SOC-B in the cuvette, transferred to a microfuge tube and subsequently incubated overnight at 37°C with agitation.
7. One hundred microliters of cell suspension is spread on the surface of TSA-Km plate and incubated at 37°C for 3 days.
8. If necessary (low efficiency), the remaining cell suspension is pelleted by centrifugation at 10,000 × *g* for 1 min, resuspended in TSB-Km and plated on TSA-Km plates.
9. Individual colonies are replica-plated onto TSA-Km and TSA-Cb.
10. Marked deletion mutants are kanamycin resistant (Km^R) and Carbenicillin sensitive (Cb^S) due to loss of the plasmid during allelic exchange.
11. Following verification (Subheading 3.15 below) individual colonies are resuspended in TSB containing 50% (v/v) glycerol and stored frozen at −80°C.

3.14. Creation of Unmarked Mutants by Electroporation

1. Marked deletion mutants are harvested from the surface of confluent plates after 3–4 days of growth at 37°C.
2. Repeat the procedure described in the previous section to prepare electrocompetent cells.
3. Electroporation is performed using plasmid pEX18Ap-containing the insert composed of 5′ and 3′ fragments (see Note 7) and bacterial suspensions are plated on TSA-Cb.
4. Individual colonies are replica-plated onto three different solid media: TSA-Cb, TSA-Km, and sucrose plates.
5. Co-integrants form due to homologous recombination between genomic and plasmid gene copies, and are Cb^R-, Km^R- and sucrose-sensitive.

6. Sucrose-sensitive colonies are used to inoculate 5 ml of sucrose broth and incubated at 37°C for 24 h with agitation (see Note 8).
7. The culture is diluted 10- to 100-fold with TSB-sucrose and 100 µl of undiluted and diluted cultures are plated onto TSA-sucrose plates.
8. The plates are incubated at 37°C for 3–5 days.
9. Sucrose-tolerant colonies are replica-plated onto TSA-sucrose and TSA-Km plates.
10. Unmarked deletion (and the original parental organism) mutants are sucrose-tolerant and kanamycin-sensitive (see Note 9) and require genetic analysis to distinguish.

3.15. Confirmation of Gene Deletion

1. Genomic DNA is extracted from sucrose-tolerant Kan^S colonies and PCR amplification of the target gene uses the 5′-upstream and 3′ reverse primers described above in step 6c of Subheading 2.4.
2. The choice of primers and size of the amplification product depends on the gene deleted and the sequence flanking the gene of interest.
3. Amplification of a deleted locus produces a smaller PCR product that may be distinguished from either revertants to wildtype and/or the parental strain.

3.16. Confirmation of Mutant Attenuation

1. Stock cultures of mutants or the parental wildtype 16M are inoculated into 5 ml TSB or TSB-Km and incubated at 37°C for 48–72 h with agitation.
2. Fresh cultures are prepared by diluting the stock cultures (1:1000) into fresh TSB or TSB-Km and incubated at 37°C for 24 h with agitation.
3. Macrophages are seeded into 24-well plates (2.5×10^5/well in 0.5 ml DMEM) 1 day prior to infection.
4. The bacteria are pelleted by centrifugation and resuspended in an equal volume of PBS. This step is repeated twice and following the last centrifugation the bacterial suspension is diluted fivefold.
5. Approximately 10 µl of bacterial suspension is added to each well of the microtiter dish reflecting bacteria to macrophage ratio or MOI of 50:1.
6. The microtiter dishes are centrifuged at $200 \times g$ for 5 min at room temperature and then incubated at 37°C for 20 min.
7. The supernatant is removed and the infected cell monolayer is washed with PBS. This step is repeated three times.
8. Fresh DMEM is added to each well containing gentamicin (50 µg/ml) to destroy extracellular bacteria.

9. The dishes are incubated at 37°C up to 48 h. At that time the DMEM is removed and 1.0 ml 0.5% (v/v) Tween-20 is added to lyse macrophages and release intracellular bacteria.
10. Tenfold dilutions of the lysate are prepared using PBS and the dilutions are plated in triplicate on TSA with or without kanamycin.
11. The plates are incubated at 37°C for 3 days, at which time the bacterial recovery is determined, i.e., CFU/well.
12. Recovery of mutants is compared to wildtype organism to observe attenuation.

3.17. Summary

The work described has been developed in or laboratory over the past 10 years and is the culmination of the work of several individuals as well as collaborators who generously provided reagents, cloning and delivery vehicles, and signature-tagged transposons. The methods presented are meant to describe their use in the development of *Brucella* mutants and not to imply their original development that is described in those works provided in the bibliography. The work has resulted in the identification of several important virulence factors and vaccine candidates currently under evaluation.

4. Notes

1. It is best to passage the bacteria through a host organism, including mice or small ruminants. In this way, the parent organism may be expected to exhibit the highest level of virulence.
2. To ensure that the starting organism is fully virulent, it is also best to streak for isolation and to select a smooth colony to inoculate either fresh plates or broth. This will help to minimize the presence of spontaneously appearing rough organisms.
3. These mutants will be screened for susceptibility to ampicillin to eliminate strains (fewer than 2% of mutants) carrying cointegrates of the suicide vector pUT inserted in the chromosome. As *B. melitensis* is not allowed to replicate before plating the transformants, the isolation of siblings with this procedure is unlikely.
4. This number of organisms is sufficient to provide recovery of virulent organism with 95% confidence.
5. Primers used contain restriction sites for subsequent cloning.
6. Plasmids used for electroporation should be eluted into water prior to use in electroporation, since high salt concentrations present in commercial elution buffers negatively affects the procedure.

7. The unmarked plasmid, containing the *sacB* gene, insert, and *bla* gene (same function as Carbenicillin) is used for electroporation into marked deletion strains. Utilizing the newly created marked strain enhances selection, since loss of kanamycin resistance identifies unmarked mutants formed via allelic exchange.
8. After 24 h of growth, the cultures will not look saturated because the sucrose was toxic to the majority of the cells, but there are enough cells for plating.
9. Unmarked deletion mutants are sucrose tolerant, resulting from the loss of the integrating plasmid-containing *sacB*, and kanamycin-sensitive, since the original kanamycin cassette is replaced during plasmid integration.

Acknowledgments

We would like to acknowledge the support of NIH (R01048496) and the WRCE (1U54AI057156-0100) for the construction of *B. melitensis* mutant banks and the USDA (99-35204-7550) for the development of *B. abortus* mutant banks. We gratefully acknowledge the contributions of Chris Allen, Renée Tsolis, Priscilla Hong, and Carol Turse without whose efforts this publication would not be possible.

References

1. Van Bressem MF, Van Waerebeek K, Raga JA, Godfroid J, Brew SD, MacMillan AP (2001) Serological evidence of *Brucella* species infection in odontocetes from the south Pacific and the Mediterranean. Vet Rec 148:657–661
2. Retamal P, Blank O, Abalos P, Torres D (2000) Detection of anti-*Brucella* antibodies in pinnipeds from the Antarctic territory. Vet Rec 146:166–167
3. Ohishi K, Zenitani R, Bando T et al (2003) Pathological and serological evidence of *Brucella*-infection in baleen whales (Mysticeti) in the western North Pacific. Comp Immunol Microbiol Infect Dis 26:125–136
4. Seleem MN, Boyle SM, Sriranganathan N (2008) *Brucella*: a pathogen without classic virulence genes. Vet Microbiol 129:1–14
5. Goldstein J, Hoffman T, Frasch C et al (1992) Lipopolysaccharide (LPS) from *Brucella abortus* is less toxic than that from *Escherichia coli*, suggesting the possible use of *B. abortus* or LPS from *B. abortus* as a carrier in vaccines. Infect Immun 60:1385–1389
6. Pei J, Turse JE, Wu Q, Ficht TA (2006) *Brucella abortus* rough mutants induce macrophage oncosis that requires bacterial protein synthesis and direct interaction with the macrophage. Infect Immun 74:2667–2675
7. Foulongne V, Bourg G, Cazevieille C, Michaux-Charachon S, O'Callaghan D (2000) Identification of *Brucella suis* genes affecting intracellular survival in an in vitro human macrophage infection model by signature-tagged transposon mutagenesis. Infect Immun 68:1297–1303
8. Wu Q, Pei J, Turse C, Ficht TA (2006) Mariner mutagenesis of *Brucella melitensis* reveals genes with previously uncharacterized roles in virulence and survival. BMC Microbiol 6:102
9. Allen CA, Adams LG, Ficht TA (1998) Transposon-derived *Brucella abortus* rough mutants are attenuated and exhibit reduced intracellular survival. Infect Immun 66:1008–1016
10. Hong PC, Tsolis RM, Ficht TA (2000) Identification of genes required for chronic

persistence of *Brucella abortus* in mice. Infect Immun 68:4102–4107

11. Lai F, Schurig GG, Boyle SM (1990) Electroporation of a suicide plasmid bearing a transposon into *Brucella abortus*. Microb Pathog 9:363–368
12. Burkhardt S, Jimenez de Bagues MP, Liautard JP, Kohler S (2005) Analysis of the behavior of eryC mutants of *Brucella suis* attenuated in macrophages. Infect Immun 73:6782–6790
13. den Hartigh AB, Sun YH, Sondervan D et al (2004) Differential requirements for VirB1 and VirB2 during *Brucella abortus* infection. Infect Immun 72:5143–5149
14. Kahl-McDonagh MM, Ficht TA (2006) Evaluation of protection afforded by *Brucella abortus* and *Brucella melitensis* unmarked deletion mutants exhibiting different rates of clearance in BALB/c mice. Infect Immun 74:4048–4057
15. Ocampo-Sosa AA, Garcia-Lobo JM (2008) Demonstration of IS711 transposition in *Brucella ovis* and *Brucella pinnipedialis*. BMC Microbiol 8:17
16. Haine V, Dozot M, Dornand J, Letesson JJ, De Bolle X (2006) NnrA is required for full virulence and regulates several *Brucella melitensis* denitrification genes. J Bacteriol 188:1615–1619
17. Valderas MW, Alcantara RB, Baumgartner JE et al (2005) Role of HdeA in acid resistance and virulence in *Brucella abortus* 2308. Vet Microbiol 107:307–212
18. Elzer PH, Phillips RW, Kovach ME, Peterson KM, Roop RM II (1994) Characterization and genetic complementation of a *Brucella abortus* high-temperature-requirement A (htrA) deletion mutant. Infect Immun 62:4135–4139
19. Robertson GT, Reisenauer A, Wright R et al (2000) The *Brucella abortus* CcrM DNA methyltransferase is essential for viability, and its overexpression attenuates intracellular replication in murine macrophages. J Bacteriol 182:3482–3489
20. Pei J, Wu Q, Kahl-McDonagh M, Ficht TA (2008) Cytotoxicity in macrophages infected with rough *Brucella* mutants is type IV secretion system dependent. Infect Immun 76:30–37
21. Godfroid F, Taminiau B, Danese I et al (1998) Identification of the perosamine synthetase gene of *Brucella melitensis* 16M and involvement of lipopolysaccharide O side chain in *Brucella* survival in mice and in macrophages. Infect Immun 66:5485–5493
22. McQuiston JR, Schurig GG, Sriranganathan N, Boyle SM (1995) Transformation of *Brucella* species with suicide and broad host-range plasmids. Methods Mol Biol 47:143–8
23. Brown JS, Aufauvre-Brown A, Brown J, Jennings JM, Arst H Jr, Holden DW (2000) Signature-tagged and directed mutagenesis identify PABA synthetase as essential for *Aspergillus fumigatus* pathogenicity. Mol Microbiol 36:1371–1380
24. Hoang TT, Karkhoff-Schweizer RR, Kutchma AJ, Schweizer HP (1998) A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. Gene 212:77–86
25. Schweizer HP, Hoang TT (1995) An improved system for gene replacement and xylE fusion analysis in *Pseudomonas aeruginosa*. Gene 158:15–22
26. Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW (1995) Simultaneous identification of bacterial virulence genes by negative selection. Science 269:400–403
27. Murphy KC, Campellone KG, Poteete AR (2000) PCR-mediated gene replacement in *Escherichia coli*. Gene 246:321–330

Chapter 3

Random Mutagenesis Strategies for *Campylobacter* and *Helicobacter* Species

Duncan J.H. Gaskin and Arnoud H.M. van Vliet

Abstract

Campylobacter and *Helicobacter* species are important pathogens in man and animals. The study of their virulence and physiology has been difficult due to the lack of tractable genetic tools, since many of the techniques established in *Escherichia coli* and related species were found to be non-functional in *Campylobacter* and *Helicobacter* species. The advent of functional genomics techniques in the last decade has been accompanied by the development of genetic tools, which take advantage of specific features of *Campylobacter* and *Helicobacter*, like natural transformation. This has allowed for the construction of random mutant libraries based on in vitro transposition or ligated loops followed by natural transformation and recombination, thus circumventing selection against sequences when cloning or passaging libraries through *E. coli*. Uses of the techniques have been in the study of motility, gene expression, and gene essentiality. In this chapter, we discuss the approaches and techniques used for the construction of random mutant libraries in both *Campylobacter* and *Helicobacter*.

Key words: *Campylobacter*, *Helicobacter*, Natural transformation, Random mutagenesis, In vitro transposition, Homologous recombination, Antibiotic resistance cassettes

1. Introduction

Members of the genera *Campylobacter* and *Helicobacter* colonize the gastrointestinal tract of a broad range of mammals and birds, where they can be either commensal or pathogenic (1, 2). They are generally characterised by requiring microaerobic growth conditions, and have a cork-screw motility with unipolar or bipolar flagella which allows them to rapidly move through viscous environments like the mucus layer in the gastrointestinal tract. The genera *Campylobacter* and *Helicobacter* are phylogenetically relatively close, as they are both members of the epsilon subdivision of the Proteobacteria (3), and many of the genetic tools

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_3,

and tricks work in both genera without requiring significant adaptation; hence, we discuss them jointly in this chapter.

The development of tractable genetic tools in *Campylobacter* and *Helicobacter* has been facilitated by the availability of complete genome sequences. The first complete *Helicobacter pylori* genome sequence was published in 1997, whereas the first *Campylobacter jejuni* genome sequence was published in 2000 (4, 5). The other rapid developments in functional genomics and high-throughput analysis methods in the last 10 years have contributed significantly to increasing our knowledge about the biology of *Campylobacter* and *Helicobacter*. In this chapter, we discuss the strategies for the generation and screening of random mutant libraries in both *Campylobacter* and *Helicobacter*, and also discuss the current state of the art in technical developments. Since most of the research on *Campylobacter* has been focused on *C. jejuni*, we mostly discuss data on *C. jejuni*; likewise we focus on *H. pylori* when discussing *Helicobacter*, unless specifically indicated.

1.1. Targeted Gene Inactivation

Research on *Campylobacter* and *Helicobacter* has suffered from a lack of sophisticated genetic tools such as those available for model organisms like *Escherichia coli*. There are no phage-based mutagenesis or delivery systems for *Campylobacter* or *Helicobacter*, and the construction of unmarked mutations using sucrose sensitivity is still not straightforward and has only been reported in isolated cases (6). Targeted inactivation of genes has, however, been very successful in both *Campylobacter* and *Helicobacter*, and is based on double homologous recombination whereby the genomic intact copy of the target sequences is replaced by sequences or insertionally inactivated with an antibiotic resistance marker (7).

1.2. Transformation

One aspect that facilitates mutagenesis in *Campylobacter* and *Helicobacter* species is that they are naturally transformable, i.e. they have the ability to acquire foreign double-stranded DNA from the environment, and can incorporate it into their own genome (8). This ability is thought to contribute significantly to the genetic heterogeneity of *Campylobacter* and *Helicobacter*, and is thought to be mediated via several independent systems. In contrast to other naturally transformable bacterial species, such as streptococci, *Haemophilus*, and *Neisseria*, natural transformation in *Campylobacter* and *Helicobacter* is not dependent on growth phase, competence factor, or specific uptake sequences in the DNA (9).

1.3. Random Mutagenesis Approaches

When coupled to the availability of complete genome sequences, random mutagenesis is a very powerful approach that can allow the identification of genes involved in processes without prior

knowledge or hypothesis. It does require the availability of a suitable selection assay which mimics the condition or phenotype of interest to screen a library of mutants. The application of random mutagenesis to *Campylobacter* and *Helicobacter* species has been somewhat slow due to the absence of functional transposons in vivo, a problem that was solved in the relatively recent past by the development of in vitro transposon mutagenesis (see below). Before the development of in vitro transposon mutagenesis, alternative methods of (semi-)random mutagenesis had been employed with variable degrees of success.

Generation of (semi-)random mutant libraries in *Campylobacter* and *Helicobacter* relies on the insertion of antibiotic resistance marker genes into the genome of either species, using homologous recombination. This has been achieved by the generation of libraries of chromosomal DNA via traditional restriction enzyme sites using ligated chromosomal loops (10–12), insertion of plasmids into the genome using single homologous recombination (13–16) or *E. coli*-based shuttle transposon mutagenesis (17, 18). These libraries were mostly used to identify genes involved in easily screenable phenotypes such as motility, urease activity, or amino acid auxotrophy. Later, the use of in vitro transposon-based methods for constructing libraries were reported, based on the in vitro activity of different transposases (19–24). All of these techniques use the natural transformation capability of both *Campylobacter* and *Helicobacter* species for the introduction of mutated genes into the chromosome. Mutagenesis by single homologous recombination (plasmid insertion) (13, 15) is to date only possible in specific strains of *C. jejuni* (480), *C. coli* (UA585), and *H. pylori* (1061), and this significantly limits the applicability of this technique. In addition, both this technique and shuttle transposon mutagenesis rely on passaging through *E. coli*, and due to such limitations they have been superseded by in vitro transposon mutagenesis, and are not further discussed in this chapter.

1.4. Ligated Chromosomal Loops

This technique is based on the restriction enzyme digestion of purified chromosomal DNA to fragments of 1–10 kb, which are then ligated at low concentration to promote intramolecular ligation to form loops. Secondary restriction enzyme digestion is subsequently performed on these loops, and these fragments are then ligated with the antibiotic resistance cassette produced by PCR, hence resulting in DNA containing mostly adjacent homologous chromosomal DNA interrupted by an antibiotic resistance cassette lacking methylation. This mixture is then transformed by natural transformation (25). This technique has been used successfully in *C. jejuni* to identify the *pflA* gene resulting in paralysed flagella (12), in *H. pylori* to identify genes involved in IL-12 induction (26), and in *H. mustelae* to identify genes involved in

adhesion to epithelial cells (11). Disadvantages of this method are semi-randomness, as it is based on the presence of restriction enzyme sites, and the potential introduction of artefacts at the ligation steps.

1.5. In Vitro Transposon Mutagenesis

This approach relies on the in vivo and in vitro activity of different transposases (19–21). It requires a mini-transposon (i.e. the minimally required sequences for transposition without the ability to excise and reintegrate at other locations) modified, so that it carries an antibiotic selection marker functional in *Campylobacter* and *Helicobacter*. So far only the use of kanamycin resistance has been reported for *Campylobacter* and *Helicobacter*. Many of the described transposases (e.g. Himar) show little or no sequence bias in their insertion sites (27, 28). This results in a more random distribution of transposon insertions throughout the target DNA.

2. Materials

2.1. Preparation of Genomic DNA

1. Rich broth agar plate or Rich broth (Brucella, Mueller-Hinton, see supplier (e.g. Oxoid, Difco, Becton Dickinson) for information).
2. For cultivation of *Helicobacter* species: β-cyclodextrins (Sigma-Aldrich, Fluka, or other commercial supplier) or either Newborn or Foetal Calf Serum (Invitrogen or other supplier of sera for tissue culture).
3. Microaerobic incubator or jars: 10% CO_2, 5% O_2, and 85% N_2 (see Note 1).
4. Centrifuge with 50 ml tube capacity and microcentrifuge.
5. 2 ml screw-cap tubes.
6. TE buffer: 10 mM Tris-HCl pH 8.0, 1 mM EDTA.
7. 10% sodium dodecyl sulphate (SDS) in water (see Note 2).
8. 20 mg/ml proteinase K in water (stored in small single-use aliquots at –20°C).
9. 5 M NaCl in water.
10. CTAB/NaCl solution: 10% CTAB in 0.7 M NaCl; Dissolve 4.1 g NaCl in 80 ml water and slowly add 10 g CTAB (hexadecyltrimethyl ammonium bromide) while heating and stirring. If necessary, heat to 65°C to dissolve. Adjust final volume to 100 ml. Before use, preheat to 65°C.
11. 24:1 chloroform/isoamyl alcohol.
12. 25:24:1 phenol/chloroform/isoamyl alcohol.
13. Isopropanol.

14. 70% ethanol.
15. 10 mg/ml RNase in water.

2.2. Production and Purification of Himar1 C9 Transposase

1. IPTG stock solution: 1 M IPTG made in water, filter sterilised using a 0.22 µm filter, aliquoted and stored at −20°C.
2. Cell resuspension buffer: 20 mM Tris–HCl pH 7.6, 2 mM $MgCl_2$, and 1 mM DTT (see Note 3).
3. Lysozyme solution: 25 mg/ml lysozyme in water, store at −20°C.
4. Lysis buffer: 20 mM Tris–HCl pH 7.6, 4 mM EDTA, 200 mM NaCl, 1% (w/v) sodium deoxycholate, 1% (v/v) NP-40, 1 mM DTT, and 0.6 mM PMSF (see Note 4).
5. DNase I solution: 1 mg/ml DNase I (Sigma) in water. This should be prepared fresh before use.
6. Wash buffer: 1% (v/v) NP-40, 1 mM EDTA.
7. DEAE Sepharose (GE Healthcare).
8. Column buffer: 4 M Guanidine HCl, 20 mM Tris–HCl pH 7.6, 50 mM NaCl, 5 mM DTT, and 1 mM PMSF.
9. 5 ml Spectra/Por Float-A-Lyzer G2 8–10 kDa cut off (Spectrum Laboratories, Inc.).
10. Dialysis buffer: 50 mM Tris–HCl pH 7.6, 100 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT.

2.3. In Vitro Transposition Using Campylobacter Genomic DNA

1. 2× Transposition buffer: 25 mM HEPES pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 10% (v/v) glycerol, 250 µg/ml BSA, and 2 mM DTT.
2. 100 mM dATP, 100 mM dGTP, 100 mM dCTP, and 100 mM TTP stock solutions (Promega).
3. T4 DNA polymerase (Promega).
4. 10× ligase buffer (Promega) with ATP.
5. T4 DNA ligase (Promega).

2.4. Preparation of Ligated Loops

1. Restriction enzymes, T4 DNA ligase, and corresponding buffers dependent on supplier.
2. 25:24:1 phenol/chloroform/isoamyl alcohol.
3. 100% ethanol.
4. 70% (v/v) ethanol.
5. Plasmid containing antibiotic resistance cassette flanked by suitable recognition sites for restriction enzymes, for example pJMK30 (kanamycin cassette) and pAV35 (chloramphenicol cassette) (29).
6. Alternative for ethanol precipitation: commercial kits equivalent to the QiaQuick DNA purification kit (Qiagen).

2.5. Electrotransformation of Campylobacter jejuni

1. *Campylobacter* transformation buffer: 272 mM sucrose and 15% (v/v) glycerol.
2. Skirrow plates: 39 g Columbia Blood Agar Base (Oxoid) per litre, autoclave, cool, add 1 vial Campylobacter Selective Supplement (Oxoid) per 500 ml agar.
3. Electroporation device: example – Bio-Rad Gene Pulser II.
4. Electroporation cuvettes: example – Bio-Rad cuvettes with a 2 mm electrode gap.
5. Blood plates: 40 g Blood Agar Base (Oxoid) per litre, autoclave, cool, add 5% v/v defibrinated Horse Blood (Oxoid). Do not use plates which are more than a week old.
6. Blood plates containing kanamycin, final concentration 50 µg/ml. Do not use plates which are more than a week old.
7. Micro-aerobic incubators: 5% O_2, 10% CO_2, and 85% N_2 (see Note 1).
8. Sterile microcentrifuge tubes.
9. Brucella (Becton Dickinson).

2.6. Natural Transformation of Helicobacter pylori

1. Dent agar plates: Columbia agar supplemented with 7% saponin-lysed horse blood and Dent selective supplement (30). Use plates not more than a week old for the first incubation period.
2. Dent plates containing the selective antibiotics required for the second incubation period. Concentrations of antibiotics commonly used are 10 or 20 µg/ml kanamycin, 10 µg/ml chloramphenicol, or 5 µg/ml erythromycin, depending on the antibiotic resistance cassette used. Use plates not more than a week old.
3. Brucella broth (Difco or Oxoid or other major supplier).
4. Micro-aerobic incubators or jars: 5% O_2, 10% CO_2, and 85% N_2 (see Note 1).

3. Methods

3.1. Preparation of Genomic DNA of Campylobacter and Helicobacter

1. Preparation of genomic DNA from either species is often achieved by the use of commercial kits, like the DNeasy kit (Qiagen). However, genomic DNA of high quality can also be purified using a protocol based on CTAB/NaCl purification (31), and is given below for a small-scale culture.
2. Plate out 10–20 µl glycerol stock of bacteria on a rich agar plate. Incubate plate O/N at 37°C in microaerophilic incubator.

3. Re-streak a number of colonies on a rich agar plate. Incubate plate O/N at 37°C in microaerophilic incubator.
4. Remove colonies and use them to inoculate 100 ml of Brucella broth (for *Helicobacter* species add 3–5% foetal or newborn calf serum, or 0.2–1% β-cyclodextrins). Grow at 37°C in microaerophilic incubator on a shaker at 200 rpm for 15–18 h.
5. Pellet cells by centrifuging at 3,000 × *g* for 10 min. Discard supernatant and resuspend pellet in 1 ml PBS. Transfer cell suspension to a 2 ml screw-cap tube. Centrifuge tube for 3 min at 14,000 × *g*.
6. Discard supernatant and resuspend pellet in 400 µl TE. Add 70 µl 10% SDS and 5 µl 10 mg/ml proteinase K. Mix by inversion (do not vortex) and incubate at 65°C for 10 min.
7. Add 100 µl 5 M NaCl, mix and then add 100 µl 10% CTAB in 0.7 M NaCl (preheated to 65°C). Vortex until the suspension turns white. Incubate at 65°C for 10 min.
8. Add 500 µl chloroform/isoamyl alcohol (24:1). Vortex for at least 10 s and centrifuge for 5 min at 14,000 × *g*. Transfer upper aqueous phase to a fresh microcentrifuge tube and add 0.6 volumes of isopropanol. Mix and incubate at –20°C for at least 30 min.
9. Centrifuge for at least 10 min at 14,000 × *g*. Discard supernatant and wash pellet with 500 µl cold 70% ethanol. Centrifuge briefly, and then carefully remove remaining supernatant using a pipette.
10. Dry the pellet in air and resuspend in 50 µl TE supplemented with 1 µl 10 mg/ml RNase. Incubate at 37°C for 30 min. If necessary, let dissolve for longer at room temperature of 4°C.
11. Check DNA by running 2 µl sample on 1% agarose gel or measuring absorbance at 260 nm. Store DNA at –20°C.

3.2. Production and Purification of Himar1 C9 Transposase

1. The Himar1 C9 transposase expression plasmid pET29C9 was described by Lampe et al. (32). This plasmid carries the gene encoding a hyperactive mutant of the Himar1 transposase under the control of a T7 RNA polymerase promoter.
2. Streak the *E. coli* BL21(DE3) pET29C9 stock onto a LB agar plate supplemented with kanamycin to a final concentration of 50 µg/ml. Incubate overnight at 37°C.
3. Inoculate a single colony into 10 ml LB with 50 µg/ml kanamycin and grow overnight at 37°C with shaking.
4. Inoculate 1 l of LB supplemented with kanamycin to a final concentration of 50 µg/ml, in a 2 l baffled flask with 10 ml of the overnight culture. Grow at 37°C with shaking (approximately 200 rpm) until the OD_{600} reaches approximately 1.

Add 1 ml of IPTG stock solution and grow for 2 more hours with shaking at 37°C.

5. Take 1 ml of the culture for analysis by SDS–PAGE to confirm the presence of a band corresponding to the Himar transposase (approximately 35 kDa).
6. Harvest the cells by centrifugation at greater than 3,000 × *g* for 20 min at 4°C and discard the supernatant. Determine the cell pellet wet mass and resuspend in cell resuspension buffer to a concentration of approximately 0.5 g/ml. Divide into 1 ml aliquots in 2 ml centrifuge tubes and store at –80°C.
7. Thaw a 1 ml aliquot at room temperature. Add 25 µl of lysozyme solution, mix by inversion and incubate at room temperature for 5 min.
8. Add 1 ml lysis buffer, mix by inversion and incubate at room temperature for 15 min. The suspension should become viscous as cell lysis occurs.
9. Add 60 µl of DNase I solution and 20 µl of 1 M $MgCl_2$ solution, mix by gentle pipetting until the suspension loses its viscosity. Incubate for an additional 20 min at room temperature.
10. Centrifuge at greater than 14,000 × *g* for 5 min at 4°C and discard the supernatant.
11. Wash the pellet by resuspending it in 1 ml of wash buffer.
12. Centrifuge at greater than 14,000 × *g* for 5 min at 4°C and discard the supernatant.
13. Repeat the washing step.
14. Dissolve the pellet in 500 µl of column buffer and centrifuge at greater than 14,000 × *g* for 5 min.
15. Apply the supernatant to the surface of an 8 ml DEAE Sepharose column (see Note 5).
16. As soon as the sample has entered the column, add 1 ml of column buffer and allow it to run into the column.
17. Add a further 8 ml of column buffer to the top of the column and collect 0.5 ml fractions from the bottom of the column in 1.5 ml centrifuge tubes until the column stops running.
18. Analyse 10 µl of each fraction by SDS–PAGE to identify those containing the transposase.
19. Pool the fractions containing the majority of the transposase, usually eight fractions (4 ml).
20. Load the pooled fractions into a 5 ml Spectra/Por Float-A-Lyzer G2 and dialyse against 1 l of dialysis buffer for 2 h at 4°C. Replace with a fresh 1 l of dialysis buffer and dialyze overnight at 4°C.

21. Recover the dialysed protein solution from the DispoDialyzer into a suitable centrifuge tube and centrifuge at >5,000 × *g* for 15 min.
22. Remove the supernatant and mix with an equal volume of glycerol.
23. Dispense 50 µl aliquots into 0.2 ml PCR tubes and store at –80°C (see Notes 6 and 7).

3.3. In Vitro Transposition Using Campylobacter Genomic DNA

1. The in vitro transposition reaction results in random insertion of the mini-transposon into the target genomic DNA with single-stranded gaps and nicks at the insertion sites that need to be repaired prior to transformation.
2. Mix 2.5 µg of *C. jejuni* genomic DNA (see Note 8) with 2 µg of mini-transposon donor plasmid (see Note 9), 70 µl of 2× transposition buffer and water to 130 µl.
3. Add 10 µl of purified HimarC9 transposase (see Subheading 3.2, step 23). Incubate at 30°C for 4 h. Heat at 75°C for 15 min to inactivate the transposase.
4. Add 150 µl Phenol:Chloroform, mix by vortexing for 30 s.
5. Centrifuge at greater than 14,000 × *g* for 2 min. Carefully transfer the upper aqueous layer to a fresh tube. Add 150 µl chloroform, mix by vortexing for 30 s.
6. Centrifuge at greater than 14,000 × *g* for 2 min. Carefully transfer the upper aqueous layer to a fresh tube. Add 14 µl 3 M sodium acetate, mix and add 300 µl ethanol.
7. Incubate at –80°C for 10 min.
8. Centrifuge at greater than 14,000 × *g* for 10 min at 4°C and carefully remove the supernatant and discard.
9. Wash the pellet (which may be invisible) with 500 µl of 70% (v/v) ethanol.
10. Centrifuge at greater than 14,000 × *g* for 10 min at 4°C and carefully remove the supernatant and discard.
11. Dry the pellet.
12. Resuspend the pellet in 40 µl of 10 mM Tris–HCl, pH 7.4.
13. Add 6 µl Restriction Buffer 2 (New England Biolabs), 4.8 µl of 1.25 mM dNTPs (Promega), 8.7 µl water, and 0.5 µl (4.5 U) T4 DNA polymerase (Promega). Incubate at 11°C for 20 min followed by 10 min at 75°C to inactivate the T4 DNA polymerase.
14. Add 10 µl 10× ligase buffer (Promega), 47.5 µl water, and 0.5 µl (5 U) T4 DNA ligase (Promega). Incubate overnight at 16°C.
15. Prior to transformation, the DNA needs to be purified by ethanol precipitation.

3.4. Preparation of Ligated Loops of Genomic DNA

1. Digest 1–10 µg genomic DNA to completion by digestion with the first restriction enzyme to leave fragments of 0.5–10 kb (see Note 10 for a discussion of restriction enzyme choice).
2. Phenol/chloroform extract and ethanol precipitate the DNA sample.
3. Ligate the digested DNA in a large volume (greater than 200 µl), promoting intramolecular ligation. DNA concentrations used previously (10–12) range from 1 to 20 µg/ml with 1 U of T4 DNA ligase, at 15–18°C.
4. Phenol/chloroform extract and ethanol precipitate the DNA sample, or alternatively use commercial kits like the QiaQuick kit to remove the ligase, and concentrate the DNA to a volume of 10 µl.
5. Digest the ligated loops with a secondary restriction enzyme of choice (see step 1 for a discussion on restriction enzyme choice), followed by purification via phenol/chloroform extraction and ethanol precipitation or QiaQuick column.
6. Digest the pJMK30-derived antibiotic resistance cassette (29) with *Bam*HI or *Eco*RI to leave compatible ends for ligation to *Bam*HI/*Bgl*II/*Sau*3AI or *Apo*I-compatible ends (see Note 10), respectively, followed by purification via phenol/chloroform extraction and ethanol precipitation or QiaQuick kit.
7. Ligate digested loops from step 6 ("vector") and digested antibiotic resistance cassette from step 7 ("insert") using the standard 3:1 molecular ratio, using standard ligation protocols.
8. Ethanol precipitate the ligation mixture, resuspend in 10 µl water and transform to *C. jejuni* or *H. pylori* using electrotransformation (see Subheading 3.5) or natural transformation (see Subheading 3.6).

3.5. Electrotransformation of Campylobacter jejuni

1. Two days before electrotransformation of *Campylobacter*: Set up lawn of *C. jejuni* from a working stock on a Skirrow plate, incubate O/N at 37°C microaerobically (see Note 11 for alternatives to electrotransformation).
2. One day before: Resuspend the lawn in 2 ml of Brucella broth, pre-warmed to 37°C.
3. Prepare a lawn, using 200 µl of the suspension and inoculate onto a fresh Skirrow plate (see Notes 12 and 13).
4. Grow the plates overnight at 37°C under microaerobic conditions.
5. On the day of electrotransformation of Campylobacter, place the correct number of Blood plates (without antibiotic) into a 37°C incubator to pre-warm.

6. Harvest the bacteria from the Skirrow plate using 2 ml of Brucella broth and a spreader to scrape the cells from the surface of the plate. Transfer the suspension into a 2 ml microcentrifuge tube (see Note 14).
7. Pellet the bacteria by centrifugation for 3 min at 14,000 × *g* and discard the supernatant. Resuspend the pellet with 1 ml of transformation buffer.
8. Repeat step 7 twice. On the last resuspension use 0.5 ml of transformation buffer.
9. Aliquot 100 µl of the competent cells into a 1.5 ml centrifuge tube, add the DNA to be transformed in 10 µl of water, mix by gentle pipetting and transfer the cells plus DNA to a new electroporation cuvette (see Notes 15 and 16).
10. Transfer the cuvette to the electroporation apparatus and pulse at 2.5 kV, 200 Ω, and 25 µF for cuvettes with 2 mm electrode gap (see Note 17).
11. Remove the cuvette from the apparatus. Add 250 µl of Brucella broth to the cuvette and mix.
12. Transfer the cell mix from the cuvette onto the surface of a pre-warmed labelled Blood plate and spread the cells using a spreader.
13. Incubate plates for approximately 5 h at 37°C under microaerobic conditions (see Note 18).
14. Recove the cells from the plate using 800 µl of Brucella broth and a spreader.
15. Spread the suspension on the surface of a Blood plate containing kanamycin.
16. Incubate the Blood plate at 37°C under microaerobic conditions for 2–5 days until colonies are visible (see Note 19).

3.6. Natural Transformation of Helicobacter pylori

1. Inoculate *H. pylori* or other *Helicobacter* species onto a Dent plate, and grow for 24–48 h (see Note 20).
2. Harvest cells from the plate using a sterile inoculation loop, and transfer as thick patch of cells of approximately 1 cm diameter on a fresh plate. Several independent transformations can be performed on a single plate when sufficient distance is kept to avoid cross-contamination.
3. Incubate for 5–8 h at 37°C in microaerobic conditions.
4. Add approximately 1–5 µg of DNA in a maximum of 8–10 µl TE or water to the patch, and let dry for 2–3 min.
5. Incubate for 5–20 h at 37°C in microaerobic conditions.
6. Transfer cells with a cotton swab to a fresh plate with selective antibiotics, and spread cells over the surface of the plate.
7. Incubate at 37°C under microaerobic conditions for 3–7 days.

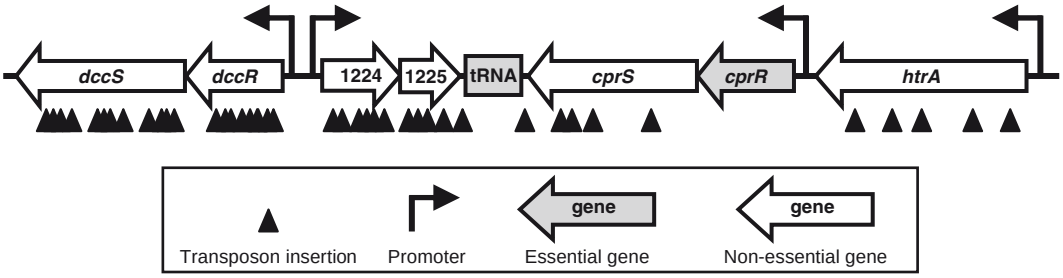


Fig. 1. Example of Genomic Analysis and Mapping by In vitro Transposition (GAMBIT) (33) screening of *C. jejuni* NCTC 11168 transposon libraries for essential genes. Transposon insertions in the region *cj1222c* (*dccS*) to *cj1228c* (*htrA*) are shown. From previous studies, it was known that the *dccS*, *dccR*, *cprS*, and *htrA* could be inactivated, whereas *cprR* is an essential gene (35–37). This is visible in the transposon insertions, which are found in all genes except *cprR.*

3.7. Screening of Libraries of Random Mutants

There are several ways in which libraries of random mutants can be used to provide valuable information about gene function. It is beyond the scope of this chapter to discuss detailed protocols as they will vary depending on selection criteria. However, in general there are two ways to use such libraries. Firstly, individual mutants can be screened for a particular phenotype, with those exhibiting or lacking it being further characterised and the location of the insert determined (see Notes 21 and 22). Secondly, techniques such as Genomic Analysis and Mapping by In vitro Transposition (GAMBIT) (33) and Signature Tagged Mutagenesis (STM) (34) can be used to map the location of all inserts in a genome to identify genes essential for survival under defined conditions. An example of GAMBIT used to map transposon insertions into the genome of *C. jejuni* NCTC 11168 is shown in Fig. 1. This study confirmed the essentiality of the *cprR* response regulator gene (*cj1227c*) (35), whereas other genes around it can be inactivated by transposon insertions (36, 37).

4. Notes

1. Several methods are available to create the required microaerobic conditions for growth of *Campylobacter* and *Helicobacter*. Commonly used are gaspacks (Oxoid and several other suppliers), the Anoxomat™ system (Mart Microbiology B.V) or microaerobic workstations (Don Whitley Scientific).
2. Unless stated otherwise, all solutions should be prepared in water that has a resistivity of 18.2 MΩ cm and total organic content of less than five parts per billion. This standard is referred to as "water" in this text.
3. DTT should be added fresh when making up solutions containing it. Either as powder or from frozen aliquots of 100 mM stock solution stored at –20°C.

4. PMSF should be added just before use from a 1 M stock solution in isopropanol stored at –20°C in aliquots.
5. The DEAE Sepharose column is prepared by loading 8 ml of DEAE Sepharose into a suitable column and equilibrating with 50 ml of column buffer (without the PMSF).
6. The typical protein concentration of the refolded transposase is approximately 100 µg/ml, and can be determined by any suitable protein determination method.
7. The transposase can be stored at –80°C for approximately 1 year without loss of activity.
8. We have observed the best transformation efficiencies with genomic DNA purified using Qiagen Genomic Tips.
9. Several suitable donor plasmids have been described (38).
10. The ligated loops strategy is difficult to contain in a single protocol. The choice of restriction enzymes strongly depends on the average size of the fragments obtained, methylation patterns of the DNA, and compatible overhangs of the antibiotic resistance cassette used. The combinations that were used in *C. jejuni* (12) are first digestion with *Cla*I, second digestion with *Bam*HI or *Bgl*II (to leave compatible ends for a *Bam*HI-digested kanamycin resistance cassette). For *H. pylori* (10) we used *Afl*II or *Ssp*I for the first digestion, and *Apo*I or *Sau*3AI for the second digestion (leaving compatible ends for an *Eco*RI or *Bam*HI-digested kanamycin resistance cassette, respectively). Finally, in *H. mustelae* (11) the enzyme combination *Cla*I/*Bgl*II was used, similar to *C. jejuni*.
11. *C. jejuni* is both naturally transformable and electrotransformable. In our experience the latter is more reproducible, and hence the first method of choice. Additional protocols for natural transformation are available (7).
12. Longer incubation results in a mixture of *Campylobacter* morphology (coccoids, straight and wiggly). It has been observed that cells taken from a 24 h plate are more competent as opposed to a 48 h plate.
13. Each plate will give enough cells for approximately five transformations.
14. It is important to avoid scraping bits of agar from the surface of the plate as these can lead to arcing in the electroporation cuvettes and reduced transformation efficiencies.
15. Do not reuse electroporation cuvettes.
16. It was observed that mixing the cells and DNA prior to transfer to the cuvette gave more reproducible transformation efficiencies than mixing in the cuvettes.
17. Settings for a BioRad Gene Pulser II are: 2.5 kV, 200 Ω, and 25 µF. This should give a time constant of less than 5 ms.

Sometimes the sample will arc when the cuvette is electroporated. This may happen because there is too much DNA, incorrect equipment settings or foreign bodies in the cuvettes. In extreme cases, this can result in the lid of the cuvette being blown off and the inside of the electroporation apparatus being contaminated. Experience has shown that even after such an event transformants can be obtained.

18. Plates can be left incubating overnight. However, it is not advisable to do this if trying to create a library, for this will create problems by producing an outgrowth of cells, potentially resulting in multiple copies of the same clones.
19. If individual colonies are to be screened, then they need to be picked and grown up to allow long-term storage. For pooled library applications, the colonies from transformation plates can be recovered using 2 ml of Brucella broth and a spreader. This pooled library can be stored for subsequent growth and analysis.
20. Although *H. pylori* can be electrotransformed, there is often a lower frequency of transformation than that observed by natural transformation. While the exact cause of this is unknown, there may be an effect of the more fragile nature of *H. pylori* when compared to *C. jejuni*. Natural transformation is highly efficient in *H. pylori*, and the protocol described works in several *Helicobacter* species (*H. pylori*, *H. mustelae*, and with some adaptation also in *H. hepaticus*).
21. Southern blotting of genomic DNA prepared from individual mutants and probed with labelled probes specific for the antibiotic resistance gene will confirm the randomness of the insertions and show whether there are any multiple insertion events.
22. Screening of libraries: commonly, libraries are first screened for easily determined phenotypes. With *Campylobacter* and *Helicobacter*, motility is often used by screening in motility agar (Brucella broth with 0.4% agar). An alternative screening for *Helicobacter* species is by measuring urease activity.

Acknowledgements

The *Campylobacter* research in the Institute of Food Research is supported by the Core Strategic Grant from the Biotechnology and Biological Sciences Research Council (BBSRC). We thank Dr Jerry Wells, Dr Kay Fowler, and other past and present members of our laboratories for their contributions to the development of these protocols.

References

1. Young KT, Davis LM, Dirita VJ (2007) *Campylobacter jejuni*: molecular biology and pathogenesis. Nat Rev Microbiol 5:665–679
2. Kusters JG, van Vliet AHM, Kuipers EJ (2006) Pathogenesis of *Helicobacter pylori* infection. Clin Microbiol Rev 19:449–490
3. Nakagawa S, Takaki Y, Shimamura S, Reysenbach AL, Takai K, Horikoshi K (2007) Deep-sea vent epsilon-proteobacterial genomes provide insights into emergence of pathogens. Proc Natl Acad Sci U S A 104:12146–12150
4. Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD et al (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature 388:539–547
5. Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D et al (2000) The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature 403:665–668
6. Copass M, Grandi G, Rappuoli R (1997) Introduction of unmarked mutations in the *Helicobacter pylori vacA* gene with a sucrose sensitivity marker. Infect Immun 65: 1949–1952
7. van Vliet AHM, Wood AC, Henderson J, Wooldridge KG, Ketley JM (1998) Genetic manipulation of enteric *Campylobacter* species. Methods Microbiol 27:407–419
8. Sheppard SK, McCarthy ND, Falush D, Maiden MC (2008) Convergence of *Campylobacter* species: implications for bacterial evolution. Science 320:237–239
9. Wang Y, Taylor DE (1990) Natural transformation in *Campylobacter* species. J Bacteriol 172:949–955
10. de Jonge R, Bakker D, van Vliet AHM, Kuipers EJ, Vandenbroucke-Grauls CM, Kusters JG (2003) Direct random insertion mutagenesis of *Helicobacter pylori*. J Microbiol Methods 52:93–100
11. Croinin TO, McCormack A, van Vliet AHM, Kusters JG, Bourke B (2007) Random mutagenesis to identify novel *Helicobacter mustelae* virulence factors. FEMS Immunol Med Microbiol 50:257–263
12. Yao R, Burr DH, Doig P, Trust TJ, Niu H, Guerry P (1994) Isolation of motile and non-motile insertional mutants of *Campylobacter jejuni*: the role of motility in adherence and invasion of eukaryotic cells. Mol Microbiol 14:883–893
13. Bijlsma JJ, Vandenbroucke-Grauls CM, Phadnis SH, Kusters JG (1999) Identification of virulence genes of *Helicobacter pylori* by random insertion mutagenesis. Infect Immun 67:2433–2440
14. de Vries N, Kuipers EJ, Kramer NE, van Vliet AHM, Bijlsma JJ, Kist M et al (2001) Identification of environmental stress-regulated genes in *Helicobacter pylori* by a *lacZ* reporter gene fusion system. Helicobacter 6:300–309
15. Dickinson JH, Grant KA, Park SF (1995) Targeted and random mutagenesis of the *Campylobacter coli* chromosome with integrational plasmid vectors. Curr Microbiol 31:92–96
16. Smeets LC, Bijlsma JJ, Boomkens SY, Vandenbroucke-Grauls CM, Kusters JG (2000) *comH*, a novel gene essential for natural transformation of *Helicobacter pylori*. J Bacteriol 182:3948–3954
17. Labigne A, Courcoux P, Tompkins L (1992) Cloning of *Campylobacter jejuni* genes required for leucine biosynthesis, and construction of leu-negative mutant of *C. jejuni* by shuttle transposon mutagenesis. Res Microbiol 143: 15–26
18. Haas R, Meyer TF, van Putten JP (1993) Aflagellated mutants of *Helicobacter pylori* generated by genetic transformation of naturally competent strains using transposon shuttle mutagenesis. Mol Microbiol 8:753–760
19. Colegio OR, Griffin TJ IV, Grindley ND, Galan JE (2001) In vitro transposition system for efficient generation of random mutants of *Campylobacter jejuni*. J Bacteriol 183: 2384–2388
20. Golden NJ, Camilli A, Acheson DW (2000) Random transposon mutagenesis of *Campylobacter jejuni*. Infect Immun 68: 5450–5453
21. Hendrixson DR, Akerley BJ, DiRita VJ (2001) Transposon mutagenesis of *Campylobacter jejuni* identifies a bipartite energy taxis system required for motility. Mol Microbiol 40: 214–224
22. Baldwin DN, Shepherd B, Kraemer P, Hall MK, Sycuro LK, Pinto-Santini DM, Salama NR (2007) Identification of *Helicobacter pylori* genes that contribute to stomach colonization. Infect Immun 75:1005–1016
23. Chang KC, Yeh YC, Lin TL, Wang JT (2001) Identification of genes associated with natural competence in *Helicobacter pylori* by transposon shuttle random mutagenesis. Biochem Biophys Res Commun 288:961–968

24. Salama NR, Shepherd B, Falkow S (2004) Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*. J Bacteriol 186:7926–7935
25. Sharetzsky C, Edlind TD, LiPuma JJ, Stull TL (1991) A novel approach to insertional mutagenesis of *Haemophilus influenzae*. J Bacteriol 173:1561–1564
26. de Jonge R, Kuipers EJ, Langeveld SC, Loffeld RJ, Stoof J, van Vliet AH, Kusters JG (2004) The *Helicobacter pylori* plasticity region locus *jhp0947-jhp0949* is associated with duodenal ulcer disease and interleukin-12 production in monocyte cells. FEMS Immunol Med Microbiol 41:161–167
27. Griffin TJ IV, Parsons L, Leschziner AE, DeVost J, Derbyshire KM, Grindley ND (1999) In vitro transposition of Tn552: a tool for DNA sequencing and mutagenesis. Nucleic Acids Res 27:3859–3865
28. Lampe DJ, Churchill ME, Robertson HM (1996) A purified mariner transposase is sufficient to mediate transposition in vitro. EMBO J 15:5470–5479
29. van Vliet AHM, Wooldridge KG, Ketley JM (1998) Iron-responsive gene regulation in a *Campylobacter jejuni fur* mutant. J Bacteriol 180:5291–5298
30. van Vliet AHM, Kuipers EJ, Waidner B, Davies BJ, de Vries N, Penn CW et al (2001) Nickel-responsive induction of urease expression in *Helicobacter pylori* is mediated at the transcriptional level. Infect Immun 69:4891–4897
31. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (2001) Current protocols in molecular biology. Wiley, Malden, MA
32. Lampe DJ, Akerley BJ, Rubin EJ, Mekalanos JJ, Robertson HM (1999) Hyperactive transposase mutants of the Himar1 mariner transposon. Proc Natl Acad Sci U S A 96:11428–11433
33. Akerley BJ, Rubin EJ, Camilli A, Lampe DJ, Robertson HM, Mekalanos JJ (1998) Systematic identification of essential genes by *in vitro* mariner mutagenesis. Proc Natl Acad Sci U S A 95:8927–8932
34. Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, Holden DW (1995) Simultaneous identification of bacterial virulence genes by negative selection. Science 269:400–403
35. Svensson SL, Davis LM, MacKichan JK, Allan BJ, Pajaniappan M, Thompson SA, Gaynor EC (2009) The CprS sensor kinase of the zoonotic pathogen *Campylobacter jejuni* influences biofilm formation and is required for optimal chick colonization. Mol Microbiol 71:253–272
36. Brondsted L, Andersen MT, Parker M, Jorgensen K, Ingmer H (2005) The HtrA protease of *Campylobacter jejuni* is required for heat and oxygen tolerance and for optimal interaction with human epithelial cells. Appl Environ Microbiol 71:3205–3212
37. MacKichan JK, Gaynor EC, Chang C, Cawthraw S, Newell DG, Miller JF, Falkow S (2004) The *Campylobacter jejuni dccRS* two-component system is required for optimal in vivo colonization but is dispensable for in vitro growth. Mol Microbiol 54:1269–1286
38. Hendrixson DR, DiRita VJ (2004) Identification of *Campylobacter jejuni* genes involved in commensal colonization of the chick gastrointestinal tract. Mol Microbiol 52:471–484

Chapter 4

Mutagenesis of the Repeat Regions of Herpesviruses Cloned as Bacterial Artificial Chromosomes

Yuguang Zhao and Venugopal Nair

Abstract

Cloning of infectious and pathogenic herpesvirus genomes in a bacterial artificial chromosome (BAC) vector greatly facilitates genetic manipulation of their genomes. BAC-based mutagenesis strategies of viruses can advance our understanding of the viral gene functions and determinants of pathogenicity, and can ultimately help to develop molecularly defined improved vaccines against virus diseases. Unlike the virus stocks, where continuous passage in tissue culture can lead to phenotypic alterations such as loss of virulence or immunogenicity, viral genomes can be stably maintained with high fidelity as BAC clones in bacteria. Thanks to the "RecA" or the inducible phage "λ Red" homologous recombination systems and a variety of positive and negative selection strategies, viral genomes cloned as BAC can be efficiently manipulated in *E. coli*. All the manipulations, including DNA fragment deletion or insertion, point mutations, or even multiple modifications in repeat regions can be carried out accurately in *E. coli*, and the mutated DNA can be used directly to reconstitute mutant viruses in transfected host cells. Furthermore, using self-excision strategies, the non-viral bacterial replicon sequence can be excised automatically during virus reconstitution, thus generating recombinant viruses virtually identical to the wild-type parent viruses. Here, we describe the various technologies of manipulating the infectious BAC clones of a group E herpesvirus as an example through a combination of different approaches.

Key words: Herpesviruses, Bacterial artificial chromosome, Mutagenesis, Repeat region, λ Red recombination

1. Introduction

Family *Herpesviridae* consists of over 100 distinct viruses that are important pathogens for both animals and humans. All of those herpesviruses are large, enveloped viruses, consisting of a double-stranded DNA genome, ranging from 100 to 300 kb in size. At least six classes of herpesvirus genome types termed group A to F that vary according to the existence and locations of reiterations of terminal sequences are known. Although complete sequences

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_4, © Springer Science+Business Media, LLC 2010

of the genomes of many herpesviruses have been determined, the precise functions for many of the extensive repertoires of viral genes remain to be elucidated. One of the essential ways of getting insights into viral functional genomics is to generate mutant viruses that carry mutations in one or more of the genes. The conventional strategy for mutagenesis of herpesviruses employs eukaryotic cell homologous recombination pathways. However, most of the eukaryotic cells do not show high efficiency of homologous recombination, and there are no controlled/inducible systems to promote recombination in eukaryotic cells, making the application of the conventional methods of recombination very time-consuming and labour-intensive. Additionally, if the mutated genes are essential for *in vitro* viral replication, complementing cell lines are required beforehand. The establishment of such complementing cell lines are often difficult, particularly if primary cells are essential to support viral replication. Moreover, continued herpesvirus replication in cultured eukaryotic cells often leads to viral attenuation.

The recent developments in the technologies to clone the full-length genomes of herpesviruses as bacterial artificial chromosomes (BAC) have revolutionised the field of reverse genetics and genome manipulation of herpesviruses. Viral genomes are stably maintained in the bacteria, where controlled inducible high-efficiency homologous recombination events are carried out. The construction of BAC cloned viral mutants can be precise, fast, and independent of the biological fitness of either the parent virus or the recombinant progeny. This allows the mutations to be introduced anywhere in the genome for any number of genes irrespective of the essential nature or the requirements of these genes for the virus. With the ability for the excision of the extra bacterial replicon sequence after the reconstitution of the recombinant/mutant viruses in eukaryotic cells (1–3), viruses virtually identical to the wild-type viruses can be generated. Thus, this approach combines the highly efficient prokaryotic viral genome manipulation system with all the advantages of the eukaryotic cells that allows natural viral replication. Genomes of several herpesviruses belonging to the three subclasses have now been cloned as BAC. These include herpes simplex virus-1 (HSV-1) (4), varicella-zoster virus (VZV) (5), equine herpesvirus type 1(EHV-1) (6), pseudorabiesvirus (PRV) (1), Marek's disease virus (MDV) (7), herpes virus turkey (HVT) (8), koi herpesvirus (KHV) (9), human cytomegalovirus (HCMV) (2), Epstein-Barr virus (EBV) (10), Kaposi's sarcoma virus (KSHV) (11), and Rhesus rhadinovirus (RRV) (12).

In bacteria, two main homologous recombination promoting schemes using different recombinases are employed for BAC mutagenesis. One is based on bacteriophage λ Red and the other is based on bacterial RecA. The λ Red recombination system

utilises temperature inducible expression of the three λ Red encoded genes: Red α (Exo), Red β (Bet), and Red γ (gam). The earlier version of this, called ET recombination, was based on RecE (Red α equivalent) and RecT (Red β equivalent) of the RAC prophage. The new λ Red system is outperforming the ET recombination system as it is more efficient, and the additional Red γ (gam) encodes an inhibitor of the *Escherichia coli* RecBCD exonuclease, thereby protecting the linear DNA targeting cassette from degradation. This allows even the use of short homology arms (about 30–50 nucleotides) to mediate homologous recombination. The short homology arms can be easily incorporated within oligonucleotides, and recombination can be easily achieved using PCR-amplified selectable markers, with no need for cloning into a targeting vector. Furthermore, the λ red genes can be expressed from a stably integrated defective λ phage under tight control of a temperature-sensitive strong phage promoter (13, 14). This temporal but efficient induction of homologous recombination system ensures a good balance between the highly efficient homologous recombination and high levels of viral genome stability in the host bacteria.

RecA is a bacterial homologous recombination promoting enzyme. For maintaining stability of the BAC cloned viral genome in bacteria, the bacterial strains used were usually RecA negative. However, temporal expression of RecA can be employed for improving homologous recombination without losing the viral genome stability. The RecA system uses a temperature-sensitive shuttle vector containing a RecA gene, a positive selection marker, a negative selection marker, and a stretch of homologous sequences. Recombination is achieved in two steps: first, the shuttle vector is transfected into the bacteria harbouring the BAC DNA and following the positive selection and temperature shift, co-integrated forms are selected; and second, the resolution of the co-integrate is carried out through a negative selection marker, followed by the identification of the desired modification (15). This system requires a significant homologous region, and thus a cloning step cannot be avoided. Even if homologous sequences of 2 or 3 kb are provided, the resolution of the "co-integrates" to a desired modification occurs only with a low frequency, especially when the modified region is significantly large or has multiple changes. However, when the introduced mutations are small (for example involving only a few nucleotides), or when the mutation can introduce new identifiable markers such as a new restriction site, this strategy can be efficiently applied for BAC mutagenesis. Both λ Red- and RecA-based systems can be used to insert, delete, or introduce point mutations into a target gene of a viral genome in the BAC. As the λ Red-based inducible system is more efficient and precise, it is preferable to the RecA-based system. However, the RecA-based system can be further used within the λ Red-based

bacterial strain. This so-called "hybrid" system can further increase homologous recombination (16).

Homologous recombination requires the BAC containing bacteria to receive a piece of DNA of homologous sequence, which directs the recombination events. As the recombination events are rare, selection markers are used to select the recombinant clones. The vector in most of the herpesvirus BAC clones carries the chloramphenicol-resistance gene as a primary selection marker. Other powerful positive selection markers, such as Kanamycin (Kan), Spectinomycin (Spc), Ampicillin (Amp), or Tetracycline (Tet), can also be used for BAC manipulation (see Note 1). Any of these positive selection marker cassettes can be amplified by PCR using oligonucleotides containing short (~50 nt) homologous sequences to target genes in their 5′ end, and the PCR products can be directly introduced into λ Red-based bacteria. Recombinant clones selected with appropriate antibiotics can be confirmed by PCR using specific primers. The selection markers in the recombinant BAC DNA can be excised using inducible *Flpe* or *Cre* recombinases expressed in different bacterial strains (e.g. SW105 and SW106, respectively) if FRT or LoxP sites are engineered into the constructs flanking the selection markers. After the excision, only a short FRT or LoxP site is left behind as "scar" or "seam" sequences. The persistence of at least one copy of the FRT or LoxP sites limits the repeatability of the procedure and may interfere with gene expression in some circumstances. To construct a "scar-less" or "seam-less" mutation, such as introducing a single amino acid mutation in a coding region, counter-selectable negative selection markers have to be used.

Different negative selection markers are used to counterselect for *E. coli* recombinants. *Bacillus subtilis sacB* gene coding for levansucrase, the presence of which confers bacterial toxicity to sucrose (17) is widely used as a negative selection marker. It can be used in most of the gram-negative bacteria directly. However, as there is no means to positively ensure that the sacB gene is intact, spontaneous point mutations in the *sacB* gene can sometimes give rise to high background. Two other selection markers, *Gal*K and *Thy*A, can be used both for positive and negative selection (18, 19). The *gal*K gene product, galactokinase, phosphorylates galactose and converts it to galactose-1-phosphate. In *gal*K-deficient bacteria such as SW102, when galactose was supplied as the only carbon source in the minimal medium, only those bacteria carrying the *gal*K can survive, thus acting as a positive selection marker. On the other hand, galactokinase can also efficiently catalyse the phosphorylation of a galactose analogue, 2-deoxy-galactose (DOG) to a toxic product 2-deoxy-galactose-1-phosphate. In a DOG-containing minimal medium, only those bacterial strains which have lost the *gal*K gene can survive, making it function as a negative selection marker. Similarly, a thymidylate

synthase A (*Thy*A) gene can be used for both positive and negative selections in a *Thy*A-deficient bacterial strain such as QW1. *Thy*A is involved in the *de novo* synthesis of TTP from dUMP. The *Thy*A null-mutant bacteria can only grow in minimal medium when the exogenous *Thy*A gene is introduced, or in a medium containing thymine. On the other hand, *Thy*A also converts tetrahydrofolate (THF) to dihydrofolate (DHF). THF is essential for bacterial growth when trimethoprim is used to block the pathway to replenish THF from DHF. When bacteria lack *Thy*A, THF is maintained, and TTP can be synthesised from thymine allowing the bacteria to grow in a media containing thymine and trimethoprim. Among the *galK* and *Thy*A selection systems, the *galK* system is more convenient in our hands. In this chapter, only the *galK* selection details are included.

In addition to these negative selection markers, the rare cutting restriction site, 18-bp recognition I-*Sce*I can also be used for negative selection (20). When a I-*Sce*I site is engineered into a BAC genome, it will not be able to replicate if the bacteria also contain an I-*Sce*I enzyme expression plasmid, which will continuously digest the BAC genome leaving it linear and unable to circularise. This strategy can be further enhanced by duplicating a short homologous sequence next to the I-*Sce*I site to create a double-strand break and homologous recombination substrate at the same time (21). In this modified strategy, a positive selection marker that was used to introduce the desired target modification in a first step is removed in the second step by the combination of I-*Sce*I cleavage and intra-molecular Red recombination through previously introduced sequence duplication. This strategy needs the use of either very long synthetic oligonucleotides (see Subheading 3.2), or a cloning step to introduce a positive selection marker with the I-*Sce*I site and a duplicated homologous region (see Note 2).

2. Materials

2.1. Bacterial Strains

Different recombination/engineering strategies, especially when using different selection markers, may require using different bacterial strains. However, in the majority of cases, one could use SW102 or its related series, SW105 or SW106 (19) strains. For experiments using *Thy*A selection, QW1 strain (18) is required. All of these stains have been derived from DY380, a modification of DH10B, with the λRed genes under tight control of temperature sensitive λ repressor (cI857). SW102 or its related series have the further *galK* mutation, and QW1 has the *Thy*A mutation. SW105 has an additional *Flpe* gene under the control of *AraC* and P_{BAD} which is used for flipping off a selection marker flanked by FRT

sites once the bacteria are induced by arabinose. SW106 has an additional *Cre* gene under the control of *AraC* and P_{BAD}, for easy "*flox*ing out" of any sequence between two *Lox*P sites in the same orientation. SW102, SW105, and SW106 can all be grown in normal Luria–Bertani (LB) medium.

2.2. Bacterial Media

1. *LB medium* (For 1 L): 10 g Tryptone, 5 g Yeast extract, and 5 g NaCl, suspend in ddH_2O up to 1 L and autoclave for use.
2. *LB plates* (For 1 L): 10 g Tryptone, 5 g Yeast extract, 10 g NaCl, and 15 g Bacto-agar, suspend in ddH_2O up to 1 L, autoclave, cool to 55°C, add appropriate antibiotics and pour into sterile Petri dishes.
3. *M9 stock solution* (1 L 5×): 30 g Na_2HPO_4, 15 g KH_2PO_4, 5 g NH_4Cl, 2.5 g NaCl, suspend in ddH_2O up to 1 L. Autoclave before use.
4. *M63 stock solution* (5× 1 L): 10 g $(NH_4)2SO_4$, 68 g KH_2PO_4, and 2.5 mg $FeSO_4{\cdot}7H_2O$, dissolve in 800 ml of water and adjusted to pH 7 with KOH; final volume is 1 L. Autoclave before use.
5. *Other stock solutions*: 0.2 mg/mL d-biotin (sterile filtered), 20% galactose (autoclaved), 20% DOG (autoclaved), 20% glycerol (autoclaved), 10 mg/mL L-leucine (dissolve by addition of HCl), 10 mg/mL L-isoleucine (dissolve by addition of HCl), 10 mg/mL L-valine, 1 M $MgSO_4{\cdot}7H_2O$, 500× MM1 0.5 mM each of $ZnSO_4$, $CuSO_4$, Na_2MoO_4, $CoCl_2$, $MnSo_4$, $CrCl_3$, $NiCl_2$ (see Note 3), and 25 mg/mL chloramphenicol in ethanol (1:1,000).
6. *Gal*K *positive selection plates*: Autoclave 7.5 g agar in 400 mL H_2O in a 1 L bottle and let cool down a little. Add 100 mL autoclaved 5× M63 medium and 0.5 mL 1 M $MgSO_4{\cdot}7H_2O$. Adjust volume to about 500 mL with H_2O if necessary. Put the bottle in a 50°C water bath. Add 5 mL of 20% galactose (final conc. 0.2%), 2.5 mL of 0.2 mg/mL biotin, 2 mL of each 10 mg/mL L-leucine, 10 mg/mL L-isoleucine, 10 mg/mL L-valine, 1 mL of 500× MM1, and 500 µL of 25 mg/mL chloramphenicol (If the plates are used for the second copy, and the first copy is already replaced with a kanamycin selection cassette, add additional kanamycin at 50 µg/mL). Pour into sterile petri dishes.
7. *Gal*K *negative selection plates*: Autoclave 7.5 g agar in 400 mL H_2O in a 1 L bottle and let it cool down a little. Add 100 mL autoclaved 5× M63 medium and 0.5 mL 1 M $MgSO_4{\cdot}7H_2O$. Adjust the volume to about 500 mL with H_2O if necessary. Place the bottle in a 50°C water bath. Add 5 mL 20% DOG (autoclaved) (1:100), and 5 mL 20% glycerol (autoclaved) (1:100), 2.5 mL biotin, 2 mL of each 10 mg/mL L-leucine,

10 mg/mL L-isoleucine, 10 mg/mL L-valine, 1 mL 500× MM1, and 500 µL chloramphenicol. Pour the plates into sterile petri dishes.

8. *MacConkey indicator plates*: Prepare MacConkey agar (Oxoid, CM0007) according to manufacturer's instructions. Suspend 26 g in 500 mL ddH_2O. After autoclaving and cooling to 50°C, add 5 mL 20% galactose and 500 µL chloramphenicol, and pour the plates into sterile petri dishes.

2.3. Plasmids

1. p*gal*K plasmid (19) (see Note 4).
2. pBAD-I-*Sce*I and pEPKan-S (21) (see Note 4).

2.4. DNA Preparation Kits

1. QIAprep® Spin miniprep kit, Cat. No.27106.
2. QIAGEN® Plasmid Maxi kit, Cat. No. 12163.
3. MinElute™ Gel Extraction kit, Cat. No. 28604.

3. Methods

3.1. Strategies for Modification of Two Copies of Genes in the Repeat Regions

Modifying a unique gene region of the virus in the BAC clone is straightforward and less complicated. However, in the majority of herpesviruses, there are several important genes in the repeat regions in their genome. The group E herpesviruses, for example, have sequences from both termini repeated in an inverted orientation and juxtaposed internally, dividing the genomes into two components, each flanked by inverted repeats, as illustrated in Fig. 1. Herpes simplex virus (HSV) and Marek's Diseases virus (MDV) belong to this group. The procedures for modifying the

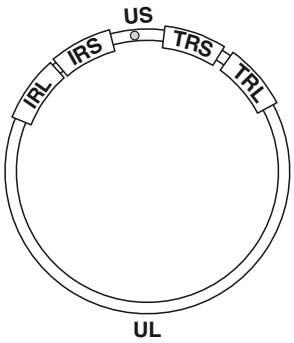


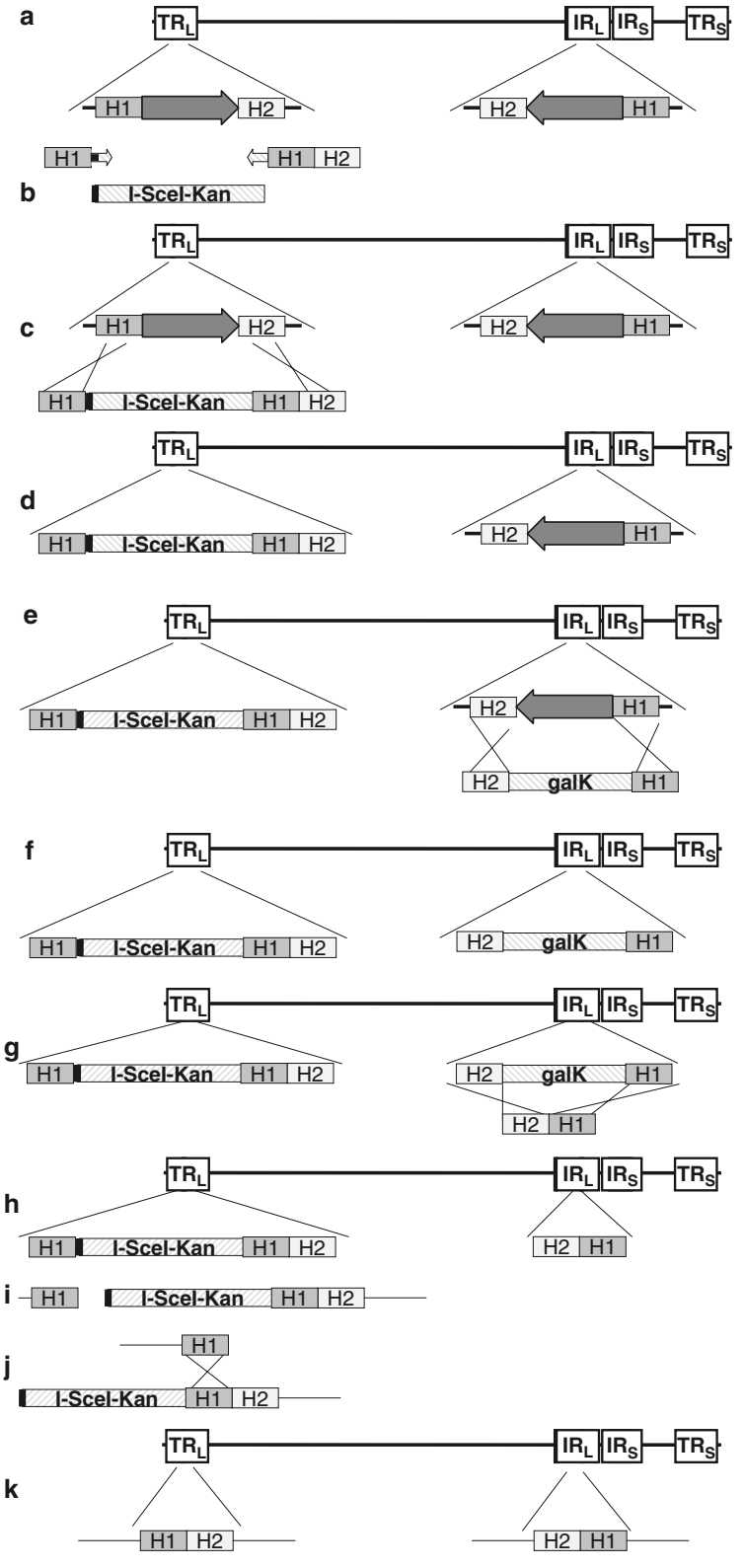
Fig. 1. Schematic illustration showing the BAC clone of group E herpesvirus genome. The circular form of double-stranded DNA viral genome contains unique long (UL) and unique short (US) as well as terminal repeat regions. Terminal (TRL) and internal (IRL) repeats from the unique long region, as well as the terminal (TRS) and internal (IRS) repeats from the unique short region are indicated. The *small circle* within the US region represents the bacterial mini-F replicon.

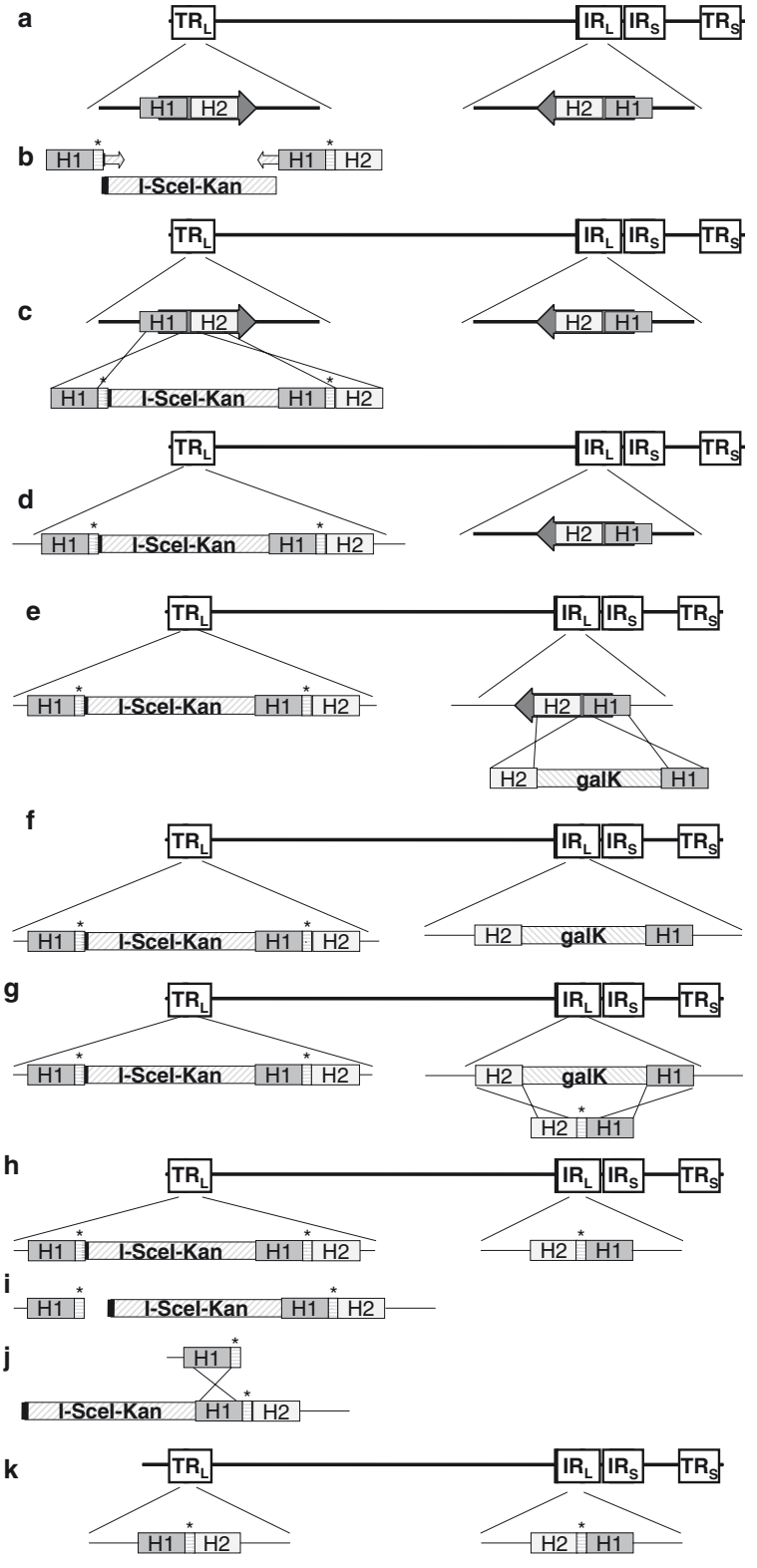
first copy of a gene in the repeat region are the same for modifying a unique gene in the BAC clone. The procedure described here is aimed to achieve "seamless" modification of two copies of genes in the repeat region through combined *galK* and I-*Sce*I-kanamycin selection (see Note 5). However, any single selection procedure can be applied to a unique gene modification.

Deletion of a viral gene is the most frequently used strategy to understand the viral gene function. Assuming the whole viral genome is cloned in a mini-F-based BAC amplicon with chloramphenicol resistance as the primary selection in the bacterial strain SW102, then two copies of seamless deletion can be achieved by the following steps: *First*, an I-*Sce*I-Kan selection cassette with a 50-bp duplication of homologous sequence is introduced to replace one copy of the gene through λ Red recombination using Kanamycin selection. *Second*, a *galK* selection cassette is introduced into the other copy by λ Red recombination with *galK* positive selection, while keeping the kanamycin selection pressure. *Third*, the desired deletion is introduced using long oligonucleotides or cloned sequences as the homologous arms to replace the *galK* gene by *galK* negative selection. *Finally*, pBAD-I-*Sce*I inducible expression plasmid is introduced into the bacterial strain to induce homologous recombination of the I-*Sce*I-Kan copy through the 50-bp duplication sequence, again using λ Red recombination. The strategy and the individual steps are illustrated in Fig. 2. For the introduction of point mutations, the same approach for the deletion of the genes as described above can be carried out by designing the point mutation in the 50-bp duplicated sequence, as well as in the oligonucleotides or cloned mutated sequence of the homologous arms to replace the *galK*. The individual steps and the strategy are illustrated in Fig. 3.

In some applications for BAC mutagenesis, it may be required to introduce foreign genes into the BAC clones of the viral genomes. This could include the introduction of other viral genes into

Fig. 2. Schematic representation of principle steps for seamless deletion of two copies of viral genes in the TRL/IRL regions. (**a**) Gene of interest to be deleted (*arrow*) shown in an inverted orientation within the repeat regions of the viral genome. H1 and H2 represent two homologous regions used for designing PCR primers and/or targeting vectors. (**b**) PCR primers designed to amplify I-*Sce*I-Kan template. (**c**) Induction of homologous recombination after the electroporation of the I-*Sce*I-Kan-homology PCR products into SW102 cells carrying the BAC clone. (**d**) The expected structure of the genome after the first I-*Sce*I-Kan recombination. (**e**) Second copy recombination through the *galK* PCR product while maintaining the Kan selection pressure results in the replacement modification shown in (**f**) with galK replacing the gene of interest by positive selection. (**g**) With a gene deleted homologous sequence, either a cloned sequence or annealed oligonucleotides are introduced into the bacteria to induce homologous recombination and *galK* negative selection, it will result in a galK deleted copy as shown in (**h**). When the I-*Sce*I restriction enzyme expression plasmid pBAD-I-*Sce*I is introduced into the bacteria harbouring the above modified BAC, a double-strand break at I-*SceI* site is created as illustrated in (**i**). This double-strand break will encourage homologous recombination between the duplicated homologous regions originally flanking the I-*Sce*I-Kan cassette as in (**j**). This homologous recombination finally results in seamless deletion in both copies as shown in (**k**).


a
TRL
IRL
IRS
TRS
H1
H2
b
I-SceI-Kan
c
d
e
galK
f
g
h
i
j
k


a
TRL
IRL
IRS
TRS
H1
H2
b
I-SceI-Kan
c
d
e
galK
f
g
h
i
j
k

herpesvirus vectors to make recombinant vaccines, addition of N-terminal or C-terminal tags to viral genes, generating fluorescent proteins, tandem affinity purification tag (TAP-tag), or an inducible gene expression system. These types of modifications can be carried out in four steps as described above. First step would require an additional cloning procedure to insert the I-*Sce*I-Kanamycin cassette in the middle of the gene of interest, with 50-bp duplication sequence surrounding the I-*Sce*I-kan cassette. The whole cassette is then amplified by PCR with primers containing 50-bp homology sequence of the target. The PCR products are then used to introduce insertion of the gene-of-interest together with I-*Sce*I-Kan. The second step is to introduce *galK* into the other copy by *galK* positive selection together with the Kanamycin selection pressure on. In the third step, the insert with homologous arms is used to replace the *galK* cassette by *galK* negative selection. In the final step, the I-*Sce*I expression plasmid (pBAD-I-*Sce*I) is introduced to induce homologous recombination through duplicated sequence surrounding the I-*Sce*I-Kanamycin cassette as demonstrated in Fig. 4.

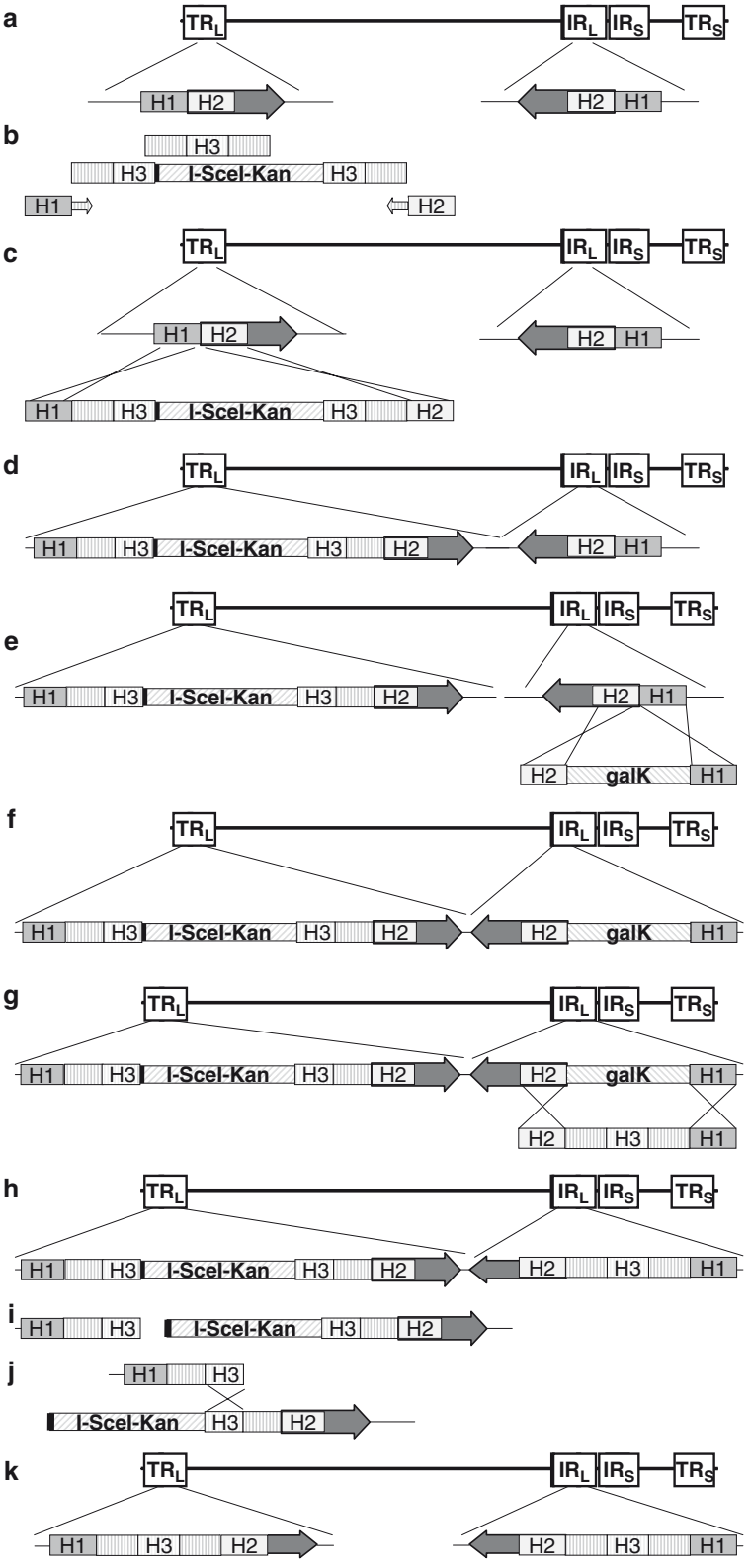
3.2. Design of Oligonucleotides

3.2.1. Oligonucleotides for I-SceI-Kan Selection (see Note 6)

I-*Sce*I-*Kan* cassette used in the experiments is derived from the plasmid pEPkan-S containing a Kanamycin selection cassette and an adjoining I-*Sce*I site.

For deletion type of modification, the transfer constructs can be generated by PCR amplification of the I-*Sce*I-Kan cassette with a forward primer containing approximately 50 bp homologous sequence before the gene to be deleted followed by sequence to amplify I-*Sce*I-Kan. The Reverse primer contains 40–50 bp of homologous sequence after the gene to be deleted followed by 50 bp of homologous sequence before the gene to be deleted and then the sequence to amplify I-*Sce*I-Kan.

Fig. 3. Schematic representation of principle steps for seamless point mutation of the two copies of viral genes in terminal repeats region. (**a**) Gene of interest to be mutated (*arrow*) in an inverted orientation within the viral repeats region. H1 and H2 represent two homologous regions used for designing PCR primers and/or targeting vectors. The homologous regions are chosen next to the sites where mutations are to be introduced. (**b**) PCR primers designed to amplify I-*Sce*I-Kan template. *Asterisks* represent the mutations introduced in the sequence. (**c**) Induction of homologous recombination after the electroporation of the I-*Sce*I-Kan-homology PCR products into SW102 cells carrying the BAC clone. (**d**) The expected structure of the genome after the first I-*Sce*I-Kan recombination (**e**) Second copy recombination through the *galK* PCR product while maintaining the Kan selection pressure result in the modification shown in (**f**) with *galK* inserted into the position where point mutation is to be introduced. (**g**) With a gene mutation and flanking homologous sequence, either a cloned sequence or annealed oligos are introduced into bacteria with the help of induced homologous recombination and *galK* negative selection, the result will be a mutated sequence replacing the *galK* gene as shown in (**h**).When the I-*Sce*I restriction enzyme expression plasmid pBAD-I-*Sce*I is introduced into the bacteria harbouring the above modified BAC, a double-strand break at the I-*SceI* site is created as illustrated in (**i**). This double-strand break will encourage homologous recombination between the duplicated homologous regions originally flanking the I-*Sce*I-Kan cassette as in (**j**). This homologous recombination finally results in a seamless mutation in both copies as in (**k**).


a
TRL
IRL
IRS
TRS
H1
H2
H2
H1
b
H3
H3
I-SceI-Kan
H3
H1
H2
c
TRL
IRL
IRS
TRS
H1
H2
H2
H1
H1
H3
I-SceI-Kan
H3
H2
d
TRL
IRL
IRS
TRS
H1
H3
I-SceI-Kan
H3
H2
H2
H1
e
TRL
IRL
IRS
TRS
H1
H3
I-SceI-Kan
H3
H2
H2
H1
H2
galK
H1
f
TRL
IRL
IRS
TRS
H1
H3
I-SceI-Kan
H3
H2
H2
galK
H1
g
TRL
IRL
IRS
TRS
H1
H3
I-SceI-Kan
H3
H2
H2
galK
H1
H2
H3
H1
h
TRL
IRL
IRS
TRS
H1
H3
I-SceI-Kan
H3
H2
H2
H3
H1
i
H1
H3
I-SceI-Kan
H3
H2
j
H1
H3
I-SceI-Kan
H3
H2
k
TRL
IRL
IRS
TRS
H1
H3
H2
H2
H3
H1

They should have the following overall design.

Forward: 5′------40–50bp_Homology_1-----AGGATGACG-ACGATAAGTAG-3′

Reverse: 5′----40–50bp_homology_2_complementary strand---50 bp_homology_1_ complementary strand------ CAACCAATTAACCAATTCTGATTAG-3′

For point mutation types of modification, the transfer constructs can be generated by PCR amplification of the I-*Sce*I-Kan cassette with a forward primer containing a 40–50-bp homologous sequence before the gene to be mutated, then with 50-bp sequence with designed mutation followed by sequence amplifying I-*Sce*I-Kan. The Reverse primer contains 40–50-bp homologous sequence adjacent to the gene to be mutated followed by 50-bp designed mutation sequence and finally with the sequence amplifying I-*Sce*I-Kan. The overall design of the oligonucleotides will be as follows:

Forward: 5′---------40–50bp_Homology_1----------50bp_with point mutation----------AGGATGACGACGATAAGTAGGG-3′

Reverse: 5′------40–50bp_homology_2_complementary strand ------50bp_with point mutation _complementary strand ------CAACCAATTAACCAATTCTGATTAG-3′.

For inserting a gene of interest, a cloning step is required. First, a unique restriction site within the gene of interest has to be identified. Then, a pair of primers to amplify I-*Sce*I-kan with 50-bp duplication around the restriction site will have to be designed. The cassette is then inserted into the gene of interest through the unique restriction site. Finally, another pair of primers to amplify this gene of interest with target homology has to be designed.

Fig. 4. Schematic representation of principle steps for seamless addition of a gene of interest into the two copies of viral genes in terminal repeats region. (**a**) Viral genomic location where gene intends to be introduced (*arrow* represents a viral gene, or a gene of interest such as the tandem-affinity purification (TAP) tag or GFP to be introduced at its N-terminus). H1 and H2 represent two homologous regions used for designing PCR primers and/or targeting vectors. The homologous regions are chosen next to the sites where a new gene is to be introduced. (**b**) Represents a new gene to be introduced, and a 50-bp homology region H3 is chosen in the middle of the gene where a unique restriction enzyme cleavage site is available. Clone the I-*Sce*I-Kan into the unique restriction site with duplicated 50-bp homologous region H3, and the whole cassette is amplified by PCR with primers having H1 and H2 homologous regions. (**c**) Induction of homologous recombination after the electroporation of the I-*Sce*I-Kan-homology PCR products into SW102 cells carrying the BAC clone. (**d**) The new gene with I-*Sce*I-kan, as well as duplicated sequences H3, is introduced into the BAC. (**e**) The second copy recombination through a *galK* PCR product with homologous regions while maintaining the Kan selection pressure will result in *galK* insertion as shown in (**f**) at the site where the new gene is to be introduced. (**g**) The new gene with flanking homologous sequences (can be achieved by PCR primers with added 50-bp homologous sequence, or cloning) are introduced into bacteria, with the help of induced homologous recombination and *galK* negative selection, and result in the new gene replacing the *galK* gene as in (**h**) and located in frame with the N-terminal viral gene. (**i**) When the I-*Sce*I restriction enzyme expression plasmid pBAD-I-*Sce*I is introduced into the bacteria harbouring the above modified BAC, a double-strand break at I-*Sce*I site is created. This double-strand break will encourage homologous recombination between the duplicated homologous regions originally flanking the I-SceI-Kan cassette as shown in (**j**). This homologous recombination finally results in the introduction of a new gene at the N-terminal of a viral gene in both copies as in (**k**).

The first pair of primers for cloning I-*Sce*I-Kan into the gene of interest can have the following design:

Forward: 5′-Restriction site-------AGGATGACGACGATAAGTAG-3′
Reverse: 5′-Restriction site------50 bp_before_restriction_site_compl._strand----------CAACCAATTAACCAATTCTGATTAG-3′

The second pair for the transfer constructs has the following design:

Forward: 5′---------50bp_homology_1--------beginning_of_gene_of_interest-3′
Reverse: 5′---------50bp_homology_2_complementary_strand-------end_ gene_of_interest_complementary_strand-3′

3.2.2. Oligonucleotides for the galK Selections

The primers for *gal*K positive selection can be designed as follows:

Forward: 5′------50bp_homology_1-------CCTGTTGACAATTAATCATCGGCA-3′
Reverse: 5′--------50bp_homology_2_complementary_strand----------TCAGCACTGTCCTGCTCCTT-3′

The primers for *gal*K negative selection can be designed as follows:

For deletion:

Top strand: 5′------50bp_homology_1-------50bp_homology_2—3′
Bottom strand: is the reverse complementary of the top strand.

For introducing a point mutation:

Top strand: 5′--40–50 bp_homology_1--point mutation--40–50 bp homology_2------3′

Bottom strand is the reverse complementary of the top strand.

For insertion:

Forward: 5′------50bp_homology_1-------20bp_gene_of_interest------3′
Reverse: 5′----------50bp_homology_2_complementary_strand-------20bp_gene_of_interest_complementary strand---3′

3.3. Preparation of DNA for Electroporation

PCR amplification of I-*Sce*I-kan or *galK* cassettes with the designed primers is carried out using *Taq* DNA polymerase or a proofreading *Taq*-mix (For example, the Expand High Fidelity from Roche, or KOD enzyme from Merck) with 1–2 ng of template plasmid (for example pEPkan-S or *gal*K plasmid). The 30-cycle PCR programme consists of 94°C 20 s, 58°C 30 s and 72°C 1–2 min. The programme annealing temperature and extension time can be adjusted according to the primers used and the PCR products size. After the PCR, 1–2 µl *Dpn*I per 25 µl reaction

is added and incubated at 37°C for 1 h to digest the plasmid template without affecting the PCR products. *Dpn*I-digested PCR products can be gel-purified, and eluted in 30 µl ddH_2O (or lesser volume if the PCR products are less) using Qiagen MinElute gel extraction kit.

When using *galK* as a negative selection marker to introduce large fragments of foreign sequences of sizes more than a few kilobases, small homology arms of 50 bp are usually not sufficient for effective recombination. In such circumstances, the gene of interest can be cloned into an appropriate vector using conventional cloning techniques so as to provide longer homology arms up to 1 kb at each end to improve the efficiency of recombination. The homology region containing the DNA fragment can be released by restriction digestion of the transfer vector and can be used after gel purification to induce homologous recombination.

3.4. Induction of λ Red Recombinase and Preparation of Electrocompetent Bacteria

1. A single SW102 bacterial colony containing BAC is picked and grown overnight at 32°C with vigorous shaking (220 rpm) in 5 mL LB containing 25 µg/mL of chloramphenicol.
2. Dilute the overnight culture 1:50, i.e. transfer 500 µL into a sterile 50 mL Erlenmeyer baffled flask with 25 mL LB and 25 µg/mL of chloramphenicol. Incubate at 32°C for 3–5 h with vigorous shaking until the density reaches an OD_{600} of 0.5–0.6.
3. For the induction of homologous recombination using the heat shock step, induce 15 mL culture in a 100 mL flask at 42°C water bath with shaking, for 15 min (if the induction step is not to becarried out, go directly to step 4) (see Note 7).
4. Cool down the bacterial culture in ice/water bath with shaking for 2–5 min and subsequently transfer into a pre-cooled 14 mL round bottom Falcon tube.
5. Spin down the bacteria in a cold centrifuge (0–4°C) for 7–8 min at 4,000 rpm (3,000 g) or 5 min at 5,000 rpm (4,500 g) in a pre-cooled rotor.
6. Pour off all the supernatant, return to ice and use the pipette to gently remove the rest of the liquid without disturbing the pellet. After the addition of 1 mL ice-cold ddH_2O (see Note 8), the pellet is resuspended by gently moving the tubes around in circles while keeping them in the ice/water slurry. This can take a few minutes. When fully resuspended, add another 10 mL of ice cold ddH_2O, invert the tube twice, and spin again for 8 min at 4,000 rpm (3,000 g).
7. Repeat the procedure in the step 6 once.
8. Gently remove all supernatant by inverting the tube on a paper towel very briefly or shake out the final amount (be careful not to lose the pellet). Store the competent cells on

ice. Gently resuspend the pellet as before in the residual small amount of ddH_2O left in the tube. The final volume should be around 50 µl (see Note 9).

3.5. Transformation of Bacteria by Electroporation

1. Transfer 25 µL of the freshly made electro-competent cells to a pre-cooled 0.5 ml Eppendorf tube and mix with the DNA to be transformed (1–3 µL DNA). Transfer to a pre-cooled 0.1 cm gap cuvette (BioRad).
2. Electroporation of the bacteria in the chilled 0.1 cm cuvette is performed with a Bio-Rad Gene Pulser (or equivalent electroporator) set at capacity 25 µF, resistance at 200 Ω (ohms), and voltage set at 1.80 kV (If using 0.2 cm gap cuvette, set the voltage at 2.5 kV). Pulse once by pressing the two red buttons at the same time until the Gene Pulser beeps. Check that the final pulse time is between 0.42 and 0.47 ms. Transfer the bacteria to a tube containing 1 mL LB medium. Incubate at 32°C with shaking for 1–1.5 h. When using *galK* as negative selection, resuspend the bacteria in 10 mL of LB and shake the bacteria at 32°C for 4.5 h. After the recovery period, spread bacteria onto plates according to the selection requirement (see subheading 3.6 and Note 10).

3.6. Modifying Two Copies of Genes in the Repeat Regions of Herpesvirus Genomes

3.6.1. Insertion of I-SceI-Kan into One Copy of the Gene

Design the PCR primers as described in Subheading 3.2.1 and carry out PCR amplification as described in Subheading 3.3. After *Dpn*I digestion and gel purification, the PCR products are electroporated into SW102 bacteria containing the BAC clones as indicated in Subheading 3.4. After the recovery period, 10 and 100 µL of the bacteria are spread directly onto separate LB plates containing chloramphenicol 25 µg/mL and kanamycin 50 µg/mL. Spin down the remaining bacteria in an Eppendorf tube at 13,000 rpm (16,200 *g*) for 15 s and resuspend in 100 µL of LB medium, and spread onto another plate. Incubate at 32°C for 1–2 days. Pick about ten well-separated bacterial colonies, and inoculate half of each colony (see Note 11) into LB medium with chloramphenicol (25 µg/mL) and kanamycin (50 µg/mL). The other half of the colony is used directly for PCR screen using a pair of primers designed from the site outside the homology region used in the transfer vector, so that PCR product size is different once the I-*Sce*I-Kan is inserted into target locus. At this stage, the PCR products should be two bands, one represents the I-*Sce*I-Kan inserted copy, and the other represents the non-modified copy. It is very unlikely that both copies of the gene are targeted simultaneously in a single step. Choose 1–4 clones with the expected PCR product and make mini-preparations of BAC, digest with a restriction enzyme and check the integrity of the viral genome (see Note 6). Then, choose one clone for the next step.

3.6.2. Insertion of galK into the Second Copy of the Gene

Design the PCR primers as described in Subheading 3.2.2 and carry out PCR amplification using *galK* plasmid as template. Electroporate *Dpn*I digested and gel-purified PCR products into the heat-induced modified SW102 as prepared in Subheading 3.4 using the same procedures as described in Subheading 3.5. After the recovery period, spin down the bacteria in an Eppendorf tube at 13,200 rpm (16,800 g) for 15 s, resuspend in 1 mL of ddH_2O or 1 x M9 salts and pellet the bacteria again. Resuspend the bacteria in 500 μL of ddH_2O or 1 x M9, spread 100 μL of serial 1 x M9 dilutions (that is 100 μL, 100 μL of a 1:10 dilution, and 100 μL 1:100 dilution) onto *galK* positive selection plates, with chloramphenicol 25 μg/mL and kanamycin 50 μg/mL. Wrap the plates with cling film and incubate at 32°C for 3–5 days. Pick about 2–6 relatively large colonies, grow and screen the bacteria by PCR as described above. At this stage, the size of the two bands of the PCR products should be different from those of the non-modified constructs. If the two modified copy bands are similar in size, the PCR product can be digested with a restriction enzyme that can distinguish between the products to ensure that both copies do exist and are modified.

3.6.3. Modifying the galK Inserted Copy into Desired Sequence

After PCR screening and restriction digestion analysis, a colony purification step is required (see Note 12). A single colony can be grown in 5 mL LB culture at 32°C for a few hours and streaked onto MacConkey indicator plates. After incubation of the plates at 32°C for 1–2 days, a single well-separated bright red/pink colony can be picked up and grown up at 32°C in 5 mL LB containing 25 μg/mL chloramphenicol and 50 μg/mL kanamycin. After the induction of λ Red recombinase, electrocompetent bacteria can be prepared from these cells as described in Subheading 3.4.

If the transformation is to be carried out using double-stranded DNA oligonucleotides (as designed in Subheading 3.2.2), two complementary oligonucleotides are to be annealed in vitro. For this, mix 10 μg of each oligonucleotide in a 100 μL volume of 1× PCR buffer and after boiling the samples for 5 min, the reaction is allowed to cool slowly to room temperature over a 30 min period. The sample is precipitated by the addition of 10 μL of 3 M sodium acetate and 250 μL absolute ethanol. The precipitate is pelleted, washed once in 70% ethanol and air-dried before dissolving in ddH_2O to obtain a final concentration of 200 ng/μL.

Transform the bacteria using the DNA or the annealed oligonucleotides as described in Subheading 3.5. Grow bacteria in 10 mL LB in a 50 mL baffled conical flask by incubating in a 32°C shaker for 4.5 h. After 4.5 h recovery time, spin 1 mL bacterial culture down at 13,000 rpm (16,200 g) for 15 s, remove the supernatant and resuspend the pellet in 1 mL of 1× M9, and spread 100 μL of serial dilutions (that is 100 μL of a 1:10 dilution; 100 μL 1:100; 100 μL 1:1,000; 100 μL 1:10,000) onto *galK*

negative selection plates. Wrap the plates with cling film and incubate at 32°C for 3–5 days. Analyse about ten colonies by PCR as in earlier steps and this time the PCR product should be different from the *galK* and the I-*Sce*I-Kan insertion copies. The PCR product can be sequenced directly to confirm the seamless modification of the previously *galK* inserted copy. At this stage, the bacteria should still be kanamycin-resistant, although the *galK* negative selection plates did not have the kanamycin selection pressure. There is little chance that both copies can undergo homologous recombination at same time with perfect viral genome integrity. Prepare plasmid DNA from the putative clones and digest with an appropriate restriction enzyme to check the integrity of viral genome by comparing with the parental BAC DNA digests. Finally, one of the clones is selected for the final step of the procedure (see Note 13).

3.6.4. Modifying the I-SceI-Kan Inserted Copy into Desired Sequence

Grow a colony of bacteria from the previous step and prepare electrocompetent cells as described in Subheading 3.4, but omitting the heat-inducing step. 1–10 ng of pBAD-I-*Sce*I is then electroporated into these cells as outlined in Subheading 3.5. After 1 h recovery in LB medium, the bacteria are plated onto LB plates containing 25 μg/mL chloramphenicol and 50 μg/mL ampicillin, and incubated at 32°C for 1–2 days. Pick a single colony and grow overnight in LB medium with 25 μg/mL chloramphenicol and 50 μg/mL ampicillin, then dilute 1:50 (400 μL into an autoclaved 50 mL Erlenmeyer baffled flask with 20 mL LB with chloramphenicol and ampicillin) and incubate for 2–4 h at 32°C with vigorous shaking (220 rpm). Add 200 μL of 10% l-arabinose to the culture (final concentration 0.1%) and continue to incubate at 32°C with shaking for 1 h. Transfer the culture into 42°C water bath shaker for 15 min. Return the culture to 32°C and shake for another 1–4 h. Plate 100 μL each of the serial dilutions (e.g. 1:10, 1:100, 1:1,000) on LB plates containing 25 μg/mL chloramphenicol and 1% l-arabinose and incubate at 32°C for 1–2 days. About ten colonies can then be checked by PCR, sequencing and restriction fragment length polymorphism (RFLP) analysis.

3.7. BAC DNA Mini-Preparation

1. Grow a single bacterial colony in 5 mL LB with chloramphenicol (in a 20 mL universal tube) at 32°C overnight with shaking. Take 1.5 mL into an Eppendorf tube and pellet the bacteria at 13,000 rpm (16,200 g) for 1 min and remove the supernatant.
2. Resuspend the pellet with 70 μl of STET buffer (8% Sucrose, 5% Triton X-100, 50 mM EDTA, 50 mM Tris–HCl, pH 8.0) by continuous vortexing.
3. While vortexing the Eppendorf, add 200 μl of P2 solution contained in a Qiagen plasmid mini-prep kit, followed immediately by the addition of 150 μl of 7.5 mM ammonium acetate.

4. Incubate the mixture in ice for 5 min and then centrifuge at 13,000 rpm (16,200 g) for 20 min at 4°C.
5. Pour the supernatant into a new eppendorf tube, add 240 µl of isopropanol, and mix by vortexing.
6. Pellet the DNA by centrifugation 13,000 rpm (16,200 g) at room temperature for 5 min.
7. Discard the supernatant and wash the pellet with 200 µl of 70% ethanol. Centrifuge for 3 min and discard the supernatant, air dry the pellet and dissolve it in 50 µL TE with 5 µg/mL RNAse A.

3.8. BAC DNA Maxi-Preparation

Grow about 300 mL LB culture of bacteria at 32°C for 20–24 h, and make Maxi plasmid DNA with a Qiagen Plasmid Maxi kit, following the low copy plasmid protocol, but with the final elution step with QF solution warmed at 65°C.

4. Notes

1. Antibiotic concentration used for BAC selection should be about half amount used for multi-copy plasmids. For example: chloramphenicol (cm): 12.5–25 µg/mL, kanamycin (kan): 25–50 µg/mL, spectonomycin (Spc): 50 µg/mL, ampicillin (amp): 50 µg/mL, tetracycline (tet): 12.5 µg/mL.
2. If a frequently used gene, such as GFP, TAP-tag to be repeatedly used, it becomes worthwhile to clone the I-SceI-Kan cassette into the gene of interest. To do this, a unique restriction site needs to be identified in the gene of interest. Then design a pair of PCR primers as in Subheading 3.2.1 for the insertion type of modification, following the guide line of the first pair of primers in that section, and clone in the PCR product into this unique restriction site, and finally check the orientation.
3. This solution is not absolutely required for making *galK*-based selection plates, but the bacteria grow better in a plate with it.
4. The p*galK* with plasmid detailed information is available from the recombineering website maintained by Dr Neal G. Copeland' group: http://recombineering.ncifcrf.gov/Plasmid.asp. The pBAD-I-SceI and pEPKan-S2 can be obtained from Dr Nikolaus Osterrieder, Institut für Virologie, Freie Universität Berlin, Berlin, Germany. E-mail: no34@cornell.edu.
5. The procedure described is aimed to achieve seamless modifications in both copies. Sometimes, however, when a bacterial selection marker or even a smaller "scar" sequence such as a *Lox*P or FRT sequence can be tolerated, the procedure can be

simplified. For this, just design two sets of primers, with 50-bp homologous sequences linked to sequences that are able to amplify two different positive selection markers such as kanamycin and spectinomycin. Replace one copy with kanamycin and the other copy with spectinomycin through two steps of λ Red recombination. If the selection markers are flanked by FRT sites (as in plasmids pKD13 or pL451), then bacterial strain SW105 can be used for easily "flipping" off the selection marker after the recombination events. If the selection marker is flanked by *Lox*P site (as in plasmid pL452), then bacterial strain SW106 can be used for "floxing" out the selection marker.

6. Oligonucleotides designed for PCR amplification of transfer constructs are usually long. For oligonucleotides less than 80 bp in length, sequence errors are usually scarce. For oligonucleotides more than 100 bp long, choosing a reputable oligonucleotide provider is important. Oligonucleotides need not be specially purified as just desalting is sufficient in most cases. The major drawback for I-*Sce*I-Kan selection is that the oligonucleotides need to be very long, and errors can occur. Thus, once the I-*Sce*I-Kan cassette is in the BAC, it is very important to screen a few clones by PCR and sequencing across the primers region to choose the correct clone.
7. When the bacteria are used to introduce a BAC clone, or a plasmid (like pBAD-I-*Sce*I), there is no need for the heat shock step. To transform with a BAC, use 1 μg DNA, ideally made freshly. For plasmids, 1–10 ng of plasmid DNA is often sufficient.
8. The water needs to be chilled in ice for at least 2 h. If the competent cells are to be used right away, use ddH_2O. If the competent cells are going to be stored at –70°C for later use, you must use 10% glycerol.
9. Efficiency of electrocompetent bacteria is important for successful homologous recombination. Try to keep the bacteria as close to 0°C as possible in order to get efficient competent cells. Test the transfection efficiency of your electrocompetent SW102 bacteria with an AmpR plasmid with a colE1 origin. The efficiency needs to be higher than 1×10^7 per μg of DNA.
10. When spreading the bacteria onto *gal*K selection plates (both positive and negative), extra care has to be taken not to damage the soft agar surface. A wet spreader and reduced force while spreading can help to reduce the damage. Spread by rotating the plate and hold the spreader steady. Stop once the culture is absorbed into the agar.
11. The bacterial colonies are usually appeared after 24 h of incubation at 32°C. For easy colony picking, plates are usually incubated for 36–48 h. A plate of well-separated bacterial

colonies that result from right bacteria dilutions is chosen. A fine-tip bacterial inoculation needle or a pipette tip is used to touch a bacterial colony and put it into LB liquid culture. The same colony is picked again with another tip. This time a bit more bacteria are required. The tip with eye visible bacteria is dipped into PCR reaction mixture and mixed well to allow the bacteria dispersed into the solution.

12. After the *galK* positive selection, the bacterial colonies are picked from the minimal selective plates and grown in LB liquid culture. As the LB is not a selective medium for *galK* anymore, the colonies that are picked are not 100% pure and therefore give a very high background when using *galK* negative selection for the loss of *galK*. Hence, a colony purification step must be used for this step.
13. When *galK* is used for positive selection, it is efficient and powerful, with little background. When the *galK* is used for negative selection, there will be some background. This is because the strong counterselection force which enriches the rare spontaneous mutations or deletions in the *galK*. Increasing the length of homology arms can reduce the background. Longer homology arms can be added by traditional cloning.

Acknowledgments

The authors would like to thank Dr. Nikolaus Osterrieder for providing pBAC-I-*Sce*I and pEPKan-S plasmids; Dr. Neal G. Copeland for p*galK* plasmid and bacterial strain SW102, SW105, and SW106. The protocols are modified from the publication of Warming et al. (2005) (19) and Tischer et al. (2006) (21).

References

1. Smith GA, Enquist LW (2000) A self-recombining bacterial artificial chromosome and its application for analysis of herpesvirus pathogenesis. Proc Natl Acad Sci U S A A97:4873–4878
2. Yu D, Smith GA, Enquist LW, Shenk T (2002) Construction of a self-excisable bacterial artificial chromosome containing the human cytomegalovirus genome and mutagenesis of the diploid TRL/IRL13 gene. J Virol 76:2316–2328
3. Zhao Y, Petherbridge L, Smith LP, Baigent S, Nair V (2008) Self-excision of the BAC sequences from the recombinant Marek's disease virus genome increases replication and pathogenicity. Virol J 5:19
4. Tanaka M, Kagawa H, Yamanashi Y, Sata T, Kawaguchi Y (2003) Construction of an excisable bacterial artificial chromosome containing a full-length infectious clone of herpes simplex virus type 1: viruses reconstituted from the clone exhibit wild-type properties in vitro and in vivo. J Virol 77:1382–1391
5. Tischer BK, Kaufer BB, Sommer M, Wussow F, Arvin AM, Osterrieder N (2007) A self-excisable infectious bacterial artificial chromosome clone of varicella-zoster virus allows analysis of the essential tegument protein encoded by ORF9. J Virol 81:13200–13208
6. Rudolph J, O'Callaghan DJ, Osterrieder N (2002) Cloning of the genomes of equine herpesvirus type 1 (EHV-1) strains KyA and

racL11 as bacterial artificial chromosomes (BAC). J Vet Med B Infect Dis Vet Public Health 49:31–36

7. Petherbridge L, Brown AC, Baigent SJ et al (2004) Oncogenicity of virulent Marek's disease virus cloned as bacterial artificial chromosomes. J Virol 78:13376–13380
8. Baigent SJ, Petherbridge LJ, Smith LP, Zhao Y, Chesters PM, Nair VK (2006) Herpesvirus of turkey reconstituted from bacterial artificial chromosome clones induces protection against Marek's disease. J Gen Virol 87: 769–776
9. Costes B, Fournier G, Michel B et al (2008) Cloning of the koi herpesvirus genome as an infectious bacterial artificial chromosome demonstrates that disruption of the thymidine kinase locus induces partial attenuation in Cyprinus carpio koi. J Virol 82:4955–4964
10. Kanda T, Yajima M, Ahsan N, Tanaka M, Takada K (2004) Production of high-titer Epstein-Barr virus recombinants derived from Akata cells by using a bacterial artificial chromosome system. J Virol 78:7004–7015
11. Zhou FC, Zhang YJ, Deng JH et al (2002) Efficient infection by a recombinant Kaposi's sarcoma-associated herpesvirus cloned in a bacterial artificial chromosome: application for genetic analysis. J Virol 76:6185–6196
12. Estep RD, Powers MF, Yen BK, Li H, Wong SW (2007) Construction of an infectious rhesus rhadinovirus bacterial artificial chromosome for the analysis of Kaposi's sarcoma-associated herpesvirus-related disease development. J Virol 81:2957–2969
13. Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL (2000) An efficient recombination system for chromosome engineering in Escherichia coli. Proc Natl Acad Sci U S A 97:5978–5983
14. Lee EC, Yu D, Martinez de Velasco J et al (2001) A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics 73:56–65
15. Borst EM, Hahn G, Koszinowski UH, Messerle M (1999) Cloning of the human cytomegalovirus (HCMV) genome as an infectious bacterial artificial chromosome in Escherichia coli: a new approach for construction of HCMV mutants. J Virol 73:8320–8329
16. Sopher BL, La Spada AR (2006) Efficient recombination-based methods for bacterial artificial chromosome fusion and mutagenesis. Gene 371:136–143
17. Gay P, Le Coq D, Steinmetz M, Ferrari E, Hoch JA (1983) Cloning structural gene sacB, which codes for exoenzyme levansucrase of Bacillus subtilis: expression of the gene in Escherichia coli. J Bacteriol 153:1424–1431
18. Wong QN, Ng VC, Lin MC, Kung HF, Chan D, Huang JD (2005) Efficient and seamless DNA recombineering using a thymidylate synthase A selection system in Escherichia coli. Nucleic Acids Res 33:e59
19. Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG (2005) Simple and highly efficient BAC recombineering using galK selection. Nucleic Acids Res 33:e36
20. Posfai G, Kolisnychenko V, Bereczki Z, Blattner FR (1999) Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome. Nucleic Acids Res 27:4409–4415
21. Tischer BK, von Einem J, Kaufer B, Osterrieder N (2006) Two-step red-mediated recombination for versatile high-efficiency markerless DNA manipulation in *Escherichia coli*. Biotechniques 40:191–197

Chapter 5

An Efficient Protocol for VZV BAC-Based Mutagenesis

Zhen Zhang, Ying Huang, and Hua Zhu

Abstract

Varicella-zoster virus (VZV) causes both varicella (chicken pox) and herpes zoster (shingles). As a member of the human herpesvirus family, VZV contains a large 125-kb DNA genome, encoding 70 unique open reading frames (ORFs). The genetic study of VZV has been hindered by the large size of viral genome, and thus the functions of the majority of these ORFs remain unclear. Recently, an efficient protocol has been developed based on a luciferase-containing VZV bacteria artificial chromosome (BAC) system to rapidly isolate and study VZV ORF deletion mutants.

Key words: Varicella-zoster virus, Bacterial artificial chromosome, Deletion mutagenesis, Bioluminescence

1. Introduction

Varicella-zoster virus (VZV) is a common human herpesvirus that is a significant pathogen in the United States, with more than 90% of the US population harboring the virus (1). Primary infection of VZV leads to varicella (chicken pox). VZV establishes lifelong latency in the host, specifically in trigeminal ganglia and dorsal root ganglia (2). The VZV reactivation results in herpes zoster (shingles), which often leads to chronic postherpetic neuralgia (2, 3). As a member of human alpha-herpesvirus subfamily, VZV has a 125-kb long double-stranded DNA genome, which encodes at least 70 unique open reading frames (ORFs). The genomes of several different VZV strains were sequenced and a few of the VZV genes genetically analyzed (4).

It has been extremely difficult to generate VZV site-specific mutations using conventional homology recombination methods. This was mainly due to the high cell-associated nature of VZV infection in vitro, which leads to the difficulties in isolating

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_5,

viral DNA and purifying recombinant virus away from wild-type virus. In the last few years, a popular method for VZV in vitro mutagenesis involves a four-cosmid system covering the entire viral genome (5–7). Using the cosmid system to generate recombinant VZV variants involves technically challenging steps such as co-transfection of four large cosmids into permissive mammalian cells and multiple homologous recombination events within a single cell to reconstruct a full-length viral genome. The highly cell-associated nature of VZV also makes the downstream applications of traditional virology methods such as plaque assay-based titering and plaque purification difficult. To date, the functions of the majority of VZV ORFs remain uncharacterized (8).

In order to create recombinants of VZV more efficiently, the full-length VZV (P-Oka strain, a cloned clinical isolate of VZV) genome has been successfully cloned as a VZV bacteria artificial chromosome (BAC) (9, 10). This VZV BAC combined with a highly efficiently *E. coli* homologous recombination system allows quick and easy generation of recombinant VZV. To further ease the downstream virus quantification assays, a firefly luciferase reporter gene, was inserted into the VZV BAC to generate a novel luciferase-expressing VZV (10). In this protocol, we show the generation and analyses of VZV full-length ORF deletion mutants and genetic revertants as examples to demonstrate the utility and efficiency of this versatile system for VZV mutagenesis in vitro. Furthermore, this protocol can be easily modified to broaden its applications to a variety of genetic maneuvers including making double ORF deletions, partial ORF deletions, insertions, and point mutations.

2. Materials

2.1. Cells, VZV_{luc}, Plasmids, and E. coli Strain

1. Human melanoma (MeWo) cells were grown in DMEM supplemented with 10% fetal bovine serum, 100 U penicillin–streptomycin/ml, and 2.5 µg amphotericin B/ml at 37°C in a humidified incubator with 5% CO_2. All tissue culture reagents were obtained from Sigma (St. Louis, MO).
2. VZV_{luc} was recently developed in the laboratory (10). It contains a full-length VZV P-Oka genome with a firefly luciferase cassette (see Note 1). The BAC vector was inserted between VZV ORF60 and ORF61, which includes a green fluorescent protein (GFP) expression cassette and a chloramphenicol resistance cassette (Cm^R).
3. pGEM-oriV/kan was previously constructed (11) in the laboratory and used as a PCR template to generate the expression cassettes for the kanamycin or ampicillin resistance genes (Kan^R and Amp^R).

4. pGEM-lox-zeo was derived from pGEM-T (Promega, Madison, WI) (12) and was used to generate the rescue clones of VZV ORF deletion mutants.
5. *E. coli* strain DY380 was obtained from Neal Copeland and Craig Stranthdee and used for mutagenesis (13).
6. A cre recombinase expression plasmid pGS403 was a gift from L. Enquist (Princeton University, NJ).

2.2. Primers, PCR, PCR purification, DpnI Treatment, and Electroporation

1. All primers were synthesized by Sigma-Genosys (Woodlands, TX) and stored in TE buffer (100 μM).
2. HotStar Taq DNA polymerase (Qiagen, Valencia, CA) was used for PCR reactions and Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA) could be used for optional hi-fidelity PCR reactions (see Note 2).
3. PCR purification was carried out using a PCR purification kit (Qiagen, Valencia, CA).
4. The amplified linear DNAs were suspended in sterile ddH_2O and were quantified by spectroscopy (NanoDrop Technologies, Wilmington, DE).
5. DpnI (New England Biolabs, Ipswich, MA) restriction treatment following PCR was carried out in order to eliminate circular template DNA.
6. Electroporation was carried out with a Gene Pulser II Electroporator (Bio-Rad, Hercules, CA).

2.3. Antibiotics Selection and BAC DNA Purification

1. All antibiotics were obtained from Sigma (St. Louis, MO). LB plates containing specific antibiotics were used for appropriate selections (Table 1).
2. A 37°C air shaker and a 37°C water bath shaker were used for bacterial culturing.

Table 1
Antibiotics concentrations for selection

For BACs (single or low copy numbers)	For plasmids (high copy numbers)
Chloramphenicol (cm): 12.5 μg/ml	
Hygromycin B (hyg): 50 μg/ml	
Kanamycin (kan): 30 μg/ml	Kanamycin (kan): 50 μg/ml
Ampicillin (amp): 50 μg/ml	Ampicillin (amp): 100 μg/ml
Zeocin (zeo): 50 μg/ml	Zeocin (zeo): 100 μg/ml

3. NucleoBond Xtra Maxi Plasmid DNA purification kits (Clontech Laboratories, Inc., Palo Alto, CA) were used to purify VZV BAC DNA from *E. coli*.
4. Kimwipes (Kimberly-Clark Global Sales, Inc., Roswell, GA) were used as small filters in BAC DNA Mini-preparations.
5. Phenol/chloroform, isopropanol, and ethanol were obtained from Sigma (St. Louis, MO) and were used as additional reagents in BAC DNA preparations.
6. *Hin*dIII (New England Biolabs, Ipswich, MA) digestions were performed to check the integrity of BAC DNA.

2.4. Transfection and Subsequent Virological Assays (Tittering and Growth Curve Analysis)

1. FuGene6 transfection kit (Roche, Indianapolis, IN) was used for transfecting viral BAC DNA into MeWo cells (ATCC).
2. An inverted fluorescent microscope was used to observe and count plaque numbers.
3. Tissue culture media containing 150 μg/ml D-luciferin (Xenogen, Alameda, CA) was used as substrate for in vitro bioluminescence detection.
4. An IVIS Imaging 50 System (Xenogen, Alameda, CA) was used to record bioluminescence signal from virally infected cells.
5. Bioluminescence data were quantified by using Living Image analysis software (Xenogen, Alameda, CA).

3. Methods

3.1. Generation of VZV ORF Deletion BAC Clones

In order to generate VZV ORF deletion mutants using this new VZV_{luc} system, we took advantage of an efficient recombination system for chromosome engineering in *E. coli* DY380 strain (13). A defective lambda prophage supplies the function that protects and recombines linear DNA. This system is highly efficient and allows recombination between homologies as short as 40 bp. The experimental design is summarized in Fig. 1.

3.1.1. Making a Kan^R Cassette Targeting a Specific VZV Open Reading Frame

1. The first step in making any specific VZV ORF deletion was to amplify a Kan^R cassette containing 40-bp flanking sequences of the targeted ORF.
2. Primers were stored in TE buffer (100 μM). The Kan^R expression cassette was amplified from pGEM-oriV/Kan using a HotStar DNA polymerase kit following a standard protocol.
3. PCR product was purified using a PCR purification kit following the manufacturer's protocol.

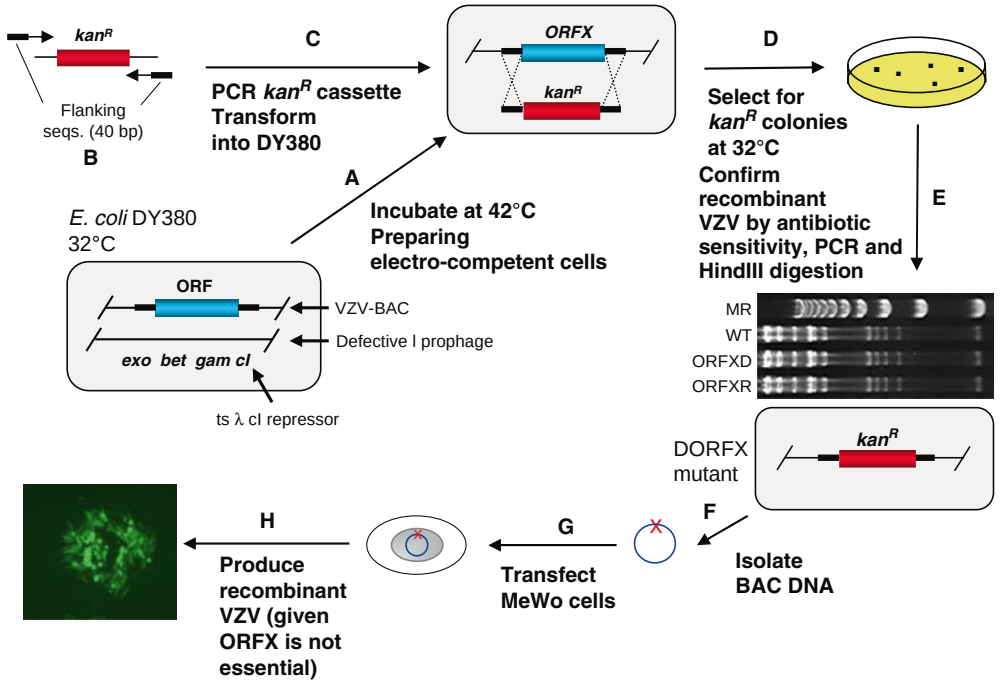


Fig. 1. Generating ORF deletion mutants (ORFD). (**a**) The *E. coli* DY380 strain provides a highly efficient homologous recombination system, which allows recombination of homologous sequences as short as 40 bp. The homologous recombination system is strictly regulated by a temperature-sensitive repressor, which permits transient switching on by incubation at 42°C for 15 min. VZV_{luc} BAC DNA is introduced into DY380 by electroporation. Electro-competent cells are prepared with homologous recombination system activation. (**b**) Amplification of the Kan^R expression cassette by PCR using a primer pair adding 40-bp homologies flanking ORFX. (**c**) About 200 ng of above PCR product are transformed into DY380 carrying the VZV_{luc} BAC via electroporation. (**d**) Homologous recombination between upstream and downstream homologies of ORFX replaces ORFX with the Kan^R cassette, creating the ORFX deletion VZV clone. The recombinants are selected on LB agar plates containing kanamycin at 32°C. (**e**) The deletion of ORFX is confirmed by testing antibiotic sensitivity and PCR analysis. The integrity of viral genome after homologous recombination is examined by restriction enzyme *Hin*dIII digestion. (**f**) VZV_{luc} BAC DNA with ORFX deletion is propagated and isolated from DY380. (**g**) Purified BAC DNA is transfected into MeWo cells. (**h**) 3–5 days after transfection, the ORFX deletion virus is visualized under a fluorescent microscope due to EGFP expression given nonessentiality of ORFX.

4. The purified PCR product was treated with DpnI in order to eliminate the template DNA. This step greatly reduces the background in later selections.
5. PCR product was purified again as above (step 3) and the amplified linear DNA was suspended in sterile ddH_2O and was quantified by spectroscopy (see Note 3).

3.1.2. Induction of the Lambda Recombination System and Preparation of Electroporation-Competent DY380

1. DY380 cells were grown at 32°C until the OD_{600nm} measurement reached 0.5 (see Note 5).
2. The culture was shifted to 42°C by placing the flask into a 42°C water bath with vigorous shaking for 10–15 min (see Note 4).

3. The culture was immediately transferred to an ice–water slurry for 30 min. (see Note 6).
4. After incubation on ice, the culture was then pelleted at $6{,}000 \times g$ for 10 min 4°C, washed with ice-cold sterile ddH_2O, and repelleted.
5. Prechilled 10% glycerol (use about 1% of original volume of culture) was used to resuspend cells, and a 40-μl aliquot ($>1 \times 10^{10}$ cells) was used for each electroporation reaction.

3.1.3. Electroporation and Recombinant Screening

1. Two microliters of Kan^R cassette DNA (greater than 200 ng) were electroporated into competent DY380 cells harboring the VZV_{luc} BAC. Homologous recombination took place between the 40-bp ORF flanking sequences and the targeted BAC ORF was replaced by the linear Kan^R cassette creating the expected VZV ORF deletion clones.
2. Electroporation was carried out at 1.6 kV, 200 Ω, and 25 μF in a Gene Pulser II electroporator. Two microliters of concentrated linear DNA cassette (greater than 200 ng) were used in each reaction.
3. The bacteria were immediately transferred to 1 ml LB medium after electroporation and incubated at 32°C for 1 h before plating. The resultant recombinants were selected on LB agar plates containing kanamycin at 32°C for 16–24 h (see Note 7).
4. Antibiotic sensitivity: it is important to further test that kanamycin-resistant colonies are resistant to kanamycin but not to ampicillin because the circular pGEM-oriV/Kan^R (containing Amp^R) was used as the PCR template. This can be tested by re-streaking single colonies on multiple LB agar plates containing different antibiotics. VZV ORFX deletion clones should be resistant to chloramphenicol (from BAC vector), hygromycin (from luciferase cassette), and kanamycin (VZV ORF replacement cassette), but sensitive to ampicillin (potentially from pGEM-oriV/Kan^R; see Note 8).

3.1.4. BAC DNA Purification and BAC Clone Verification

1. Mini-BAC DNA preparations.
   (a) A single DY380 clone containing the recombinant VZV BAC was inoculated in 5 ml LB supplemented with the appropriate antibiotics and cultured at 32°C overnight.
   (b) BAC DNA was isolated by pelleting the bacteria, resuspending in 1 ml resuspension buffer supplemented with RNase A (Buffer RES), lysing in 1 ml NaOH/SDS lysis buffer (Buffer LYS), and neutralizing in 1 ml potassium acetate neutralization buffer (Buffer NEU) for 5 min for each step (NucleoBond Xtra Maxi Plasmid DNA purification kit).

(c) The cloudy solution was centrifuged at 4,500 × *g* for 15 min at 4°C. The supernatant was filtered through a small piece (cut to 4 × 4 cm) of Kimwipe tissue (Kimberly-Clark Global Sales, Inc., Roswell, GA).

(d) The filtered solution was extracted with an equal volume of phenol/chloroform and the BAC DNA precipitated with two volumes of ethanol.

(e) After the final spin at 4,500 × *g* for 30 min at 4°C, the DNA pellet was air-dried and resuspended in 20 μl sterile ddH_2O.

2. PCR verification: multiple colonies with the correct antibiotic sensitivities were picked for the mini-BAC DNA preparations. The ORF deletions with Kan^R replacements were confirmed by PCR using a HotStar DNA polymerase kit following a standard protocol. The target ORF should be absent in ORF deletion clones while the adjacent ORFs should remain intact as positive controls.

3. Maxi-BAC DNA preparations: the large-scale BAC DNA preparations using the NucleoBond Xtra Maxi Plasmid DNA purification kit (Clontech Laboratories, Inc., Palo Alto, CA) started with 500 ml of overnight cultures. The standard manufacturer's protocol for BAC DNA purification was followed. The final DNA products were resuspended in 250 μl sterile ddH_2O and quantified by spectroscopy (see Note 9).

4. *Hin*dIII digestion profiling: PCR verified clones were selected for maxi-BAC DNA preparations. To confirm that no large VZV genomic DNA segment is deleted, *Hin*dIII digestion profiling was routinely carried out (see Note 10). Three micrograms of BAC DNA from maxi-preparations were digested with 20 U of *Hin*dIII in a 20-μl reaction at 37°C overnight. *Hin*dIII digestion patterns were compared by electrophoresis on ethidium bromide stained 0.5% agarose gels. As shown in Fig. 1, *Hin*dIII digestion patterns of each VZV ORF deletion clone were highly comparable with the parental wild-type VZV_{luc} clone (see Note 11).

3.2. Generation of VZV ORF Deletion Revertant BAC Clones

The generation of VZV ORF deletion revertants is necessary to prove that the deleted ORF is responsible for any phenotype (usually a growth defect) observed in analyses of the deletion mutants. The viral revertants should be able to fully restore the wild-type phenotype. As an example, generating the VZV ORFX deletion rescue virus is described to demonstrate the approach (see Fig. 2).

1. VZV ORFX was amplified from wild-type VZV_{luc} BAC DNA by PCR. Two unique restriction enzyme sites and two additional 6-bp random sequences were added to the ends of the PCR product. A hi-fidelity PCR kit could be used in order to minimize unwanted mutations during PCRs (see Note 2).

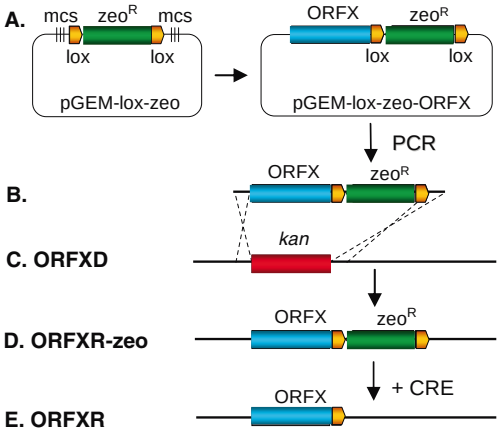


Fig. 2. Generating an ORFX deletion rescue clone (ORFXR). (**a**) To generate the ORFXR clone, ORFX was amplified by PCR from the wild-type VZV BAC DNA. The ORFX was directionally cloned into plasmid pGEM-lox-zeo to form pGEM-zeo-ORFX. (**b**) Amplification of the ORFX-Zeo^R cassette by PCR using a primer pair adding 40 bp homologies flanking ORFX. (**c**) Such PCR product was transformed into DY380 carrying the VZV$_{luc}$ ORFXD BAC via electroporation. (**d**) Homologous recombination between upstream and downstream homologies of ORFX replaced Kan^R with the ORFX-Zeo^R cassette, creating the ORFXR clone. (**e**) Zeo^R was removed while generating virus from BAC DNA by co-transfecting a Cre recombinase expressing plasmid.

2. The ORFX gene was directionally cloned into pGEM-zeo to form pGEM-ORFX-zeo. The cloned ORFX was verified by sequencing analysis.
3. ORFX-zeoR cassette was made by PCR using pGEM-ORFX-zeo as template (Fig. 2). The PCR product contained 40-bp homologies of flanking sequences of Kan^R cassette, which was also used to generate the ORFX deletion mutant.
4. The subsequent procedures are similar to producing the ORFX deletion mutant. Briefly, the linear ORFX-zeoR cassette was treated with DpnI and electroporated into competent DY380 cells harboring VZV$_{luc}$ ORFX deletion BAC. Similarly, homologous recombination functions were transiently induced by switching the culture temperature from 32 to 42°C for 10–15 min when electroporation-competent cells were prepared. The recombinants were selected on LB agar plates containing zeocin. After verification, the ORFX deletion rescue BAC DNA was isolated from *E. coli*.

3.3. Transfection and Subsequent Virological Assays

Because of VZV's highly cell-associated nature in cell culture, conventional virology techniques, including plaque purification and plaque assay, become troublesome. By developing and exploiting the new luciferase VZV BAC system, the resulting virus has a removable EGFP expression cassette and a built-in

luciferase reporter. In this protocol, an alternative bioluminescence quantification approach has been provided to significantly increase the reproducibility of results. This approach has also been successfully used in monitoring VZV growth in vivo (10).

3.3.1. Transfection of BAC DNA into MeWo Cells

1. VZV BAC DNA from maxi-preparations was transfected into MeWo cells using the FuGene6 transfection kit according to the manufacturer's standard protocol.
2. One and a half micrograms of BAC DNA and 6 µl of transfection reagent were used for a single reaction in one well of 6-well tissue culture plates (see Note 12).
3. As an option, 0.5 µg of Cre expression plasmid was co-transfected with the VZV BAC DNA to remove the BAC sequence flanked by two loxP site from the viral genome (see Note 13).
4. In order to prevent the precipitation of BAC in solution, 1.5 µg BAC DNA were diluted in serum-free tissue culture medium, and the volume of DNA solution was adjusted to 50 µl (see Note 14).
5. The DNA solution was slowly added to the transfection reagent with gentle stirring using pipet tips.
6. Because of GFP expression from the BAC vector, VZV plaques were usually visually discernable using a fluorescent microscope within 3–5 days after transfection given deleted ORF is dispensable (see Note 15). If a VZV ORF is essential for viral replication, no plaque will be observed.
7. Since VZV is highly cell-associated in tissue culture, mutant VZV-infected MeWo cells were harvested and stored in liquid nitrogen for future studies.

3.3.2. Titering by Infectious Focus Assay

Recombinant viruses were titered by infectious focus assay. MeWo cells were seeded in 6-well tissue culture plates and inoculated with serial dilutions of VZV-infected MeWo cell suspensions. Plaques were counted by fluorescent microscopy 3 days after inoculation and viral titer was determined.

3.3.3. Growth Curve Analyses Based on Bioluminescence Imaging (See Fig. 3 and Note 16)

1. MeWo cells were infected with 100 PFU of infected MeWo cell suspensions in 6-well tissue culture plates.
2. After every 24-h interval, cell culture media was replaced with media containing 150 µg/ml D-luciferin.
3. After incubation at 37°C for 10 min, the bioluminescent signal was quantified and recorded using an IVIS Imaging System following the manufacturer's instructions.
4. Fresh tissue culture medium was added to replace the luciferin-containing medium for further incubation at later time points.

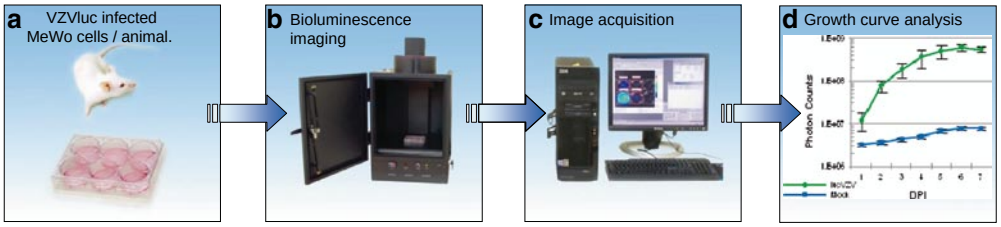


Fig. 3. Growth curve analyses based on bioluminescence imaging. (**a**) Small animals/tissue culture can be infected with VZV_{luc}. (**b**) After administration of an enzyme substrate, luciferin, bioluminescence emitting from living animals/cultured cells can be detected and monitored by using a bioluminescence imaging system (a CCD camera mounted on top of a light-tight imaging dark chamber). (**c**) Data can be stored in a connected PC and quantified by using region-of-interest analysis. (**d**) Viral growth kinetics can be analyzed based on quantification of bioluminescence signals.

5. Measurements from the same plate were repeated every day for 7 days.
6. Bioluminescence signal data from each sample was quantified by manual designation of regions of interest and analyzed using Living Image analysis software (see Note17).

4. Notes

1. The luciferase expression cassette, driven by an SV40 early promoter, was inserted between VZV ORF65 and ORF66. The cassette also contains a hygromycin B resistance gene (Hyg^R).
2. Platinum Taq DNA polymerase can be used alternatively if a hi-fidelity PCR product is preferred.
3. In order to achieve optimum results, the final concentration of the linear DNA cassette for the subsequent electroporation was adjusted to at least 100 ng/μl.
4. The 42°C temperature shift is critical for the success of the homologous recombination. The temperature needs to be adjusted accurately to 42°C and remain constant. Too much recombination system activity is detrimental to *E. coli* and harm the integrity of BAC DNA. On the other hand, inadequate induction of the recombination system in DY 380 leads to inefficient recombination. Ten to fifteen minutes might need to be adjusted carefully in order to achieve optimized efficiency of homologous recombination.
5. *E. coli* DY380 strain needs to be cultured at 32°C all the time except when the recombination system is transiently activated and expressed by shifting the culture to 42°C.
6. Beyond this point, every step needs to be carried out at a low temperature (0–4°C). All reagents, centrifuge rotor and glassware need to be prechilled.

7. Recombinants often have multiple antibiotic resistances. For instance, VZV ORFX/Kan clone will have Kan^R, Cm^R (from BAC vector), and Hyg^R (from luciferase cassette). Screening for recombinants with more than one antibiotic is optional. However, the growth rate under such conditions could be much slower than selection under one antibiotic.
8. If a clone also has Amp^R, it should count as a false positive result.
9. Due to the large size, handling BAC DNAs needs to avoid any harsh physical sheering force including vortexing or quickly passing through fine pipette tips. Freeze and thaw should also be avoided. BAC DNA solutions should always be stored at 4°C.
10. Although it has been shown that VZV_{luc} DNA is highly stable in *E. coli* (10) under the conditions described in this protocol, large undesirable deletions in the BAC clones were observed if homologous recombination system in DY380 was over-induced.
11. Since many large DNA fragments are generated by a *Hin*dIII digestion of the VZV genome, smaller genetic alterations, including replacement of an ORF by a Kan^R cassette, would be difficult to recognize by this assay.
12. The ratio of BAC DNA and FuGene6 reagent might need to be adjusted to maximize transfection efficiency.
13. The ORFX rescue clone was generated by introducing the wild-type ORFX back into the deletion viral genome along with a Zeo^R cassette flanked by two loxP sites. By following this optional step in transfection, Zeo^R will be removed from the genome by Cre-mediated recombination. The resulting virus will have a wild-type copy of ORFX restored in the same direction and location as the parental wild-type strain except a remaining loxP site (34 bp) in the 3′ noncoding region of ORFX.
14. Highly concentrated (greater than 250 μg/μl) BAC DNA solutions are viscous and BAC DNA molecules easily precipitate out of solution when added to transfection reagent solutions. When such precipitation becomes visible, it is irreversible and the result of the transfection assays is often poor. Therefore, we predilute each BAC DNA in media before gently mixing with the transfection reagent.
15. Transfection efficiency was easy to monitor because of the resulting GFP expression from the BACs.
16. Growth curve analyses were traditionally carried out by a plaque assay-based method.
17. See ref. 14 for more detailed methods and more application of in vivo bioluminescence assay.

References

1. Abendroth A, Arvin AM (1999) Varicella-zoster virus immune evasion. Immunol Rev 168:143–156
2. Gilden DH, Kleinschmidt-DeMasters BK, LaGuardia JJ, Mahalingam R, Cohrs RJ (2000) Neurologic complications of the reactivation of varicella-zoster virus. N Engl J Med 342:635–645
3. Arvin AM (2001) Varicella-zoster virus. In: Knipe DM, Howley PM (eds) Fields virology, vol 2. Lippincott Williams & Wilkins, Philadelphia, PA, pp 2731–2767
4. Davison AJ, Scott J (1986) The complete DNA sequence of varicella zoster virus. J Gen Virol 67:1759–1816
5. Cohen JI, Seidel KE (1993) Generation of varicella-zoster virus (VZV) and viral mutants from cosmid DNAs: VZV thymidylate synthetase is not essential for replication in vitro. Proc Natl Acad Sci USA 90:7376–7380
6. Mallory S, Sommer M, Arvin AM (1997) Mutational analysis of the role of glycoprotein I in varicella-zoster virus replication and its effects on glycoprotein E conformation and trafficking. J Virol 71:8279–8288
7. Niizuma T, Zerboni L, Sommer MH, Ito H, Hinchliffe S, Arvin AM (2003) Construction of varicella-zoster virus recombinants from P-Oka cosmids and demonstration that ORF65 protein is dispensable for infection of human skin and T cells in the SCID-hu mouse model. J Virol 77:6062–6065
8. Cohen JI, Straus SE, Arvin AM (2007) Varicella-zoster virus replication, pathogenesis, and management. In: Knipe DM, Howley PM (eds) Fields virology, vol 2. Lippincott Williams & Wilkins, Philadelphia, PA, pp 2773–2818
9. Nagaike K, Mori Y, Gomi Y, Yoshii H, Takahashi M, Wagner M, Koszinowski U, Yamanishi K (2004) Cloning of the varicella-zoster virus genome as an infectious bacterial artificial chromosome in *Escherichia coli*. Vaccine 22:4069–4074
10. Zhang Z, Rowe J, Wang W, Sommer M, Arvin A, Moffat J, Zhu H (2007) Genetic analysis of varicella zoster virus ORF0 to 4 using a novel luciferase bacterial artificial chromosome system. J Virol 81:9024–9033
11. Wang W, Patterson CE, Yang S, Zhu H (2004) Coupling generation of cytomegalovirus deletion mutants and amplification of viral BAC clones. J Virol Methods 121:137–143
12. Netterwald J, Yang S, Wang W, Ghanny S, Cody M, Soteropoulos P, Tian B, Dunn W, Liu F, Zhu H (2005) Two gamma interferon-activated site-like elements in the human cytomegalovirus major immediate-early promoter/enhancer are important for viral replication. J Virol 79:5035–5046
13. Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc Natl Acad Sci USA 97:5978–5983
14. Tang QY, Zhang Z, Zhu H (2010) Bioluminescence imaging for herpesvirus studies in vivo. In: Gluckman TR (ed) Herpesviridae: viral structure, life cycle and infections. Nova Science, Huntington, in press

Chapter 6

A Method for Rapid Genetic Integration into *Plasmodium falciparum* Utilizing Mycobacteriophage Bxb1 Integrase

Sophie H. Adjalley, Marcus C.S. Lee, and David A. Fidock

Abstract

Genetic manipulation of the human malaria parasite *Plasmodium falciparum* has presented substantial challenges for research efforts aimed at better understanding the complex biology of this highly virulent organism. The development of methods to perform gene disruption, allelic replacement or transgene expression has provided important insights into the function of parasite genes. However, genomic integration studies have been hindered by low transfection and recombination efficiencies, and are complicated by the propensity of this parasite to maintain episomal replicating plasmids. We have developed a fast and efficient site-specific system of integrative recombination into the *P. falciparum* genome, which is catalyzed by the mycobacteriophage Bxb1 serine integrase. This system has the advantage of providing greater genetic and phenotypic homogeneity within transgenic lines as compared to earlier methods based on episomal replication of plasmids. Herein, we present this methodology.

Key words: *Plasmodium falciparum*, Malaria, Bxb1 integrase, *attB* and *attP* sites, Integrative recombination, Transfection, Transgene expression

1. Introduction

The clinical and public health impacts of malaria are enormous, causing disease in over 500 million individuals and resulting in the death of approximately one million young children each year (1). In recent years, substantial efforts have been devoted to developing tools to genetically manipulate *Plasmodium falciparum*, the most lethal etiologic agent of malaria. Knockouts, allelic replacements, and transgene expression are now established techniques for investigating *Plasmodium* gene function (2). Procedures to manipulate the *P. falciparum* genome, however, have been hampered by the low efficiencies of transfection and

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_6,

recombination, as well as the persistence of self-replicating episomal concatemers, which limit the frequency of integration events into the parasite genome (2).

Integrative recombination constitutes an effective method to target exogenous DNA into specific genomic locations. For bacteriophages, integration into the bacterial host chromosome is typically mediated by tyrosine or serine integrases that catalyze site-specific recombination between phage *attP* and bacterial *attB* sites (3). These phage integrases, which mediate integration as well as excision reactions, can differ in their requirements for accessory proteins (4). Integration events are highly directional, thus excluding the possibility of spontaneous reversion. Serine integrases of the mycobacteriophage Bxb1 as well as of *Streptomyces* sp. phage ϕC31 have been extensively used for in vitro integration. In particular, the stability of the *attB*× *attP* locus generated by the ϕC31 system has been well documented in yeast and bacteria (5, 6).

Mycobacteriophage Bxb1 is a temperate phage that integrates its genome at an *attB* site located within the *Mycobacterium smegmatis groEL1* gene. Recombination occurs between a 48-bp *attP* and a 38-bp *attB* site, each of which contains interrupted inverted repeats that flank a conserved 5′-GT dinucleotide (7). The nonpalindromic nature of this central 5′-GT dinucleotide determines the directionality of integration. Unlike the tyrosine integrase systems, serine integrase-based recombination only requires the recombinase to efficiently catalyze the reaction, with no requirement for bacterial host cofactors, divalent cations, or DNA supercoiling (7). These properties make the serine integrase-mediated recombination system an ideal tool for targeted gene delivery in heterologous systems.

We have applied this serine integrase-driven recombination to *P. falciparum*, and find that transiently expressed mycobacteriophage Bxb1 integrase can efficiently catalyze recombination between an *attP*-containing plasmid and an *attB* site integrated into the parasite genome. This system was first established to achieve rapid integration of a *pfenr*-GFP gene fusion. *pfenr* (also known as *pffabI*) encodes *P. falciparum* enoyl acyl carrier protein ACP reductase, which is a component of the type II fatty acid synthesis pathway that functions in the parasite apicoplast. Microscopic studies localized the corresponding PfENR-GFP protein product to this intracellular organelle (8).

The integrase-based recombination offers several advantages over episomal-based transgene expression: namely, a higher proportion of fluorescent parasites when the fluorescently tagged transgene is integrated and a more uniform intensity of fluorescence (Fig. 1a–c). This provides for a more genetically and phenotypically homogeneous expression system as compared to episomal expression, where plasmid copy numbers and transgene phenotypes display higher variability (8, 9). This system has now

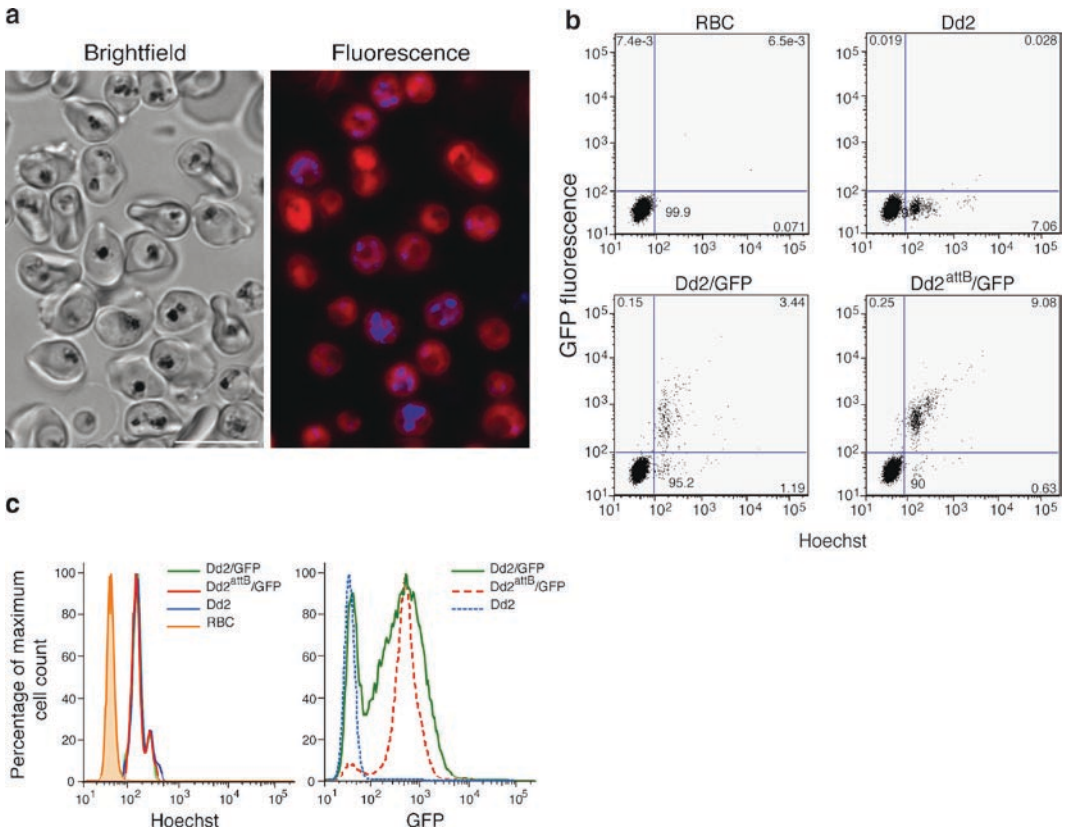


Fig. 1. Generation of homogeneous parasites by integrase-mediated *attB* × *attP* recombination. (**a**) Live cell imaging of Dd2^{attB} parasites expressing an mRFP-tagged cytosolic protein (PfVps4) encoded on a plasmid integrated at the *cg6-attB* locus. Bright field (*left panel)* and merged fluorescence images (*right panel*) showing mRFP-PfVps4 (*red*) and Hoechst 33342-stained nuclei (*blue*). Prior to imaging, infected RBCs were enriched by magnet-activated cell sorting (MACS, Miltenyi Biotec Inc., Auburn, CA) (18). Bar = 10 μm. (**b**) and (**c**) Flow cytometry analysis of parasites expressing GFP either from a plasmid integrated at the *cg6-attB* locus (Dd2^{attB}/GFP) or from episomes (Dd2/GFP). Parasites were incubated with Hoechst 33342 dye and analyzed at 16- to 20-h post-invasion. Cells were gated for nuclear staining to distinguish infected from uninfected RBCs. (**b**) GFP expression profiles show a higher percentage of GFP-expressing parasites in the integrant Dd2^{attB}/GFP population (94%) than in the episomal Dd2/GFP (74%). (**c**) Dd2^{attB}/GFP integrants show a unique peak of GFP expression with negligible variance, whereas Dd2 parasites expressing GFP from episomes display multiple peaks with a significantly larger variance. Panels (**b**) and (**c**) were derived from Fig. 4 in ref. 8 with the permission of the Nature Publishing Group.

become a standard tool in our lab to permit studies of transgene expression in phenotypically homogeneous parasite cultures.

Below, we describe the materials and methods of our integrase-based dual-plasmid co-transfection approach, which we now routinely utilize when conducting studies of transgenes expressed as a single copy from the *P. falciparum* genome. The simplicity of this integrase-based reaction, which can function without a requirement for exogenous host factors, makes it likely that such a system could be adapted to many heterologous cell systems.

2. Materials

2.1. Parasite Cultures

1. Dd2^{attB} and 3D7^{attB} parasite lines harboring an *attB* site integrated at the non-essential *cg6* locus. These lines are available from the Malaria Research and Reference Reagent Resource Center (MR4, Manassas, VA – www.malaria.mr4.org). We note that Dd2^{attB} propagates more rapidly and is our preferred host line.
2. RPMI 1640 with L-glutamine (Invitrogen, Carlsbad, CA), 25 mM N-2-hydroxyethylpiperazine-*N*-2-ethanesulfonic (HEPES; CalBiochem, San Diego, CA), 50 mg/L hypoxanthine (Sigma-Aldrich, St. Louis, MO) medium is supplemented with 0.5% Albumax II (Invitrogen), 0.25% Na_2CO_3 (Sigma-Aldrich), and 0.01 mg/mL gentamicin (Invitrogen) to constitute complete media (CM). Store at 4°C.
3. Human red blood cells (RBCs) and human sera are obtained from Interstate Blood Bank (Memphis, TN).

2.2. Parasite Transfection and Selection of Recombinant Lines

1. Plasmids: Integrase-expressing plasmid pINT, which contains a Neomycin selectable marker that confers G418 resistance; *attP*-containing plasmid pLN-ENR-GFP, which harbors a *pfenr-gfp* expression cassette as well as a *bsd* (blasticidin S-deaminase) selectable marker that confers resistance to blasticidin hydrochloride. pLN-ENR-GFP can be digested with *Afl*II and *Avr*II enzymes (New England Biolabs, Ipswich, MA) to replace the *pfenr-gfp* fusion with the gene of interest. Digestion with *Avr*II and *Bsi*WI (New England Biolabs) can also be performed to place the gene of interest in frame with *gfp*, whose product is positioned at the C-terminus of the corresponding fusion protein. Both the pINT and pLN-ENR-GFP plasmids are available from the Malaria Research and Reference Reagent Resource Center (MR4). Sequence files are available from GenBank (http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide) under the identifiers DQ813653 and DQ813654 for pLN-ENR-GFP and pINT, respectively. Plasmid maps are illustrated in Fig. 2a, b.
2. Cytomix (1×): 120 mM KCl, 0.2 mM $CaCl_2$, 2 mM EGTA, 10 mM $MgCl_2$, 25 mM HEPES, 5 mM K_2HPO_4, 5 mM KH_2PO_4; pH 7.6. Store at room temperature.
3. Gene-Pulser II (Bio-Rad, Hercules, CA) and 0.2-cm cuvettes (Bio-Rad).
4. Blasticidin HCl (Invitrogen) is dissolved in tissue culture water at 10 mg/mL, sterile-filtered and stored in aliquots at –80°C. Add to parasite cultures at a final concentration of 2.5 μg/mL.

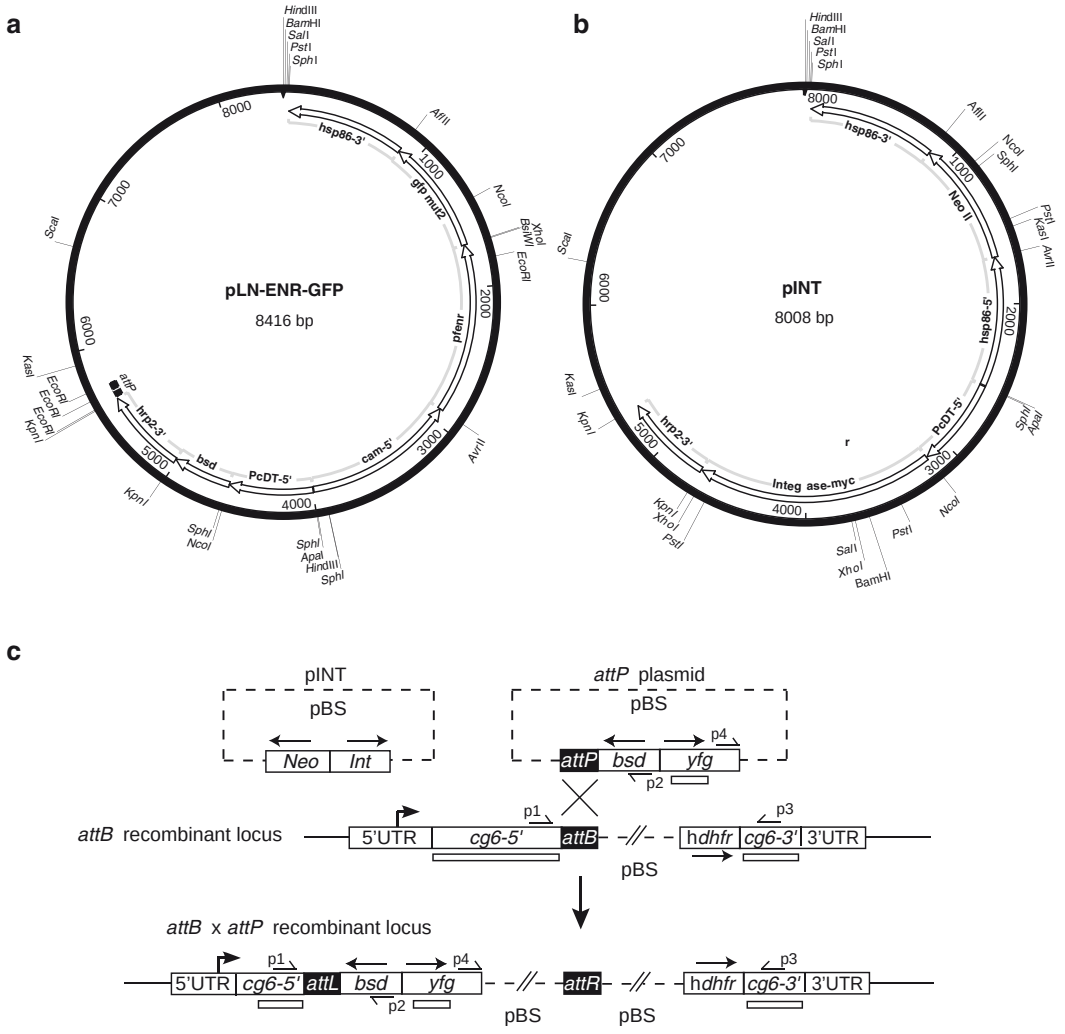


Fig. 2. Dual-plasmid approach for integrase-mediated *attB* × *attP* recombination. (**a**) Map of the *attP*-carrying plasmid pLN-ENR-GFP that can be used as a template for engineering the stable site-specific integration of a gene of interest. Digestion with *Avr*II and *Afl*II can be performed to replace the *enr-gfp* fusion by the gene of interest. Alternatively, digestion with *Avr*II and *Bsi*WI can place the gene of interest in frame with *gfp*. The plasmid carries the blasticidin S-deaminase (*bsd*) selectable marker. (**b**) Map of the integrase-expressing plasmid pINT, which harbors a neomycin selectable marker. The plasmid is usually maintained under G418 selection pressure only for the first 6 days following transfection. (**c**) Schematic representation of the *attB* × *attP* recombination allowing integration of a gene of interest (*yfg*) at the *cg6-attB* locus. Dd2^{attB} or 3D7^{attB} parasites are co-transfected with pINT and an *attP*-plasmid carrying the gene of interest, and selected with 2.5 μg/mL BSD, 2.5 nM WR99210 (to select for the human dihydrofolate reductase (hdhfr) marker in the *cg6-attB* locus) and 125 or 250 μg/mL G418 (for the Dd2^{attB} and 3D7^{attB} parasite lines respectively). Integrase expressed from the pINT plasmid catalyzes insertion of the *attP*-plasmid harboring the gene of interest into the *attB* site. PCR reactions using p1 + p2 and p3 + p4 primers as well as Southern blots analyzing digested-DNA hybridization to *cg6* or *yfg* probes (*designated as open boxes*), can be performed to confirm integration of the gene of interest at the *cg6-attB* locus. Sequences of the primers for PCR screening are as follows: p1: 5′-GAAAATATTATTACAAAGGGTGAGG, p2: 5′-TTAGCTAATTCGCTTGTAAG, p3: 5′-CTCTTCTACTCTTTCGAATTC, p4 is designed based on the sequence of the gene of interest. pBS refers to the pBluescript backbone. *UTR*: untranslated region. Panel (**c**) was adapted from Fig. 3 in ref. 8 with the permission of the Nature Publishing Group.

5. G418 (Cellgro, Manassas, CA) is dissolved in tissue culture water at 200 mg/mL, sterile-filtered and added to parasite cultures at a final concentration of either 125 µg/mL (Dd2 line) or 250 µg/mL (3D7 line).
6. WR99210 (Jacobus Pharmaceuticals, Princeton, NJ; molecular weight 394.35) is dissolved in CM or tissue culture water at 25.4 mM (10 mg/mL), sterile-filtered and stored in aliquots at −80°C. From this, we prepare working stocks of 25 µM in CM that we keep for up to 4 weeks at 4°C. Add to parasite cultures at a final concentration of 2.5 nM. This antifolate inhibits *Plasmodium* dihydrofolate reductase and selects for recombinant parasites expressing human dihydrofolate reductase (9, 10).

2.3. Plasmid Rescue

1. 0.1-cm cuvettes (Bio-Rad).
2. XL1-Blue (or other suitable) Electroporation-Competent Cells (Stratagene, La Jolla, CA).
3. SOC media (Sigma-Aldrich).
4. Luria Broth Base (Becton, Dickinson and Company, Franklin Lakes, NJ) and adequate antibiotic. Both the pINT and the *attP*-containing plasmids possess a carbenicillin resistance marker, which we select with 50 µg/mL carbenicillin disodium (Sigma-Aldrich).

2.4. Cloning Parasites by Limiting Dilution

1. CritSpin Microhematocrit Centrifuge System (StatSpin, Westwood, MA).
2. Malstat reagent: 100 mM Tris–HCl, 0.2 M L-lactic acid (Sigma-Aldrich), 0.2% (v/v) Triton X-100, and pH 9.1. Store at 4°C.
3. 3-Acetylpyridine adenine dinucleotide (APAD, Sigma-Aldrich; molecular weight 662.44) is dissolved in sterile water at 50 mg/mL, stored at −20°C, and added to Malstat reagent at 10 µM final concentration.
4. Nitroblue tetrazolium (NBT, Promega, Madison, WI) 50 mg/mL is stored at −20°C and used at a final concentration of 200 µg/mL.
5. Diaphorase (Sigma-Aldrich) is dissolved in sterile water at 8 mg/mL, stored at −20°C and added to Malstat reagent at a final concentration of 64 µg/mL.

2.5. Extraction of P. falciparum Genomic DNA from Cloned Parasites for Subsequent PCR Screening

1. Phosphate buffered saline (PBS): A 10× stock solution is prepared with 1.37 M NaCl, 26.8 mM KCl, 43 mM NaH_2PO_4, 14.7 mM KH_2PO_4, and pH 7.4. Autoclave, then store at room temperature. Use at 1× concentration by diluting with double-distilled water (ddH_2O).

2. Saponin (Sigma-Aldrich) is dissolved in ddH_2O at 10% (w/v) and used at a final concentration of 0.1% (w/v).
3. Lysis Buffer: 0.5 M Sucrose, 1 M Tris–Cl pH 7.6, 1 M $MgCl_2$, and 1% Triton X-100. Autoclave and store at 4°C.
4. Down Scale Prep (DSP) Buffer (4×): 1 M Tris–Cl pH 8.0, 1 M KCl, and 1 M $MgCl_2$. To prepare working solution, dilute to 1× and include Tween 20 and Proteinase K (Sigma-Aldrich) at final concentrations of 0.5% and 2.5 µg/mL, respectively. Our stock of Proteinase K is dissolved in sterile water at 10 mg/mL and stored at −20°C. Note that the working solution has to be made fresh each time.
5. DNeasy Blood and Tissue kits (Qiagen, Valencia, CA).

3. Methods

3.1. Overall Procedure

Parasites to be transfected are thawed from cryopreserved stocks and placed into culture for 5–7 days to obtain sufficient parasites (see below). During this period, DNA is prepared from a small aliquot (1 mL at 3–6% parasitemia, with a hematocrit of ~3%; see below) and subjected to microsatellite typing (11, 12) to confirm the identity of the strains to be used (see Subheading 3.2, step 2 and Note 1).

Transfection involves the use of two plasmids, one expressing the gene of interest (which can be prepared by modifying the *attP*-containing pLN-ENR-GFP plasmid to replace the *pfenr* coding sequence with the desired transgene; Fig. 2a) and the other expressing the integrase (such as pINT, Fig. 2b). Co-transfection of these plasmids into an *attB* line, such as $Dd2^{attB}$ or $3D7^{attB}$, results in transgene integration into the *cg6-attB* site (Fig. 2c).

Recombination mediated by mycobacteriophage Bxb1 integrase usually yields a predominant population of integrants, which can be assessed by PCR screening and plasmid rescue experiments. However, we have also experienced instances where integrative recombination did not occur in a uniform manner, but instead led to a mixed population of parasites that either underwent integration (integrants) or continued to replicate the plasmids episomally (episomal replicants). This has been observed with $3D7^{attB}$, which propagates more slowly than $Dd2^{attB}$, leading us to use the latter whenever there is a choice. In such a case of mixed parasite population, it becomes essential to obtain clones by limiting dilution so as to isolate the expected integrants and obtain a genetically homogeneous population. Clones are detected by assaying for parasite-specific lactate dehydrogenase (LDH) activity (see Subheading 3.5). The analysis of the recombinant

locus is typically performed by Southern Blot to confirm the predicted genome integration and evaluate the number of plasmid copies that can integrate in tandem.

3.2. Preparation of Cultures for Transfection

1. *P. falciparum* Dd2^{attB} and 3D7^{attB} cultures are propagated in human RBCs at a 3% hematocrit in CM under reduced oxygen conditions (90% N_2, 5% O_2, and 5% CO_2) as described (10, 13). These lines have been engineered by the introduction of an acceptor *attB* site (from *M. smegmatis*) into the non-essential *cg6* locus. This recombinant locus is maintained by continuous selection with 2.5 nM WR99210 (9).
2. In order to verify, by microsatellite typing, the identity of the lines to be transfected, genomic DNA is extracted after saponin lysis of parasite-infected RBCs. As we generally culture parasites in 6-well plates at a volume of 5 mL per well, we usually perform the extraction from a similar volume of culture. Cells are transferred into a 15-mL sterile conical tube and spun for 3 min at 650 × *g* in a tabletop centrifuge. The supernatant is aspirated and the cell pellet resuspended in 5 mL of 0.1% saponin in PBS (1×) for 5 min at room temperature, then spun at 3,150 × *g* for 5 min. The parasite pellet is washed twice in PBS before proceeding to genomic DNA extraction using a DNeasy Blood & Tissue Kit (Qiagen). We note that this kit contains two types of lysis buffer. We typically lyse parasites with the AL buffer. Also, we elute DNA from the Qiagen column in a smaller volume than suggested in the user manual, i.e. usually twice 50 µL.

3.3. Transfection and Selection

Transfection procedures are carried out in Cytomix and the final volume of infected cells and plasmids in the buffer should be 450 µL.

1. 50–100 µg of each plasmid are required per transfection. These have to be equilibrated in Cytomix beforehand to reach a final volume of 150 µL per transfection. When preparing plasmids, the DNA pellet can be either directly dissolved in 1× Cytomix, or resuspended in sterile water or 1× TE buffer (10 mM Tris–HCl, 1 mM EDTA pH8.0) and diluted 1:1 in 2× Cytomix.
2. The day before transfecting, *P. falciparum* cultures are diluted in freshly drawn, uninfected RBCs to obtain predominantly ring-stage parasites at 5–8% parasitemia on the following day.
3. On the day of transfection, parasites are first fed with fresh medium and smeared to confirm that the culture contains predominantly ring-stage parasites at greater than 5% parasitemia. Parasite-infected cells are then harvested (2.5 mL per

transfection) and centrifuged in sterile conical tubes for 3–4 min at 640 × *g* at room temperature.

4. Cell pellets are washed once in 1 × Cytomix (2.5 mL per transfection) and then resuspended in 300 μL of 1 × Cytomix.
5. Plasmid DNA and infected cells are then mixed together and transferred to electroporation cuvettes.
6. Settings for electroporation of infected RBCs are the following: voltage of 0.31 kV and capacitance of 950 μF (when set on maximum capacitance).
7. Immediately after electroporation, 1 mL of complete media (CM) is gently and aseptically added to each cuvette. Once all electroporations are completed, the contents of the cuvettes are transferred to 6-well plates that have been prepared beforehand (each well contains 3.5 mL of CM with 4% hematocrit and receives the contents of one transfection). Cuvettes are then washed with another 1 mL of CM that is added to the well, and the plates are placed at 37°C.
8. Approximately 1–4 h after transfection, cultures are transferred to sterile conical tubes and spun for 4 min at 450 × *g* at room temperature. Pellets are then gently resuspended in 5 mL of complete media (with WR99210) and returned to the 6-well plate.
9. Selection with the agents BSD and G418 (14, 15) starts the day after transfection (i.e. day 1). Cultures are smeared that day to check the post-electroporation parasitemia, which is typically between 1 and 3% (see Note 2).
10. Cultures are fed every day for the first 6 days, then every other day, with fresh medium containing the selection agents. Drug-mediated clearance of live parasites is microscopically confirmed on day 6. Integrase-mediated recombination is thought to be a rapid event, and it is assumed that maintenance (using G418 drug pressure) of the pINT plasmid expressing the integrase is only required for the first 5–6 days. After that time, G418 can be removed.
11. On day 6, a 100-μL aliquot of fresh RBCs (50% hematocrit) is added to each well. On day 12, 30–40% of each culture is replaced with fresh RBCs and media by diluting the culture 3:5. This dilution is essential as lysis of aging RBCs can occur after a prolonged period of culture. Transfections are then diluted 3:5 every 6 days and monitored by Giemsa stain every 4–5 days for reemergence of the transfected population. In our experience, Dd2^{attB}/ENR-GFP parasites (generated by transfecting the Dd2^{attB} line with the pLN-ENR-GFP plasmid) appeared within 18 days of electroporation. In general, transfected parasites that have undergone *attB* × *attP* recombination are observed 15–25 days post-electroporation.

In most experiments, the recovered parasites are predominantly integrants; however, it is not unusual to detect a subpopulation of parasites containing episomal replicating plasmids in the bulk culture.

3.4. Plasmid Rescue

This procedure can be performed to help discriminate between parasites expressing the transgene of interest either from an integrated plasmid or from episomes.

1. Proceed to genomic DNA extraction from transfected parasite cultures as described in Subheading 3.2, step 2.
2. 100 ng of genomic DNA is used to transform 50 μL of *Escherichia coli* competent cells (e.g. XL1) by electroporation. Settings for bacteria electroporation are as follows: voltage of 1.8 kV, resistance of 200 Ω, and capacitance of 25 μF.
3. Electroporated bacteria are recovered in 500 μL of SOC media (2% bacto-tryptone, 0.5% bacto-yeast extract, 9 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose, and pH 7.0) at 37°C for 1 h. 250 μL of recovered bacteria culture are plated on LB/agar with the appropriate antibiotic. We use 50 μg/mL carbenicillin, which selects for expression of beta-lactamase by the pLN-ENR-GFP and pINT plasmids and their derivatives. As an alternative, 100 μg/mL ampicillin can also be used, although this tends to produce more resistant satellite colonies upon extended bacterial culture. Quantification of the rescue efficiency provides a reasonably accurate measure of the extent to which the *attP*-containing plasmids have integrated into the *attB* site, versus their episomal replication (see Note 3).

3.5. Limiting Dilution of Parasite Clones and the Lactate Dehydrogenase Assay

As mentioned earlier, although the Bxb1 integrase-mediated recombination system is highly efficient and generally yields a predominant population of recombinants, detection of a subpopulation of parasites that express the transgene of interest as an episome is not uncommon. In the event that a highly homogeneous population of transgenic parasites is required, isolating integrant parasites from those that harbor episome-replicating plasmids can be achieved by limiting dilution cloning.

1. Before proceeding, the parasitemia of the culture has to be assessed accurately, both by Giemsa-stained smear (providing the percentage of infected RBCs), and by determining the hematocrit of the culture (indicating the percentage of RBCs per mL of culture). For the latter, we use a CritSpin Microhematocrit Centrifuge system or a standard hemocytometer. We typically consider the number of RBCs to be around 10^{10} per mL of packed cells. Consequently, 3% hematocrit, for instance, corresponds to 3×10^8 RBCs/mL.

2. Once the exact parasitemia is determined, cultures are diluted so as to inoculate the equivalent of 0.5–0.8 parasites per well. This would be predicted to yield approximately 50–80 parasitized wells per plate, but in general we tend to recover about 70% of the expected numbers of clones (i.e. approximately 30–50 clones per plate). We generally incubate 200 μL of parasite culture per well in 96-well microtiter plates with a hematocrit of 1.8%, and always use fresh blood to prevent any RBC lysis during the course of the procedure.
3. Cultures are fed once a week with a mix of media and fresh RBCs at 0.4% hematocrit. This provides additional fresh RBCs for parasite reinvasion. If parasites are cultured with medium that is supplemented with 10% human serum instead of 0.5% Albumax, cultures are fed twice a week.
4. Growth of parasite clones is assessed after 2–3 weeks using an LDH assay based on the Malstat reagent (16, 17) (see Note 4).
   (a) One 96-well plate requires 12 mL of Malstat reagent, to which 158.4 μL of APAD (50 mg/mL) is added. 6 mL of this mixture is then transferred to a 15-mL conical tube (tube 1) that is kept on ice and wrapped in aluminum foil. To this, we add 24 μL of NBT (50 mg/mL). The other 6 mL of Malstat reagent/APAD is transferred to another 15-mL conical tube (tube 2), also kept on ice and wrapped in aluminum foil. To that mix, we add 48 μL of diaphorase (8 mg/mL).
   (b) 20 μL of parasite-infected cell suspension from each well is then transferred into a new round-bottom 96-well microtiter plate.
   (c) Tubes 1 and 2 are mixed together and 120 μL of this mixed Malstat solution is added to each well using a 12-well pipetting device.
   (d) The plate is then placed in the dark (or wrapped in aluminum foil) for 30–45 min.
   (e) The purple color associated with parasitized RBCs can be detected by measuring the absorbance at 650 nm in a plate reader. If the parasitemia is high enough, the color change in positive wells can be clearly distinguished visually from the negative wells, appearing more purple than uninfected wells. We also note that this assay can distinguish the level of parasitemia in positive wells. As a rule of thumb, we tend to avoid the wells that have the strongest LDH activity as we predict that these result from multiply infected RBCs, which could therefore represent a mixture of genetically distinct parasites (see Note 5).

3.6. Extraction of P. falciparum Genomic DNA from Cloned Parasites

Once positive clones have been detected by the LDH assay and have reached an adequate parasitemia (1–2%), PCR screening can be performed to confirm the isolation of recombinant parasites. Below we list a protocol that we use to extract genomic DNA directly from the parasites that have been cloned and cultured in 96-well microtiter plates.

1. A 96-well PCR plate is prepared by aliquoting 15 μL of 1% saponin (diluted from a 10% stock in 1× PBS) in each well.
2. 150 μL of each parasite culture is transferred to the corresponding well in the saponin plate. 150 μL of fresh media with RBCs is added to the remaining volume of culture, such that parasites have been diluted 1:4, and the culture plate is returned to the incubator. The saponin plate is incubated at room temperature for 5 min and spun at 3,150 × *g* in a tabletop centrifuge using microplate adaptors.
3. Supernatants are discarded by inverting the plate and the remaining pellets are washed by resuspending in 100 μL of PBS. The plate is centrifuged once again and the supernatants are removed, as above.
4. Pellets are then resuspended in 40 μL of lysis buffer. The plate is centrifuged and the supernatants discarded, as before. The step is repeated once.
5. Finally, pellets are resuspended in 40 μL of DSP working solution and heated at 50°C for 2 h, followed by 10 min at 95°C.
6. After cooling the plate at 4°C for 10 min, 4 μL can be used as a template for PCR, and the remainder is stored at –20°C.

The PCR strategy is generally designed to confirm the generation of *attL* and *attR* flanking sites upon integrase-mediated recombination between the genomic *attB* site and the plasmid *attP* site, as well as the absence of the wild-type locus and of plasmid episomes. Primers used for these PCR screenings are sketched in Fig. 2c (see figure legend for primer sequences).

4. Notes

1. We recommend, if applicable, verifying the phenotype of the pre-transfection culture (by performing drug assays, adhesion assays, etc.), when the purpose of the genetic manipulation is to modify this phenotype. It is important that a new stock of parasites be cryopreserved at the time the transfection is performed, to provide the proper control for future analyses of the recombinant lines.

2. If the parasitemia is below 1%, it might be necessary to repeat the transfection, as the probability of having transfected parasites is reduced. If the parasitemia is above 3%, cultures have to be closely monitored for the first few days following drug pressure, and culture dilution might be required to prevent parasites dying from overgrowth rather than drug selection.
3. The rescue efficiency, considered as the number of cfu (colony-forming units) per µg of genomic DNA, correlates with the number of episomal replicating plasmids. We generally estimate that integrated plasmids will produce 10^2 cfu/µg DNA (because of a subpopulation of integrated plasmids "looping out" of the recombinant locus (9)), whereas episomal replicating plasmids will likely generate 5×10^3 to 10^5 cfu/µg DNA.
4. This assay relies on the specific detection of parasite LDH and is based on the principle that the *P. falciparum* LDH enzyme can substitute APAD (3-acetylpyridine adenine dinucleotide) for NAD (nicotinamide adenine dinucleotide) as a cofactor, whereas the human erythrocyte LDH does not have this ability.
5. To reduce the prevalence of multiply infected RBCs, the cultures can be incubated for 48 h prior to cloning with gentle rocking to maintain the RBCs in suspension. If after 3 weeks of culture, parasites are still below the detection threshold or at a density not sufficient to perform parasite extraction, cultures can be diluted 1:2 with fresh RBCs to avoid any cell lysis and cultured for an extra week.

Acknowledgments

We thank Heather Schneider for help with the figures. This work was supported by P01 AI060342 (PI. James Sacchettini).

References

1. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI (2005) The global distribution of clinical episodes of *Plasmodium falciparum* malaria. Nature 434:214–217
2. Crabb BS, Rug M, Gilberger TW, Thompson JK, Triglia T, Maier AG et al (2004) Transfection of the human malaria parasite *Plasmodium falciparum*. Methods Mol Biol 270:263–276
3. Groth AC, Calos MP (2004) Phage integrases: biology and applications. J Mol Biol 335:667–678
4. Ghosh P, Wasil LR, Hatfull GF (2006) Control of phage Bxb1 excision by a novel recombination directionality factor. PLoS Biol 4:e186
5. Thorpe HM, Wilson SE, Smith MC (2000) Control of directionality in the site-specific recombination system of the *Streptomyces* phage phiC31. Mol Microbiol 38:232–241
6. Thomason LC, Calendar R, Ow DW (2001) Gene insertion and replacement in *Schizosaccharomyces pombe* mediated by the *Streptomyces* bacteriophage phiC31 site-specific recombination system. Mol Genet Genomics 265:1031–1038
7. Ghosh P, Kim AI, Hatfull GF (2003) The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of *attP* and *attB*. Mol Cell 12:1101–1111

8. Nkrumah LJ, Muhle RA, Moura PA, Ghosh P, Hatfull GF, Jacobs WR Jr et al (2006) Efficient site-specific integration in *Plasmodium falciparum* chromosomes mediated by mycobacteriophage Bxb1 integrase. Nat Methods 3:615–621
9. Fidock DA, Wellems TE (1997) Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. Proc Natl Acad Sci USA 94:10931–10936
10. Fidock DA, Nomura T, Wellems TE (1998) Cycloguanil and its parent compound proguanil demonstrate distinct activities against *Plasmodium falciparum* malaria parasites transformed with human dihydrofolate reductase. Mol Pharmacol 54:1140–1147
11. Su XZ, Wellems TE (1997) *Plasmodium falciparum*: a rapid DNA fingerprinting method using microsatellite sequences within *var* clusters. Exp Parasitol 86:235–236
12. Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R et al (2000) Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Mol Biol Evol 17:1467–1482
13. Trager W, Jensen JB (1976) Human malaria parasites in continuous culture. Science 193:673–675
14. Mamoun CB, Gluzman IY, Goyard S, Beverley SM, Goldberg DE (1999) A set of independent selectable markers for transfection of the human malaria parasite *Plasmodium falciparum*. Proc Natl Acad Sci USA 96:8716–8720
15. Wang P, Wang Q, Sims PF, Hyde JE (2002) Rapid positive selection of stable integrants following transfection of *Plasmodium falciparum*. Mol Biochem Parasitol 123:1–10
16. Makler MT, Hinrichs DJ (1993) Measurement of the lactate dehydrogenase activity of *Plasmodium falciparum* as an assessment of parasitemia. Am J Trop Med Hyg 48:205–210
17. Goodyer ID, Taraschi TF (1997) *Plasmodium falciparum*: a simple, rapid method for detecting parasite clones in microtiter plates. Exp Parasitol 86:158–160
18. Staalsoe T, Giha HA, Dodoo D, Theander TG, Hviid L (1999) Detection of antibodies to variant antigens on *Plasmodium falciparum*-infected erythrocytes by flow cytometry. Cytometry 35:329–336

Part II

PCR Mutagenesis

Chapter 7

Random Mutagenesis by Error-Prone PCR

Elizabeth O. McCullum, Berea A.R. Williams, Jinglei Zhang, and John C. Chaput

Abstract

In vitro selection coupled with directed evolution represents a powerful method for generating nucleic acids and proteins with desired functional properties. Creating high-quality libraries of random sequences is an important step in this process as it allows variants of individual molecules to be generated from a single-parent sequence. These libraries are then screened for individual molecules with interesting, and sometimes very rare, phenotypes. Here, we describe a general method to introduce random nucleotide mutations into a parent sequence that takes advantage of the polymerase chain reaction (PCR). This protocol reduces mutational bias often associated with error-prone PCR methods and allows the experimenter to control the degree of mutagenesis by controlling the number of gene-doubling events that occur in the PCR reaction. The error-prone PCR method described here was used to optimize a de novo evolved protein for improved folding stability, solubility, and ligand-binding affinity.

Key words: Error-prone PCR, *Taq* DNA polymerase, Directed evolution

1. Introduction

Directed evolution is a powerful approach for generating synthetic molecules with new chemical and physical properties (1, 2). This technique, also referred to as *test-tube evolution*, mimics the principles of Darwinian evolution by imposing a selective pressure on a large population of molecules so that sequences with certain desirable properties increase in abundance. The advantage of directed evolution over rational design is that no prior structural or mechanistic information is required for the selection to be successful. Directed evolution has been used to study many fundamental and practical problems in chemistry and biology (3).

The power of directed evolution lies in the ability to rapidly search a large combination of sequences for rare molecules that

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_7,

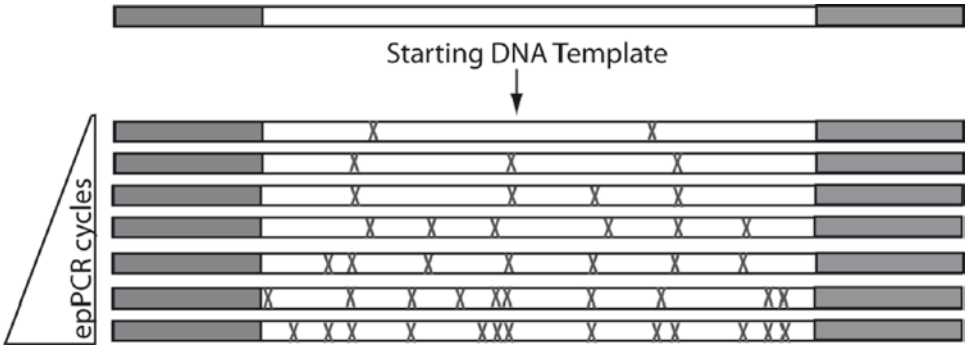


Fig. 1. Random mutagenesis by error-prone PCR. Mutations are randomly inserted into the DNA sequence under conditions that reduce the fidelity of *Taq* DNA polymerase. The number of mutations increase with the number of gene duplication events. PCR primer binding sites are denoted at the 5′ and 3′ of the DNA sequence and random point mutations are given as "X".

possess specific predetermined functions. Creating high-quality libraries is an important part of this process as the ability to go "from here to there" depends entirely on whether or not such molecules exist in the starting pool. Although there are many ways to introduce genetic diversity into a parent sequence (4), error-prone PCR (Fig. 1) is the most common method for creating a combinatorial library based on a single gene (5). This is due to the simplicity of the technique and the fact that most selection experiments aim to identify a small number of mutations that lead to improved stability or activity.

Error-prone PCR is typically performed using conditions that reduce the fidelity of *Taq* DNA polymerase during DNA synthesis. In this technique, the region of the gene undergoing mutagenesis is defined by the location of upstream and downstream PCR primer-binding sites and the number of gene doublings controls the degree of mutagenesis. In its original description by Leung and coworkers, the standard PCR protocol was modified to include: (1) increased concentration of *Taq* DNA polymerase; (2) increased polymerase extension time; (3) increased concentration of $MgCl_2$ ions; (4) increased concentration of dNTP substrates; and (5) the reaction was supplemented with $MnCl_2$ ions (5). Under these conditions, the rate of random mutagenesis is ~2% per nucleotide position per PCR reaction. Libraries produced by this method contain a large number of A→G and T→C transitions that bias the resulting sequences toward high GC content. To overcome this limitation, Cadwell and Joyce developed a modified PCR protocol that used an unbalanced ratio of nucleotides to minimize mutational bias in the amplified sequences (6). This technique provides an overall error rate of ~0.66% per nucleotide per PCR reaction, and results in sequences with no noticeable amplification bias. The protocol described below is an extension of the Leung and Joyce method and includes

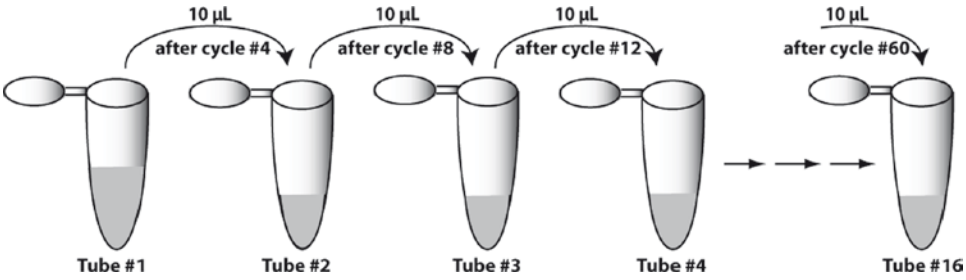


Fig. 2. Dilution and pooling technique used in epPCR. The current protocol uses a series of dilution and amplification steps to generate a mutagenic library that contains a range of single-nucleotide point mutations. After every four cycles of PCR amplification, a small portion of the PCR reaction mixture is transferred to a fresh tube and the process is continued for a total of 64 cycles of PCR or 16 serial transfer steps. The reaction products are then analyzed and pooled to create a random library for directed evolution.

several serial dilution steps (Fig. 2) that enable the experimenter to control the level of mutagenesis incorporated into the pool (7). In contrast to previous error-prone PCR methods, where a small amount of template is used as the basis for a single PCR reaction, the current protocol requires a larger amount of starting template and several serial dilution steps in which a portion of the amplified material (~10%) is successively transferred after every fourth amplification cycle to a fresh PCR reaction. Consequently, it is very easy to generate pools of variants with increasing degrees of mutations while simultaneously avoiding the PCR saturation problem. When all 16 serial dilution steps are used, this technique produces an average error rate of ~3.5% per nucleotide per PCR reaction; however, it is important to note that this number can vary between different templates (7).

We routinely use the error-prone PCR method described here to introduce random single-nucleotide mutations into synthetic genes identified by in vitro selection. In a recent example, we generated a library of random mutations from a gene that encoded an entirely synthetic, man-made protein that was previously selected from a pool of unbiased random sequences (8). We then used this library to select protein variants that enhanced the folding stability, solubility, and ligand-binding affinity of the parent gene. Following six rounds of mRNA display-based in vitro selection, we were able to identify two amino acid substitutions that appeared in almost all of the sequenced clones. These two single-point mutations, which would have been difficult, if not impossible, to identify by rational design, transformed our de novo evolved protein into a well-folded protein whose structure was solved by X-ray crystallography. This simple example demonstrates that subtle mutations can be easily identified when high-quality pools are subjected to the powers of in vitro selection and directed evolution.

2. Materials

2.1. Biological and Chemical Materials

1. DNA primers for PCR amplification (see Notes 1 and 2).
2. DNA template (either linear or plasmid).
3. Four separate solutions of 20 mM dATP (USB), TTP (Sigma), dCTP (Sigma) and dGTP (Sigma).
4. *Taq* DNA polymerase (Invitrogen), stored at −20°C (see Note 3).
5. 10× PCR buffer: 100 mM Tris–HCl, pH 8.3, 15 mM $MgCl_2$, and 500 mM KCl.
6. 1 M $MgCl_2$ (Fisher).
7. 50 mM $MnCl_2$ (Fisher) (see Note 4).
8. Agarose (EMD).
9. 10× TBE buffer: 1 M Tris base, 1 M boric acid, and 20 mM EDTA, pH 8.0.
10. 10 mg/mL ethidium bromide (US Biological) (see Note 5).
11. Loading buffer: 10 mM Tris–HCl, pH 7.6, 0.15% orange G, 0.03% xylene cyanol FF, 60% glycerol, 60 mM EDTA.
12. DNA mass ladder for quantification (Invitrogen).
13. Gel purification kit (Qiagen).

2.2. Equipment

1. Thermocycler PCR machine (Eppendorf).
2. Agarose gel electrophoresis apparatus.

3. Method (See Note 6)

1. Prepare a stock solution of the DNA template (~50–100 ng/μL) in water.
2. Label 16 thin-walled PCR tubes as “reactions 1 through 16”.
3. Combine the reagents listed in Table 1 in a conical vial and label it “PCR reaction mixture”.
4. Dispense the PCR reaction mixture from step 3 into tubes 1 through 16 by adding 96 μL to tube 1 and add 88 μL to tubes 2–16.
5. Add 2 μL of the DNA template to tube 1.
6. Place tube 1 in the thermocycler and start the PCR program (Table 2).

Table 1
Reagent mixture for error-prone PCR. Reagents in the list should be combined in a conical vial

PCR reagent mixture

Reagents	Stock concentration	Volume	Final concentration
Forward primer	100 µM	15 µL	1 µM
Reverse primer	100 µM	15 µL	1 µM
dCTP & dTTP	20 mM	75 µL/ea	1 mM
dATP & dGTP	20 mM	15 µL/ea	0.2 mM
PCR buffer w/Mg^{2+}	10×	150 µL	1×
$MgCl_2$	1 M	8 µL	~5.5 mM
Nanopure water		1,098 µL	
Final volume		1,491 µL	

Table 2
Thermocycler program for error-prone PCR. Prepare the thermocylcer program as described in the table. Annealing temperatures will vary depending on the PCR primers used in error-prone PCR

PCR thermocycler program[a,b]

	Step	Temperature	Duration
1	Denaturation	94°C	1 min
2	Annealing	60°C	1 min
3	Extension	72°C	3 min
4	Product storage	4°C	End

[a] The annealing temperature may vary between different PCR primers
[b] Repeat steps 1–3 for 64 cycles

7. Once the PCR program has reached the annealing temperature, add 1 µL of freshly prepared $MnCl_2$ solution and 1 µL of *Taq* DNA polymerase to the PCR reaction tube.
8. Perform 4 cycles of PCR amplification using the hot start procedure (see Note 3).

9. Remove the PCR tube from the thermocycler.
10. Place the tube on ice.
11. Transfer 10 µL of the PCR reaction from tube 1 to tube 2.
12. Place tube 2 in the thermocycler and start the PCR program.
13. Once the PCR program has reached the annealing temperature, add 1 µL of freshly prepared $MnCl_2$ and 1 µL of *Taq* DNA polymerase to the PCR reaction tube.
14. Perform 4 cycles of PCR amplification using the hot start procedure.
15. Remove the PCR tube from the thermocycler.
16. Place the tube on ice.
17. Repeat steps 11 through 16 using tubes 3 through 16 to create a mutagenic library by serial dilution amplification (Fig. 2).
18. Verify the quality of each PCR reaction by agarose gel electrophoresis (see Note 7).
19. Combine 50 µL of each PCR reaction into a single tube and store on ice; this is your DNA library (see Note 8).
20. Purify the DNA library by agarose gel electrophoresis.
21. Recover the DNA library using a gel purification kit.
22. Quantify the DNA library by agarose gel electrophoresis (see Note 9).

4. Notes

1. PCR primers should be designed to be noncomplementary and have similar melting temperatures. The web site given below can be used to design optimal primer sequences: http:// bioweb.uwlax.edu/GenWeb/Molecular/seq_anal/primer_design/primer_design.htm
2. Optimal PCR conditions should be determined prior to the use of this mutagenic PCR protocol. The doubling efficiency for the normal PCR reaction should be ~1.7–1.9 per PCR cycle (7) and can be analyzed by running the PCR product after each cycle on an agarose gel.
3. *Taq* DNA polymerase should be stored at –20°C and kept on ice or in a frozen metal block when taken out of the freezer. The hot start procedure refers to adding the *Taq* DNA polymerase to the PCR reaction mixture after the mixture has reached the annealing temperature. Either standard or "hot start" *Taq* DNA polymerase will work in the described method.

4. It is important to prepare a fresh $MnCl_2$ solution for the error-prone PCR experiment. This solution should be stored on ice and combined with the PCR reaction mixture at the start of each serial dilution.
5. Ethidium bromide should be stored in the dark or kept in a dark-colored vial.
6. The method described here was optimized for a 400-nucleotide gene. We highly recommend that experimenters read ref. 7 for information on the predicted average number of doublings and mutations expected as a function of template length.
7. Agarose gel electrophoresis should be performed by standard methods. A DNA ladder with an appropriate base pair range should be used to identify and estimate the quantity of the PCR product.
8. It is recommended to save a portion of the PCR product from each tube for further analysis or amplification.
9. The concentration of the DNA product can be determined by comparing the intensity of the product bands to the intensity of the bands from the DNA quantification ladder from commercial sources.

References

1. Wilson DS, Szostak JW (1999) *In vitro* selection of functional nucleic acids. Annu Rev Biochem 68:611–647
2. Arnold FH, Georgiou G (2003) Directed enzyme evolution: screening and selection methods. Humana Press, New Jersey
3. Chaput JC, Woodbury NW, Stearns LA, Williams BAR (2008) Creating protein biocatalysts as tools for future industrial applications. Expert Opin Biol Ther 8:1087–1098
4. Arnold FH, Georgiou G (2003) Directed evolution library creation: methods and protocols. Humana Press, New Jersey
5. Leung DW, Chen E, Goeddel DV (1989) A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Technique 1: 11–15
6. Cadwell RC, Joyce GF (1992) Randomization of genes by PCR mutagenesis. PCR Methods Appl 2:28–33
7. Wilson DS, Keefe AD (2000) Random mutagenesis by PCR. Current protocols in molecular biology. Wiley, New Jersey
8. Smith MD, Rosenow MA, Wang M, Allen JP, Szostak JW, Chaput JC (2007) Structural insight into the evolution of a non-biological protein: importance of surface residues in protein fold optimization. PLoS ONE 2:e467

Chapter 8

A Rapid and Versatile PCR-Based Site-Directed Mutagenesis Protocol for Generation of Mutations Along the Entire Length of a Cloned cDNA

Vincent Dammai

Abstract

Deciphering protein function is a major challenge in modern biology and continues to remain at the frontier of investigations into the molecular basis of cell behavior. With the explosion in our bioinformatics knowledge base and the now widespread use of associated software tools and database resources, we have an enormous logistic capability to identify protein domains of interest and the compelling desire to introduce mutations within these sequences in order to ultimately understand the functional aspects of a given protein and/or test its therapeutic applications. Faced with this ultimate task, a quick and efficient means to introduce desired mutations anywhere along the protein length is necessary as a first step. Here, HIF1α and HIF2α are used as examples to demonstrate the simplicity, speed, and versatility of the PCR-based mutagenesis method.

Key words: PCR, Mutagenesis, Site-directed mutagenesis, Pfu DNA polymerase, T4 DNA ligase, Chimeric protein, Domain analysis, Deletions, Duplications, Point mutations, HIF, Hif-1α, Hif-2α

1. Introduction

The heterodimeric Hypoxia-Inducible Factor (HIF) contains α- and β-subunits. The β-subunit (also called ARNT) is common to HIFs, but the α-subunit varies, namely containing Hif1α, Hif2α, or HIf3α, respectively (1). Hifα and Hifβ dimers constitute an active transcription factor complex that activates transcription from a broad range of genes containing HRE (hypoxia responsive elements). Although Hifα subunits are continually made, they are also simultaneously rapidly degraded under normal oxygen conditions through a von Hippel–Lindau protein (VHL)-dependent pathway (2). Under these conditions (normoxic, 21% O2), enzymes called prolyl hydroxylases

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_8,

posttranslationally hydroxylate specific proline residues within the ODD (Oxygen-Dependent Degradation) domains present in Hifα subunits. VHL, an E3 ubiquitin ligase, binds to this hydroxyl-modified ODD to ubiquitinate it, thereby constitutively degrading the Hifα subunits. Under hypoxia (1% O2), however, such ODD prolyl hydroxylation does not occur and therefore Hifα subunits escape destruction by VHL. This leads to stabilization and consequent dimerization of the Hifα with the Hifβ subunit under hypoxic conditions, reconstituting the active transcription factor HIF. Therefore, HRE-dependent transcription occurs under hypoxic conditions and not under normoxia. The HRE-containing genes are extremely significant molecules involved in mounting cellular response to low oxygen conditions (3). Some of these genes, such as VEGF (Vascular Endothelial Growth Factor), promote angiogenesis to increase blood supply to the hypoxic tissue. Others are involved in glycolytic pathways that are important for cell survival, and a few are involved in control of cell migration. Cancer cells usurp this pathway by finding ways to stabilize Hifα subunits even under normoxic conditions, thereby altering their metabolism, activating pathways ensuring continuous blood supply and rapid growth of the tumors and upregulating metastatic pathways (4, 5).

Mechanistically, the situation is more complicated than the simplistic view presented above. The three Hifα subunits have acetylation, other prolyl and arginine hydroxylation domains, modifications of which lead to distinct functional consequences. Cellular states that promote and/or reverse these modifications are not very well understood. Studies have also shown that pathways activated by forced expression of Hif1α do not entirely reflect those that are activated by expression of Hif2α and Hif3α, suggesting that the Hifα subunits are not exactly redundant in their function.

A time-tested and reliable way to sort out these complexities is to introduce site-directed mutations, create chimeric proteins by "domain swapping" between the Hifα subunits or employ small and large deletions of the relevant domains to study the individual function of the Hifα subunits on the HIF pathway and cellular oxygen-dependent processes. As these efforts to understand the HIF pathway are under way in our laboratory, we present here examples of our site-directed mutagenesis method to create a wide range of Hifα mutants. These examples serve to impress upon the readers the versatility and speed of the method we had described earlier (6).

2. Materials

2.1. Plasmid Templates and Primers

1. Full-length Hif1α and Hif2α cDNAs cloned in pOTB7 plasmid were purchased from Open Biosystems (Huntsville, Alabama)

as bacterial glycerol stocks. Glycerol stocks were stored at −80°C.

2. 5′-phosphorylated internal primers were complementary to the regions shown in Table 1 and in Figs. 1b and 2a. GC content of the primers is shown in Table 1 to indicate that the mutagenesis protocol works very well on a wide range of templates. The only primers that were nonphosphorylated were the outer 5′- and 3′-primers that are complementary to start and end, respectively, of the protein coding sequence. The nonphosphorylated primers also contained unique restriction enzyme sites to enable cloning of full-length Hif1α and Hif2α. All primers were synthesized at 0.2 μmole scale of synthesis from Operon Biotechnologies Inc. (Huntsville, Alabama). All primers were stored at −20°C.
3. *Optional primer purification*: Oligonucleotide primers are purified by polyacrylamide gel electrophoresis (PAGE) at low cost in the laboratory (see Note 2). PAGE purification is essential to ensure efficiency and fidelity of the procedure described here. PAGE gel for primer purification is typically run in Tris–borate–EDTA (TBE) buffer. To prepare 1 L of 10× TBE, 108 g Tris base, 55 g boric acid, and 40 mL EDTA (0.5 M) were dissolved in sterile distilled water and made up to 1 L (see Note 8). EDTA dissolves at pH 8.0, therefore pH needs to be adjusted with NaOH before EDTA completely dissolves. Store at room temperature.

2.2. Polymerases and PCR Reaction

1. Cloned Pfu DNA polymerase was from Stratagene (La Jolla, California, 2.5 U/μL, Cat#600140).
2. Pfu DNA polymerase reaction buffer 10× concentrate is supplied along with the DNA polymerase. Polymerases and buffer concentrates are stored at −20°C.
3. 100 mM $MgSO_4$ was prepared in sterile distilled water. A stock solution of 1.0 mL is more than sufficient for 1 year. Store at −20°C.
4. 10 mM PCR nucleotide mix was purchased from Promega (Cat# C1141). This serves as a 50× nucleotide mix solution. Storage is at −20°C.
5. DNA templates were plasmids containing cloned Hif1α and Hif2α. Miniprep plasmid DNA preparations were adjusted to a final concentration of 0.1 μg/μL (0.023 picomoles/μL). DNA is stored at −20°C.
6. All primers were prepared to a diluted working stock of 0.05 μM (20× working stock/50 μL reaction) from a master stock concentration of 100 μM.
7. 1% Agarose gels were run in 1× TAE buffer + 0.5 μg/mL Ethidium bromide for DNA visualization.

Table 1

Primers used in this work and relevant information. Primer sequences used to PCR-amplify various segments of HIF1α and HIF2α are indicated. Annealing conditions during PCR for all PCR reactions were 57°C, as indicated. Site-directed mutations were introduced using the modified codon located at the extreme 5′ end so as not to interfere with annealing of rest of the primer to the template and to ensure that the mutation is present through the entire PCR reaction. Primers were designed to maintain the translational reading frame, so that chimeric proteins are generated by ligating any two PCR products that contained appropriately located 5′ phosphates ("mix and match"). GC content percentages are calculated using only the sequence stretch that is complementary to template DNA (added restriction sites, mutated codons and 5′ GCGC overhang are excluded). GC% is shown to indicate that the technique works on a broad range of templates. Codon mutation numbers represent the aminoacid number of the respective proteins. Primer names in parentheses are alternate short name forms used in Figs. 1–3

Primer name	Primer sequence	5′-phosphate	% GC content	PCR annealing temperature	Codon mutation
Hif1α primers					
HIF1aFLF	GCGC *GATATC* ATGGAGG GCGCCGGCGGCGCGAA	No	78.26	57°C	None
HIF1aFLR	GCGC *GGATCC* TCAGTTAA CTTGATCCAAAGCTCTGAG	No	40.74	57°C	None
HIFA402F (402F*)	[Phos]**GCA** GCC GCT GGA GAC ACA ATC ATA T	Yes	50.0	57°C	402P-A
HIFA402R	[Phos]GGC CAG CAA AGT TAA AGC ATC AGG TTC	Yes	48.15	57°C	None
HIFA532F (532F*)	[Phos]**GCG** TTG GAA TTG GTA GAA AAA CTT TTT GCT G	Yes	28.57	57°C	532K-A
HIFA532R	[Phos]GAA TTC ATT GAC CAT ATC ACT ATC CAC	Yes	37.03	57°C	None
HIFA564F (564F*)	[Phos]**GCC** TAT ATC CCA ATG GAT GAT GAC TTC C	Yes	40.0	57°C	564P-A
HIFA564R	[Phos]AGC TAA CAT CTC CAA GTC TAA ATC TGT G	Yes	39.29	57°C	None

Hif2α primers					
HIF2aFLF	GCGC *GATATC* ATGACAG CT GACAAGGAGAAGAA AAGG	No	44.44	57°C	None
HIF2aFLR	GCGC *GGATCC* TCAG GTG GCCTGGTCCAGGGCT	No	68.18	57°C	None
HIFB402R	[Phos]AGC CAG CTG GGC CAG CTC CTC GG	Yes	73.91	57°C	None
HIFB497F (497F*)	[Phos]**GCG** ATT GAA GTG ATT GAG AAG CTC TTC GC	Yes	42.31	57°C	497K-A
HIFB497R	[Phos]CAG GTC GTT ATC CAA AGA TGT GTA ATA	Yes	37.04	57°C	None
HIFB531F (531F*)	[Phos]**GCC** TAT ATC CCC ATG GAC GGG GAA G	Yes	54.54	57°C	531P-A
HIFB531R	[Phos]TGC CAG TGT CTC CAA ATT GTC CAG CT	Yes	50.0	57°C	None

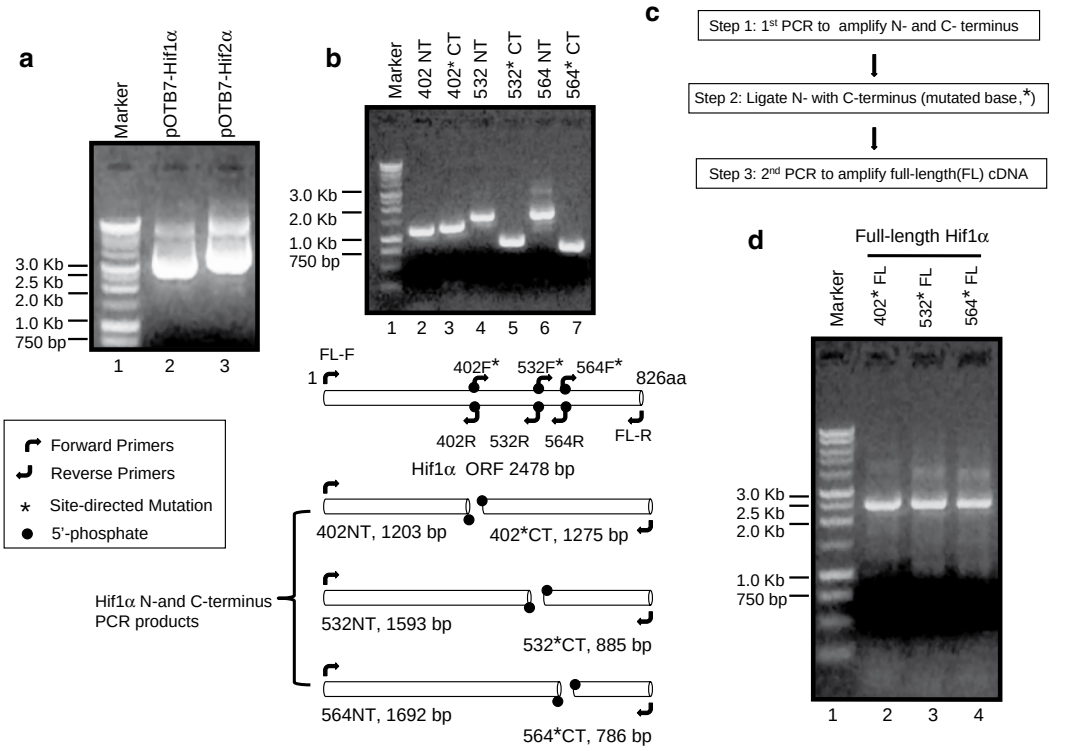


Fig. 1. Generation of point mutations of Hif1α. (**a**) Miniprep purified pOTB7 plasmids containing full-length Hif1α (*lane 2*) and Hif2α (*lane 3*) that were used as templates. Marker lane contains 1 Kb DNA ladder from Promega. (**b**) First-round NT and CT products of Hif1α using primer pairs FL-F and 402R (*lane 2*); 402*F and FL-R (*lane 3*); FL-F and 532R (*lane 4*); 532*F and FL-R (*lane 5*); FL-F and 564R (*lane 6*); 564*F and FL-R (*lane 7*). (**c**) Brief outline of the experimental scheme is shown. (**d**) Second-round PCR to regenerate full-length Hif1α now carrying single point mutations in codons 402P-A (*lane 1*); 532K-A (*lane 2*) and 564P-A (*lane 3*). The expected PCR product sizes are shown in the schematic below in (**b**).

8. 50× TAE buffer: add 242 g of Tris base (Cat# BP-152-1, Fisher) to 750 mL of distilled water. After Tris dissolves, 57.1 mL of Glacial acetic acid (Cat# AC22214-0025; Fisher) and 100 mL of 0.5 M EDTA (Cat# AC11843-2500; Fisher) are added and final volume made up to 1,000 mL. The pH of this solution should be 8.5 and there is no need to adjust pH. Storage of 50× TAE is at room temperature. The 50× stock is diluted to make 1× TAE + 1% Agarose.

2.3. Ligation Reaction

1. T4 DNA ligase (HC): 10–20 U/μL, Cat# M1794, (Promega, Madison, Wisconsin). High concentration (HC) DNA ligase is good for quick 30 min blunt-end ligations that are used in this procedure (see Note 9).
2. 10× DNA ligase reaction buffer is supplied with the enzyme (see Note 16).

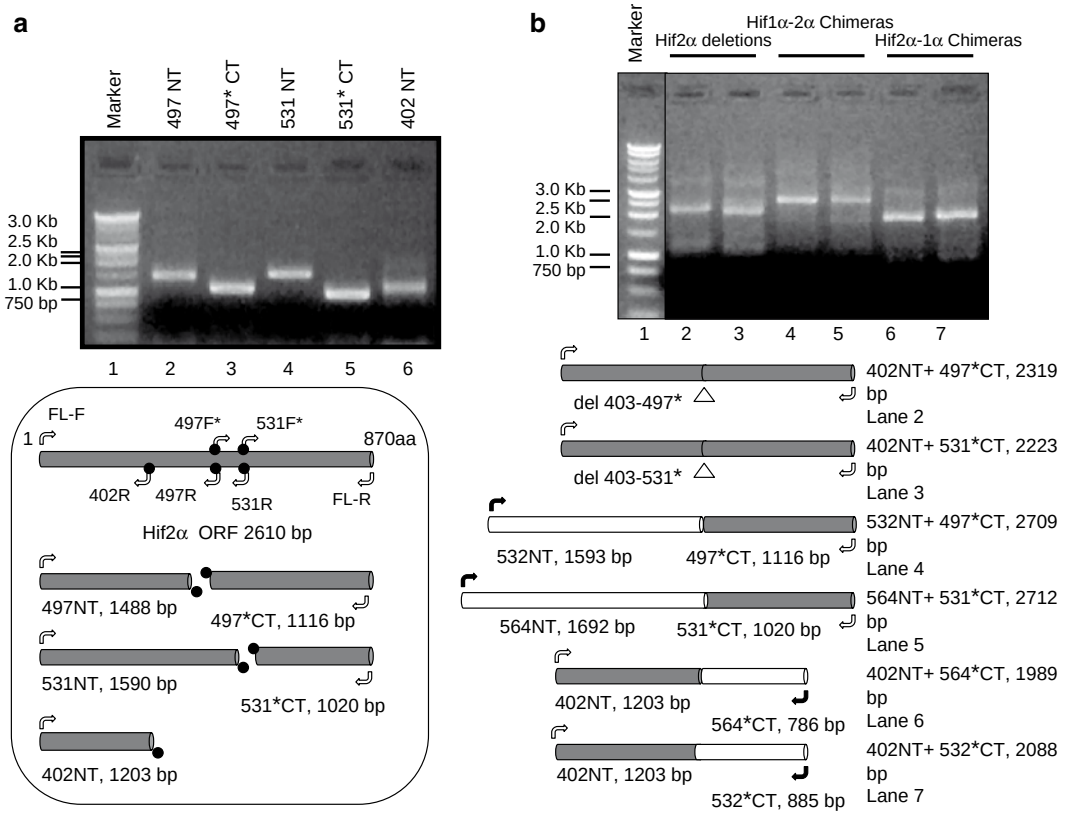


Fig. 2. Hif2α mutants and chimeras. (**a**) First-round NT and CT products of Hif2α using primer pairs FL-F and 497R (*lane 2*); 497*F and FL-R (*lane 3*); FL-F and 531R (*lane 4*); 531*F and FL-R (*lane 5*); FL-F and 402R (*lane 6*). (**b**) The products in Figs. 8.1b and 8.2a were ligated in different combinations to result in Hif2α deletions (*lanes 2–3*), Hif1α(NT)-Hif2α(CT) and Hif2α(NT)-Hif1α(CT) fusions. The expected PCR product size is shown in the schematics below.

2.4. Restriction Enzyme Digestions and Cloning

1. All restriction endonucleases: (New England Biolabs, Ipswich, Massachusetts). Generally, the concentration for the endonucleases is 20 U/μL. One NEB unit is the amount of enzyme required to completely digest 1 μg DNA in 1 h.
2. Restriction endonucleases are supplied with their respective optimal NEB reaction buffers (NEB buffer 1–4) as 10× concentrates (see Note 11). Bovine Serum Albumin (BSA) is also supplied, as some of the enzymes require BSA for optimal activity.
3. OneShot™ Top10 F' Electrocomp ™ *E. coli*: Cat# C4040-06 (Invitrogen, Carlsbad, California) cells are ideal for a rapid transformation protocol using the 5 min ligation method described below (see Subheading 3.3).
4. Luria Broth (LB, Cat# 612725000) with preset pH and Agar (Cat# S70213): (Fisher Scientific, Pittsburgh, PA). Liquid LB media was prepared by adding 20 g/L of LB powder to

distilled water and sterilized by autoclaving at 122°C for 20 min. LB + Agar was prepared by adding 15 g/L of agar powder to LB media and sterilized by autoclaving at 122°C for 20 min.

5. Typically, 25 mL of LB + Agar + antibiotics mix is used per 100 mm bacterial culture plate. For miniprep plasmid DNA isolation, a single colony from the plate is inoculated into 5 mL LB + antibiotic liquid medium. Typical antibiotics final concentrations are: ampicillin, 50 μg/mL; kanamycin, 25 μg/mL; and chloramphenicol, 25 μg/mL. Stocks of ampicillin (50 mg/mL) and kanamycin (25 mg/mL) are prepared in water and filter-sterilized by passing them through sterile 0.22 micron syringe filters. Choramphenicol stock (25 mg/mL) is prepared in 100% ethanol and used without filter sterilization.

3. Methods

Experimental design: The purpose of the experiments was to create several mutants of Hif1α and Hif2α (see Note 1). As mentioned earlier, the Hifα subunits do not have redundant functions. It is currently not known how and why the Hifα subunits regulate different sets of genes in the oxygen-sensing pathway or the global functional consequences of this regulation. To gain insight into this, we desired to introduce mutations and create chimeric fusions of Hif1α-Hif2α and study the genes that are activated by such altered HIfα proteins. Hif1α contains proline residues at positions 402 and 564 that are hydroxylated by 2-oxoglutarate (2-OG)-dependent oxygenases to result in ubiquitinylation and proteasomal destruction. Apart from this, a 532 lysine acetylation by ARD1 is shown to promote Hif1α interaction with VHL and consequent destabilization of the transcription factor. The 803 asparagine hydroxylation site is important for Hif1α biology with distinct functional consequences (7). Hydroxylation of an asparagine residue in the C-terminal transactivation domain of HIF-α prevents its interaction with p300, leading to silencing of transcriptional activation. Analogous sites in Hif2α are 405 and 532 prolyl hydroxylation sites, 497 lysine acetylation site and 847 asparagine hydroxylation site. In Figs. 1–3 we use selected examples of results involving the first three of the four sites mentioned above. Single (Fig. 1), double (Fig. 3a), and triple mutants (Fig. 3b) of Hif1α were created along with single mutants (Fig. 2a), deletions of Hif2α and various chimeric fusions between Hif1α and Hif2α (Fig. 2b). All these were achieved in 2 days, including the

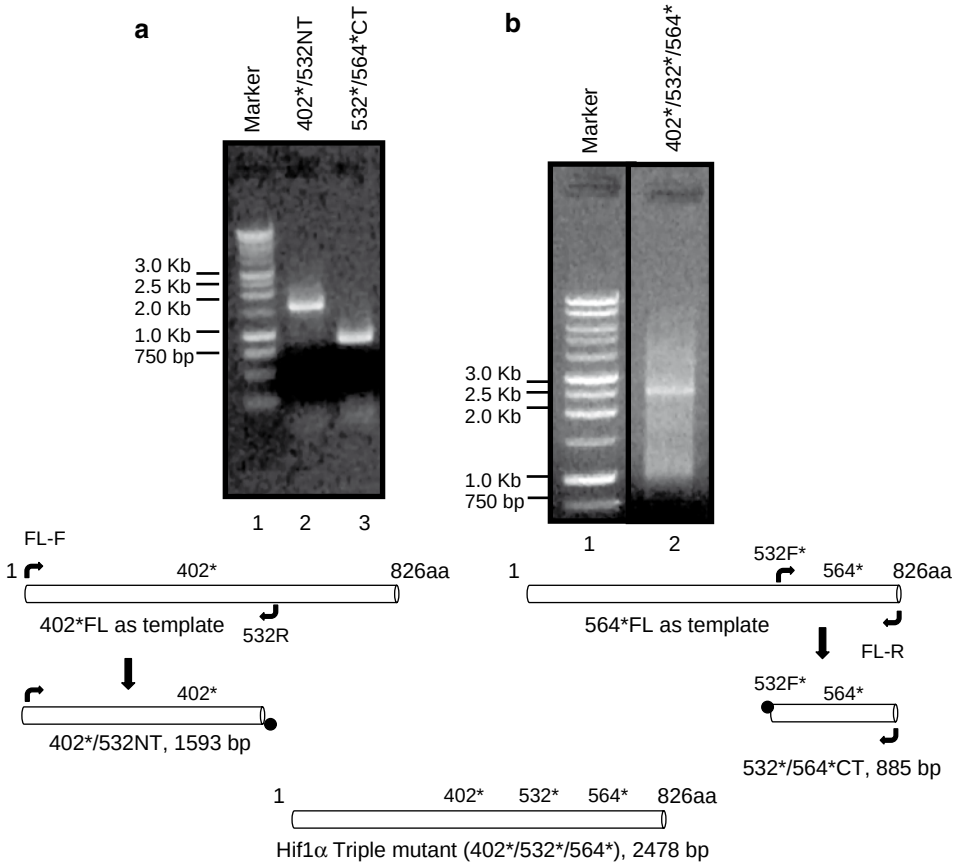


Fig. 3. Triple mutant generation in one day. (**a**) Hif1α 402*FL (from Fig. 1d, *lane 2*) was used as template to generate an NT product with primer pair FL-F and 532R. This results in a product containing 402* mutation (*lane 2*). The CT product was obtained using primer pair 532*F and FL-R but using 564*FL (Fig. 1d, *lane 4*) as template. This product now contains double mutations, 532* and 564* (*lane 3*). (**b**) Ligation of the above generated NT and CT in (**a**) and second-round amplification resulted in triple mutant 402*/532*/564* Hif1α *lane 2*, all in a single day.

experimental planning, stressing the speed of the method in generating desired mutations (see Note 14).

The scheme requires an outermost primer set (Fig. 1, FL-F and FL-R) that amplifies the full-length (FL) cDNA. These primers do not contain 5′-phosphate and do not carry mutated codons. The idea is to generate an N-terminal and a C-Terminal PCR product using a combination of internal and outermost primers (see Notes 4 and 5). All the internal primers always contain 5′-phosphate to allow for ligation of PCR products at their proper termini. The C-terminal (CT) PCR product will contain the mutated site since all internal forward primers were mutagenic primers. For example, in Hif1α 402, the original 402 codon (CCA) is mutated to GCA in the primer HIFA402F (Table 1). This primer is used in combination with FL-R to generate the CT product (Fig. 1b). The reverse primer for the 402 site, HIFA402R,

does not contain a modified codon, and the first three nucleotides of this primer (GGC) are complementary to 401 codon. The 402 reverse primer is used in combination with FL-F to generate an N-terminal (NT) product. This is the first-round PCR and it generates NT and CT from the full-length cDNA template. If a deletion were required, the reverse primer of NT would be a further upstream codon, say 300, to delete the region between 301 and 402. Figure 2b shows examples of such deletions. The next step involves T4 DNA ligase mediated joining of NT and CT. Since NT and CT have only one of their termini phosphorylated, they will be ligated as NT–CT and reform the full-length cDNA that now carries the introduced mutation. A portion of the ligation reaction is directly used as template and PCR-amplified in a second round with outermost primers, FL-F and FL-R, to generate sufficient quantities of full-length cDNA to allow cloning into desired vectors.

3.1. Preparation of Plasmid Templates and Primers

1. The pOTB7 plasmids are chloramphenicol-resistant. Bacterial frozen stocks (stored at −80°C) were scraped while still frozen with a sterile tip (200 μL tip) and immediately streaked onto 25 μg/mL chloramphenicol agar plates for antibiotic selection. A single colony from each chloramphenicol plate was picked, grown overnight in 5 mL of LB + 25 μg/mL chloramphenicol, and plasmid DNA isolated using Qiaprep protocol from Qiagen (Valencia, California). Detailed protocol of the miniprep procedure is available online at Qiagen website. The plasmids were finally eluted from the column with sterile distilled H_2O and after estimating concentration at OD_{260}, both plasmids were adjusted to 0.1 μg/μL (0.023 picomoles/μL) (Fig. 1a). We roughly estimate that a 0.1 μg/μL plasmid is at 500× concentration, and we prepare a 1:10 dilution of 0.1 μg/μL DNA to make a convenient volume (0.5–1.0 μL; 5–10 ng) for a 50 μL PCR reaction.
2. The data sheet provided by oligonucleotide manufacturers contains sufficient information to prepare primer stocks (see Note 6). It is important not to subject master primer stocks to repeated freeze–thaw cycles. Therefore, sufficient working stocks need to be immediately prepared and the main stock is stored away as lab stock. It is variously mentioned in PCR protocols that primer concentrations of anywhere between 0.1 and 1.0 μM are good for a PCR reaction. In reality, however, it is important to keep the primer concentrations close to optimal for the reaction conditions, especially for the procedure described here. Since our initial studies, we have subsequently significantly reduced our primer concentrations in PCR reactions to enable more efficient

ligation and second-round PCR amplification. For this, we prepared master stocks of 100 μM. From the master stocks, the diluted working stocks of 0.05 μM (20× working stock/50 μL PCR reaction) were used for each primer.

3.2. First- and Second-Round PCR Reactions

In principle, there is no difference between the first- and second-round PCR reactions. The only differences are that the first round uses plasmid DNA as template and the second uses ligated first-round PCR products from an intermediate ligation reaction as template (see Note 7). Of course, the first-round PCR uses a combination of outermost and internal primers while in the second-round PCR only outermost primer pair is used. All PCR reactions are cycled keeping in mind the melting temperature of the primers (T_m). The primer melting temperatures are designed to be within 4–5°C of each other so that annealing temperatures are not vastly different. Following is the PCR mixture for one 50 μL reaction.

1. 10× Pfu buffer, 100 mM $MgSO_4$, 10 mM dNTPs, template DNA stock, and the primers need to be completely thawed and mixed well. Reactions are assembled on ice.
2. A 5.0 μL volume of 10× Pfu buffer is added to a thin-walled PCR tube.
3. 10 ng DNA (1.0 μL) of the stock is added to the PCR tube. For second-round PCR, the template is the ligated product of the first-round PCR (see Subheading 3.3). In this case, 1.0 μL of the ligation reaction is used as template.
4. 1.0 μL of 10 mM dNTP mix is pipetted into the PCR tube.
5. 2.0 μL of 100 mM $MgSO_4$ is added to the PCR mix. This is sufficient for amplification of up to 3 Kb DNA. If longer products need to be amplified, $MgSO_4$ can be increased to 0.5 mM/1 Kb DNA.
6. 0.05 μM of working primer stock is at 20× concentration. 2.5 μL of the working stock is used for each primer.
7. 35.0 μL of sterile distilled water is added to the reaction mixture and briefly mixed on a vortex mixer.
8. 1.0 μL of 2.5 U/μL Pfu DNA polymerase is finally added and vortex-mixed. The reaction is collected by quick 5 s centrifugation.
9. The PCR cycling conditions (see Note 15) were Step 1: 94°C, 2 min; Step 2: 94°C, 15 s; Step 3: 57°C, 15 s; Step 4: 72°C, 2–4 min (2 min for first-round and 4 min for second-round PCR); GOTO Step 2, 36 cycles; Step 5: 72°C, 5 min; Step 6: 4°C.

3.3. Ligation of First-Round PCR Products

1. The ligation of second-round full-length product into cloning/expression vectors is the same as the ligation procedure described below. However, ligation of first-round PCR products requires careful attention (see Note 12). An aliquot of 0.5–1.0 μL of the first-round PCR products (approximately containing 75–100 ng/μL of DNA) can be directly ligated without requiring further purification. However, the success of direct ligation after first-round PCR largely depends on the absence of nonspecific bands (such as in Figs. 1b and 2a). It is highly advisable to use Zymoclean™ for rapid cleanup of reactions or gel elution for purification of the band (see Note 3) if there are nonspecific products along with the expected product.
2. Use no more than 40–60 ng of each NT and CT for ligation. Avoid using excess DNA for ligation reactions.
3. A 10.0 μL ligation reaction contained 1.0 μL 10× T4 DNA ligation buffer, 0.5 μL of NT, 0.5 μL of CT, 7.5 μL sterile distilled water, and 0.5 μL T4 DNA ligase (HC). Refer to Note 16 for ligation buffer.
4. Blunt-end ligation reaction is carried out at room temperature for 30 min (see Note 10).
5. 1.0 μL of the ligation reaction is taken for second-round PCR. In case of ligation into cloning/expression vectors, 4.0 μL of the ligation is transformed into competent *E. coli*.

3.4. Restriction Enzyme Digestions and Cloning

1. Primers amplifying the full-length cDNA contain unique restriction enzyme sites for directional cloning into any vector of choice (see Note 13). Also, the FL-primers do not contain 5′ phosphate and, therefore, before cloning the mutated cDNA, the restriction enzyme sites need to be digested to generate desired overhangs and to expose 5′-phosphate.
2. We routinely perform restriction digestions in a 20.0 μL reaction volume. As discussed earlier, restriction enzymes require specific buffer conditions and BSA to exhibit optimal activity (see Note 11). A 20.0 μL reaction contained 2.0 μL restriction enzyme buffer (10×); 0.5 μg DNA (either PCR product or vector DNA); 0.5 μL of appropriate 20 U/μL restriction enzyme(s) for single or double digestions, BSA (as required) and sterile distilled water added to make up final volume of 20.0 μL.
3. Digested products were purified using Zymoclean™ and finally eluted with water in 10 μL volume, and ligation was performed as described above using 100 ng PCR product and 50 ng vector.

4. 50 μL of competent cells were thawed from −80°C, and 2.0 μL of 0.5 M β-mercaptoethanol was added (see Note 17). Transformation into chemically competent Top10 *E. coli* cells was performed by adding 3.0 μL of the ligation product to the competent cells and incubating on ice for 30 min.
5. Transformation mix was heat-shocked at 42°C for 45 s and immediately chilled on ice for 2 min.
6. 500 μL of LB media (without antibiotics) is added to the transformation mix and incubated at 37°C for 1 h to allow expression of antibiotics resistance gene.
7. Finally, 50 μL of the bacteria is plated on selection plates and colonies are grown overnight at 37°C.

4. Notes

1. Due to the generation of many PCR products, it is very important to take the time to plan the experiment and draw up a detailed scheme of the planned reactions, their products, the ligations, and the final products (Figs. 1–3). Use of appropriate primer names add to the convenience of work and avoid significant confusion with multiple users. We have developed a laboratory practice where the mutagenic primer is always the forward primer. Also, information of the mutated site and the cDNA identity is indicated in the primer name (Table 1).
2. *PAGE-purification of PCR primers*: If cost is not an issue, all PCR primers can be ordered as PAGE-purified. If purification is to be performed in the laboratory, it is a very simple procedure and can be performed with minimal hands-on time. Vertical PAGE gels for DNA electrophoresis are cast similarly to protein gels, but without the need for any stacking gel. 10× TBE is mixed with acrylamide/bis-acrylamide (37.5:1) to give 15% gel strength. After gel polymerization in a mini-gel format, 20 μL of 100 μM each primer stock is mixed with 4 μL of 6× DNA loading dye and loaded into the well, with a 10 bp DNA ladder as marker in the next lane. Electrophoresis is performed until the tracking dye is at the bottom of the gel. The PAGE gel is removed from the cassette and stained with 0.5 μg/mL Ethidium Bromide (EtBr). PAGE gels, due to their small thickness, stain very quickly in 5–7 min. Full-length primer bands are identified based on the DNA size marker using UV transilluminator or UV-shadowing and the bands are excised. The PAGE gel piece is sliced into smaller pieces using a clean blade, placed in a sterile Eppendorf tube

and sterile distilled water is added until the gel material is fully immersed. An overnight incubation at 37°C will allow primers to elute out into the water. The water (containing the primer) is collected avoiding the gel and the concentration is calculated by measuring absorption at OD_{260} in a UV-spectrophotometer. Primers are ready for any downstream application without needing any further purification.

3. *Rapid clean up of DNA*: In our laboratory experience, the mutagenesis protocol described here is virtually fail-safe. It is a good practice to check the first-round PCR products on agarose gels to ensure the presence of expected bands, especially when using a primer set for the first time. For some users, it might be informative, particularly if the template DNA preparation contains undesirable impurities or the first-round PCR results contain additional nonspecific products, to consider a step of DNA purification. We use Zymoclean™ gel DNA recovery kit (Cat# 4002) from Zymo Research (Orange, California), which is an easy-to-use spin column DNA purification format. This kit contains agarose dissolving buffer (ADB) for the recovery of expected PCR products from the agarose gel. The buffer can also be used to repurify DNA in solution such as template DNA. Since the column matrix does not bind to <50 bp size DNA, it can be used to remove left-over primers or primer dimers from PCR reactions. We typically use the purification step before second-round PCR reaction and for ligation of full-length product into vectors. It takes less than 5 min to complete the purification steps using this kit.
4. Pfu DNA polymerase is ideal for this application, based on our experience and feedback from several investigators. Other thermophillic DNA polymerases were either found to interfere with the direct ligation step or alter the termini of the first-round PCR products. Since the integrity of the joining termini (ligation junction) is crucial for the success of the method, these issues should be carefully considered before substituting any other DNA polymerase.
5. Typically, 1 min of extension time is optimal for 1 Kb product size. However, the first-round PCR is likely to generate products of varied sizes and cycling conditions (extension times) may not be ideal for all PCR products sizes. Generally, PCR products with size differences of 750 bp between them can be amplified using the same extension time. If the expected size differences are vastly greater than this, those specific reactions need to be cycled separately.
6. If marginal extra cost is not an issue, all the PCR primers can be ordered as PAGE-purified. PAGE-purified primers significantly improve accuracy of PCR amplification and generation of the mutants.

7. Reading frame should be maintained in the primers (see Table 1) so that different N- and C-terminal products can be "mixed and matched" to create deletions, duplications, or chimeras (Fig. 2b).
8. 10× TBE stock solution is stable for 2–3 months. Higher concentrations such as 20× TBE precipitate quicker.
9. The ligation reaction performed in this method involves blunt-end joining. Therefore, DNA ligases such as T4 DNA ligase that have blunt-end joining activity should be used. *E. coli* DNA ligase works on cohesive termini only and is not suited to this protocol.
10. T4 DNA ligase is very sensitive to alcohol (used in wash buffer in Zymoclean kit). Care should be taken to spin out the wash buffer and let the column air-dry before elution.
11. Restriction enzyme activity is sensitive to alcohol and high glycerol concentrations (glycerol is used as a preservative in enzyme stocks). Care should be taken not to use too much of the restriction enzyme in a reaction as this would lead to an increase in glycerol concentrations and enzyme/DNA ratio, leading to nonspecific cleavage of the DNA.
12. High-concentration ligase (HC), NEB or Promega, significantly enhances the blunt-end ligation product in 30 min even at room temperature. The efficiency of ligation is important for second-round PCR yield.
13. Careful consideration should be given to the choice of restriction sites at primer design step since directional cloning requires the final full-length PCR products to be digested with two enzymes (double digestion) that will have optimal enzymatic activity under same reaction buffer conditions.
14. The procedure is well suited to generate mutations, including point mutations, deletions, and duplications, and unlike any other existing protocol, can generate triple or multiple mutants in a single day. An example is shown in Fig. 3. In this example, we used Hif1α 402*FL (Fig. 1d) as template to generate an NT product with FL-F and 532R. On the other hand, the CT product was obtained using 532F and FL-R but this time using 564*FL (Fig. 1d) as template such that the product now contains double mutations, 532* and 564*. Ligation of the above generated NT, CT, and second-round amplification resulted in triple mutant 402*/532*/564* Hif1α, all in a single day.
15. The short 15 s denaturing or annealing steps in PCR reactions (Subheading 3.2) is possible with newer versions of PCR machines. If using older machines, these steps need to be 30 s. The extension time of 2–4 min is the same for newer or older machines.

16. 10× ligation buffers contain dithiothreitol (DTT) and ATP as cofactor for ligation. These tend to precipitate under concentrated buffer conditions at low temperatures (−20°C). Therefore, the buffer needs to be completely thawed until no visible signs of any precipitate remain.
17. Competent cells are extremely sensitive to temperature and mechanical stress. They should be left on ice to completely thaw and pipetted gently into ice-cooled Eppendorf tubes.

Acknowledgments

This work would not have been possible without the continued grant support from NIH (NCI) 1RO1 CA128002.

References

1. Nangaku M, Inagi R, Miyata T, Fujita T (2008) Hypoxia and hypoxia-inducible factor in renal disease. Nephron Exp Nephrol 110:e1–e7
2. Maynard MA, Qi H, Chung J, Lee EH, Kondo Y, Hara S, Conaway RC, Conaway JW, Ohh M (2003) Multiple splice variants of the human HIF-3 alpha locus are targets of the von Hippel-Lindau E3 ubiquitin ligase complex. J Biol Chem 278:11032–11040
3. Fong GH (2008) Mechanisms of adaptive angiogenesis to tissue hypoxia. Angiogenesis 11:121–140
4. Champion KJ, Guinea M, Dammai V, Hsu T (2008) Endothelial function of von Hippel-Lindau tumor suppressor gene: control of fibroblast growth factor receptor signaling. Cancer Res 68:4649–4657
5. Hervouet E, Cízkova A, Demont J, Vojtiskova A, Pecina P, Franssen-van Hal NL, Keijer J, Simonnet H, Ivanek R, Kmoch S, Godinot C, Houstek J (2008) HIF and reactive oxygen species regulate oxidative phosphorylation in cancer. Carcinogenesis 29:1528–1537
6. Adereth Y, Champion KJ, Hsu T, Dammai V (2005) Site-directed mutagenesis using Pfu DNA polymerase and T4 DNA ligase. Biotechniques 38:864–868
7. Masson N, Ratcliffe PJ (2003) HIF prolyl and asparaginyl hydroxylases in the biological response to intracellular O(2) levels. J Cell Sci 116:3041–3049

Chapter 9

Rapid Sequence Scanning Mutagenesis Using *In Silico* Oligo Design and the Megaprimer PCR of Whole Plasmid Method (MegaWHOP)

Ulrich Krauss, Karl-Erich Jaeger, and Thorsten Eggert

Abstract

A wide variety of random- and site-directed mutagenesis techniques have been developed to investigate the structure–function relationship in proteins and intergenic regions like promoter sequences. Similar techniques can be employed to optimize protein properties like enantioselectivity, substrate specificity, and stability in a directed evolution approach. Due to the tremendous genetic diversity that is created by common random-mutagenesis methods, directed evolution techniques usually require the time-consuming and cumbersome screening of large numbers of variants. A gene-scanning saturation-mutagenesis approach represents one efficient way to limit the screening effort by reducing the created genetic diversity. In structure/function studies often a similar method, e.g., alanine- or arginine-scanning mutagenesis, is used to probe the role of specific amino acids in a protein. Here, we present a standardized mutagenesis strategy that can speed up the process of scanning whole proteins for structure/function studies and, furthermore, allows for the fast and efficient generation of gene-scanning saturation-mutagenesis libraries to be used in the directed evolution of enzyme functions and properties. The described method uses automated computer-assisted oligonucleotide design, and a two-step PCR-mutagenesis protocol to amplify site-specifically mutated circular plasmids that can be directly transformed in *Escherichia coli* expression strains.

Key words: Sequence-scanning mutagenesis, Saturation mutagenesis, Directed evolution, Structure/function studies

1. Introduction

Site-directed mutagenesis is employed to answer fundamental questions regarding structure–function relationships of proteins. A useful extension is scanning mutagenesis, i.e., the subsequent substitution of each amino acid in a protein or a specific target region.

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_9,

This technique constitutes a powerful strategy to reveal mechanistic aspects of protein function, including ligand and substrate binding, activity and enantioselectivity (1, 2). The same strategy can also be employed to scan DNA, e.g., intergenic regions like promoter/repressor sequences to test for DNA-target recognition by DNA-binding proteins (3). Often, cysteine (4), alanine (5), serine (6) or arginine-scanning mutagenesis (7) is performed to probe specific side-chain interactions. An illustration of a scanning-mutagenesis approach for a hypothetical 180 amino acids protein is depicted in Fig. 1.

In directed evolution experiments which aim to optimize specific enzyme properties like enantioselectivity, substrate specificity or thermostability, it is important to generate a high-quality library of enzyme variants to sample the available sequence space (8). Here, complete gene-scanning saturation mutagenesis constitutes at least an efficient starting point because each amino acid of a given enzyme is substituted for the remaining 19 amino acids (saturation) (9–11). At the same time, such an approach significantly limits the screening efforts due to limited genetic diversity that is generated. Experimentally, saturation is achieved by using degenerate mutagenesis oligonucleotides to scan (saturate) the entire protein of interest.

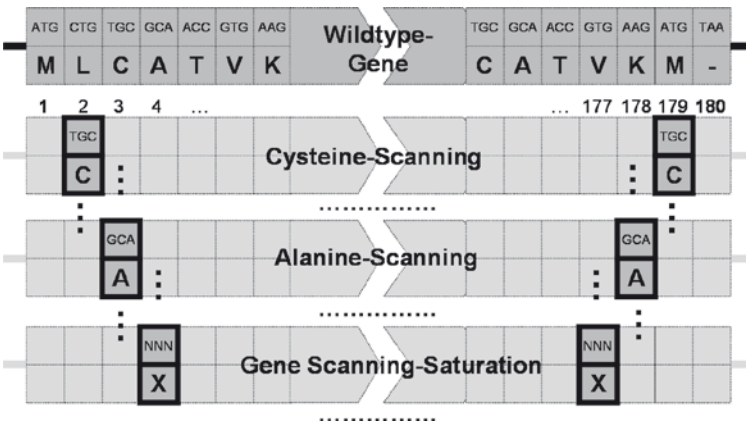


Fig. 1. *Illustration of a gene-scanning mutagenesis approach*. The figure illustrates the general strategy employed in a gene-scanning mutagenesis experiment. For a cysteine-scanning experiment, each amino acid of a given protein is subsequently exchanged against cysteine. In the depicted example, the base triplet *CTG* encoding a leucine at position 2 of the protein or codon *AAG* encoding methionine at position 179 are exchanged against the cysteine codon *TGC* to result in variants L2C or M179C, respectively. Subsequently, base exchanges at positions 3–178 will result in all respective single cysteine variants. Gene-scanning saturation mutagenesis can be performed employing the same strategy. Here, the respective codon positions are subsequently substituted such that any of the remaining 19 amino acids are incorporated by using degenerate oligonucleotide primers (corresponding to coding triplet *NNN* at the respective position).

Nevertheless, the complete scanning (or saturation) of a given protein scaffold remains a laborious task, mainly due to the need to design a large number of mutagenesis oligonucleotides, to perform a large number of different PCRs and several subsequent steps of cloning the mutated genes into plasmids for expression. Therefore, it is highly desirable to standardize the oligonucleotide design, the employed mutagenesis strategy as well as the cloning procedures.

The so-called MegaPrimer method described by Barettino *et al.* (12) is the method of choice to specifically substitute any given target amino acid position. In brief, this method makes use of a PCR-amplified mutated MegaPrimer, generated by a previous PCR that consists only of a fraction of the full-length target gene, but contains the mutated codon triplet. This MegaPrimer is subsequently used together with vector-specific forward and reverse primers to amplify the respective mutated full-length gene in a second PCR. These mutated genes then have to be cloned, with efficient ligation being the major limitation.

Several PCR mutagenesis strategies as well as ligation-independent cloning methods have been developed in recent years to cope with the issues related to inefficient PCR mutagenesis and cloning. The QuikChange™ mutagenesis method (Stratagene, LaJolla, CA, USA) uses methylated plasmid-DNA templates and complementary mutagenesis oligonucleotides to amplify the whole target gene containing plasmid in 5′and 3′direction up- and downstream of the mutagenized site. This procedure generates nicked circular PCR products that contain the desired mutation at the target site. The methylated parental-template DNA is subsequently digested by using the restriction enzyme *Dpn*I, which specifically hydrolyzes methylated (and hemi-methylated) DNA and leaves behind the nonmethylated circular PCR products. The resulting reaction mixture is thus free of parental-template DNA and can be transformed directly into any expression host strain.

We describe here a standardized mutagenesis strategy which uses automated computer-assisted oligonucleotide design to generate mutated MegaPrimers that can subsequently be used in a QuikChange™ like approach to amplify site-specifically mutated circular plasmids. After *Dpn*I digestion of the template DNA, they can directly be transformed into *E. coli* expression host strains. This mutagenesis method, called MegaPrimer-PCR of whole Plasmids (MegaWHOP), has previously been employed to efficiently generate an error-prone PCR library of the Green Fluorescent Protein (GFP) in *E. coli* (13).

This standardized mutagenesis strategy speeds up the process of scanning whole proteins for structure–function studies and, furthermore, allows for the fast and efficient generation of gene-scanning saturation-mutagenesis libraries to be used in the directed evolution of enzyme functions and properties.

2. Materials

1. Methylated template plasmid DNA, containing the gene of interest prepared from a *Dam* and/or *Dcm* positive strain (see Note 1).
2. Oligonucleotide primers P1 and P2, binding outside of the gene of interest (complementary to the plasmid sequence) and scanning-mutagenesis primer SM containing the desired codon exchange (see Fig. 2) (see Note 2).
3. A proofreading thermostable DNA polymerase, preferentially Turbo-*Pfu* Polymerase (Stratagene, LaJolla, CA, USA): 2.5 U/µl (see Note 3).
4. PCR buffer: e.g., 10× cloned *Pfu* DNA polymerase reaction 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, and 1 mg/ml nuclease-free BSA) (Stratagene, LaJolla, CA, USA).
5. Deoxynucleoside triphosphates (dNTPs) 10 mM each NTP.
6. Thermocycler.
7. Commercial Kit for the extraction of DNA from low-melting point agarose gels (see Note 4).
8. *Dpn*I restriction endonuclease (e.g., available from Fermentas, New England Biolabs, Roche, etc.).

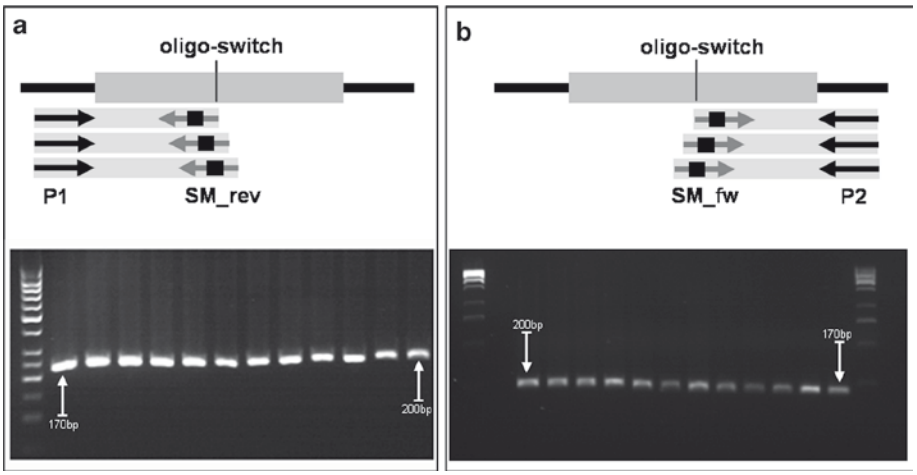


Fig. 2. *Strategy to construct mutated MegaPrimers.* The PCR-synthesized MegaPrimer should not exceed a certain length, to facilitate effective MegaPrimer extension during the second round of PCR (MegaWHOP-PCR). Thus, a so-called oligo-switch is performed: at a position halfway through the target gene a switch is conducted from (**a**) a reverse scanning-mutagenesis primer (SM_rev) (which is used together with the vector-specific forward primer P1) to (**b**) a forward scanning-mutagenesis primer (SM_fw) (which is used together with the vector-specific reverse primer P2). This results in the amplification of MegaPrimers of (**a**) increasing size (here 170 bp–210 bp) or (**b**) decreasing size (200 bp–170 bp) depending on the target amino acid position as shown in the respective agarose gels.

9. DNA-size marker.
10. Ethidium bromide (EtBr) stock solution 0.5 µg/ml.
11. Competent *E. coli* cells as recipient of the mutated target gene-containing plasmids (see Note 5).

3. Methods

3.1. Automated Primer Design

The design of oligonucleotide primers for scanning-mutagenesis purposes can be a time-consuming and laborious task, if a large number of amino acid positions must be scanned representing a protein domain or an entire protein. Therefore, we developed a web browser-based tool that enables automatic prediction of scanning (and saturation) mutagenesis primers over a defined region of the gene. The tool (*insilico.mutagenesis*) is accessible at www.insilico.uni-duesseldorf.de, and its application for primer design has been published (14).

In brief, the program takes into account several general guidelines that are also applied for manual primer design.

1. The desired codon exchange should be placed centrally in the primer, and primer sequences should extend at least 15 bp up- and downstream of this codon exchange.
2. Optimal primers have lengths of 33–42 bp.
3. The calculated melting temperature of the primer should be in the range of 60–70°C, so that annealing can be performed at 50–60°C (see Note 6).
4. If possible, the GC content should be around 50% and the last two terminal 3′ bases should be able to form GC clamps (15).

Please note that the *insilico.mutagenesis* tool optimizes primer length(s) to optimally match those criteria and tries to generate scanning-mutagenesis primers which have similar properties. Therefore, all first-round PCRs (synthesis of the megaprimer) can be performed under the same conditions. In practice, it proved advantageous if the synthesized megaprimers do not exceed a certain length (12, 16) to facilitate effective MegaPrimer extension during the second round of PCR. Therefore, a so-called oligo-switch is included in the primer design strategy (Fig. 2a). To this end, at a position halfway into the target gene a switch is conducted from a reverse scanning-mutagenesis primer (SM_rev) (which is used together with the vector-specific forward primer P1) to a forward scanning-mutagenesis primer (SM_fw) which is used together with the vector-specific reverse primer P2 (Fig. 2b).

3.2. PCR 1: MegaPrimer Synthesis

1. The DNA fragments carrying the mutated codons (= MegaPrimers) are amplified using a proofreading DNA polymerase (e.g., Turbo-*Pfu*, see Note 3). Amplifications are carried out in a total reaction volume of 50 µl. A typical reaction contains 1× PCR buffer, 200 µM of each of the four NTPs, 100–500 pg of template plasmid-DNA, 20 pmol of each primer, and 2.5 U of the DNA-polymerase. Exemplary PCR cycling conditions are as follows: initial denaturation at 95°C for 1 min, followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 50–60°C for 30 s and extension at 72°C for 1 min. Cyclic amplification was followed by a final extension step at 72°C for 10 min and finally, reaction mixtures were cooled down to 4°C (see Note 7).
2. Run the complete PCR mix on an appropriate agarose gel together with a DNA-size marker.
3. Visualize the PCR product by EtBr staining under UV light (Fig. 2b and Fig. 3c) and cut out the desired band containing the MegaPrimer.

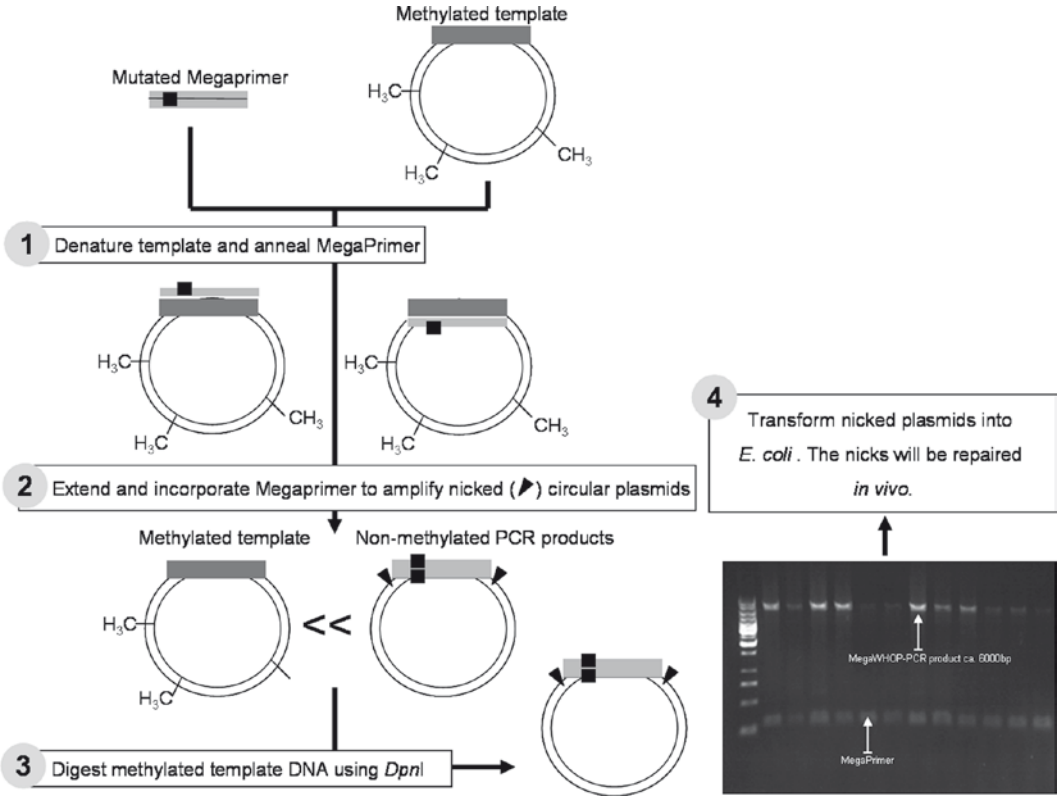


Fig. 3. *MegaPrimer PCR of whole plasmid (MegaWHOP).* In a second round of PCR, a mutated MegaPrimer is used to amplify the target gene-containing plasmid in a QuikChange™ approach. The corresponding agarose gel shows amplification products of approximately the size of the circular plasmid (ca. 6,000 bp) and a fragment of smaller size of the remaining MegaPrimer (here varying in size between approx. 170 bp and 200 bp). After *DpnI* digestion of the methylated parental template DNA, the PCR mixture can directly be transformed into *E. coli* for expression.

4. The megaprimer is purified from the gel using commercially available gel-extraction kits (see Note 4) and eluted in sterile dH_2O.

3.3. PCR 2: Whole Plasmid Amplification

In the second PCR, the purified MegaPrimer is used together with the methylated plasmid-DNA template in a QuikChange™-like approach. This technique known as Megaprimer PCR of whole plasmid (MegaWHOP) (Fig. 3) (13) allows the ligation-free cloning of mutated genes into plasmids and their expression in any *E. coli* strain. Whereas, MegaPrimer synthesis is fairly straightforward, PCR amplification of the entire plasmid may require optimization (see Note 8).

1. A typical 50 µl reaction mix contains 1× PCR-buffer, 200 µM each of four NTPs, 50–200 ng of template plasmid DNA (see Note 9), 10–50 ng of purified MegaPrimer (synthesized in the first round of PCR, subheading 3.2) (see Note 9), 2.5 U of a proofreading DNA polymerase (e.g., Turbo-*Pfu*). A primer-free reaction mix is included in each MegaWHOP PCR experiment as a negative control.

   Exemplary PCR cycling conditions are initial denaturation at 95°C for 1 min, followed by 20–25 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s and extension at 68°C for a time depending on the size of the plasmid to be amplified (see Note 8).
   Cyclic amplification is followed by a final extension step at 68°C for 10 min and the reaction mixtures were cooled down to 4°C.
2. Run 10 µl of the PCR product on an appropriate agarose gel and visualize DNA by EtBr staining und UV illumination (Fig. 3).
3. Add 10 U of *Dpn*I restriction enzyme directly to 50 µl PCR mix and incubate for 1–2 h at 37°C.

3.4. Expression of Mutant Genes in E. coli

1. Transform competent *E. coli* cells with 1–8 µl of the *Dpn*I digested PCR product and grow cells using standard laboratory protocols (17) (see Note 9).

4. Notes

1. (Hemi-)methylated template plasmid-DNA is required for the second round of PCR (MegaWHOP-PCR), but can also be used in the first round of mutagenesis. Any *dcm*, *dam* positive *E. coli* strain can be used to prepare (hemi-)methylated plasmid DNA, including the well-established host strains *E. coli* DH5α, *E. coli* XLI Blue and *E. coli* BL21 (DE3). It should be noted

here that *E. coli* JM110 and its derivatives carry a deletion of the d*am* and *dcm* methylation systems, and hence cannot be used.

2. General guidelines for oligonucleotide design should be employed, when the primers are designed manually. Design rules used in the *insilico* design strategy are summarized under Subheading 3.1.
3. Generally, Turbo-*Pfu* DNA polymerase (Stratagene, La Jolla, CA, USA) is used during both rounds of PCR. However, other proofreading DNA polymerases (e.g., Phusion High-Fidelity DNA Polymerase, Finnzymes, Espoo, Finland) can be used if amplifications using Turbo-*Pfu* prove to be difficult.
4. Extraction of DNA fragments from agarose gels should be performed using commercially available kits. When DNA fragments of low molecular weight (e.g., below 50 bp) were to be extracted, the QIAEX II Gel extraction kit (QIAGEN, Hilden, Germany) worked best in our hands.
5. It is suggested to use highly competent *E. coli* cells, e.g., electrocompetent cells prepared according to standard protocols (17), because the transformation efficiency is low for nicked circular MegaWHOP-synthesized plasmids.
6. The *insilico mutagenesis* tool calculates the oligonucleotide melting point based on the equation of Breslauer et al. (18) using the nearest neighbor thermodynamic parameter set as described by Allawi and SantaLucia (19).
7. Annealing temperatures might be subject to optimization; however, with oligonucleotides predicted by the *insilico mutagenesis* tool an annealing temperature of 55°C worked well for most cases.
8. Several factors influence the efficient amplification of the full-length circular mutation-carrying plasmids during the second round of PCR. Optimization strategies may include the addition of PCR additives such as dimethylsulfoxide (DMSO) and betaine, adjustment of the template/megaprimer ratio, testing of different DNA polymerases to be used for amplification as well as various cycling conditions and annealing temperatures. Furthermore, for the second round of PCR (MegaWHOP-PCR), 20–25 cycles should not be exceeded. This is due to the fact that the error rate of Turbo-*Pfu* DNA polymerase tends to increase with the cycle number during the amplification of circular plasmids (20).
9. Although recombinant clones should contain only the mutated MegaWHOP-PCR products, it is advisable to sequence several of the mutated genes to verify correct mutations. In theory, the methylated template DNA should completely be hydrolyzed by *Dpn*I; however, in practice we observed that about 80% of the obtained clones contained the correct mutations.

References

1. Valeva A, Siegel I, Wylenzek M et al (2008) Putative identification of an amphipathic alpha-helical sequence in hemolysin of *Escherichia coli* (HlyA) involved in transmembrane pore formation. Biol Chem 389:1201–1207
2. Hosaka Y, Iwata M, Kamiya N et al (2007) Mutational analysis of block and facilitation of HERG current by a class III anti-arrhythmic agent, nifekalant. Channels (Austin) 1:198–208
3. Shao Y, Feldman-Cohen LS, Osuna R (2008) Functional characterization of the *Escherichia coli* Fis-DNA binding sequence. J Mol Biol 376:771–785
4. Toloue MM, Woolwine Y, Karcz JA, Kasperek EM, Nicholson BJ, Skerrett IM (2008) Site-directed mutagenesis reveals putative regions of protein interaction within the transmembrane domains of connexins. Cell Commun Adhes 15:95–105
5. Sen J, Jacobs A, Caffrey M (2008) Role of the HIV gp120 conserved domain 5 in processing and viral entry. Biochemistry 47:7788–7795
6. Weis R, Gaisberger R, Gruber K, Glieder A (2007) Serine scanning: a tool to prove the consequences of N-glycosylation of proteins. J Biotechnol 129:50–61
7. Zupnick A, Prives C (2006) Mutational analysis of the p53 core domain L1 loop. J Biol Chem 281:20464–20473
8. Wong TS, Zhurina D, Schwaneberg U (2006) The diversity challenge in directed protein evolution. Comb Chem High Throughput Screen 9:271–288
9. Brissos V, Eggert T, Cabral JM, Jaeger KE (2008) Improving activity and stability of cutinase towards the anionic detergent AOT by complete saturation mutagenesis. Protein Eng Des Sel 21:387–393
10. Funke SA, Otte N, Eggert T, Bocola M, Jaeger KE, Thiel W (2005) Combination of computational prescreening and experimental library construction can accelerate enzyme optimization by directed evolution. Protein Eng Des Sel 18:509–514
11. Eggert T, Funke SA, Andexer JN, Reetz MT, Jaeger KE (2009) Evolution of Enantioselective *Bacillus subtilis* Lipase. In: Lutz S, Bornscheuer UT (eds) Protein engineering handbook. Wiley-VCH, Weinheim, pp 441–451
12. Barettino D, Feigenbutz M, Valcarcel R, Stunnenberg HG (1994) Improved method for PCR-mediated site-directed mutagenesis. Nucleic Acids Res 22:541–542
13. Miyazaki K, Takenouchi M (2002) Creating random mutagenesis libraries using megaprimer PCR of whole plasmid. Biotechniques 33:1033–1038
14. Krauss U, Eggert T (2005) insilico.mutagenesis: a primer selection tool designed for sequence scanning applications used in directed evolution experiments. Biotechniques 39:679–682
15. Lowe T, Sharefkin J, Yang SQ, Dieffenbach CW (1990) A computer program for selection of oligonucleotide primers for polymerase chain reactions. Nucleic Acids Res 18:1757–1761
16. Ling MM, Robinson BH (1997) Approaches to DNA mutagenesis: an overview. Anal Biochem 254:157–178
17. Sambrook J, Russell D (2001) Molecular cloning a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
18. Breslauer KJ, Frank R, Blocker H, Marky LA (1986) Predicting DNA duplex stability from the base sequence. Proc Natl Acad Sci U S A 83:3746–3750
19. Allawi HT, SantaLucia J Jr (1997) Thermodynamics and NMR of internal G.T mismatches in DNA. Biochemistry 36:10581–10594
20. Bichet A, Bureik M, Lenz N, Bernhardt R (2004) The "Bringer" strategy: a very fast and highly efficient method for construction of mutant libraries by error-prone polymerase chain reaction of ring-closed plasmids. Appl Biochem Biotechnol 117:115–122

Chapter 10

Insertion and Deletion Mutagenesis by Overlap Extension PCR

Jehan Lee, Myeong-Kyun Shin, Dong-Kyun Ryu, Seahee Kim, and Wang-Shick Ryu

Abstract

Mutagenesis by the overlap extension PCR has become a standard method of creating mutations including substitutions, insertions, and deletions. Nonetheless, the established overlap PCR mutagenesis is limited in many respects. In particular, it has been difficult to make an insertion larger than 30 nt, since all sequence alterations must be embedded within the primer. Here, we describe a rapid and efficient method for creating insertions or deletions of any length at any position in a DNA molecule. This method is generally applicable, and therefore represents a significant improvement to the now widely used overlap extension PCR method.

Key words: Site-directed mutagenesis, Overlap extension PCR, Insertion, Deletion, chimeric primers

1. Introduction

Various PCR-based site-directed mutagenesis protocols have been described. In particular, a procedure called "overlap extension PCR mutagenesis" has been established as a standard method (1–7). It represents a powerful method of creating mutations such as substitutions, insertions, or deletions without needing restriction sites and involves generation of PCR products overlapping each other, as its name implies. The overlap extension PCR mutagenesis is essentially composed of two sequential steps of PCR. In the first step, typically two pairs of primers are used to generate DNA fragments with overlapping ends containing the desired alteration. These two DNA fragments are then mixed and annealed to get hybrid duplexes. In the second step, the resulting hybrids are then extended and amplified to yield recombined PCR products that can be directly cloned into an appropriate plasmid vector

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_10, © Springer Science+Business Media, LLC 2010

following restriction digest. Although it is simple and efficient, the established overlap PCR mutagenesis is still limited in many respects. For instance, to make insertions, all sequence alterations must be embedded within the primer (1, 2, 8). Therefore, insertions larger than 30 nt. or so have been difficult to create owing to the limitations of oligonucleotide synthesis.

Here, we described an "overlap extension PCR mutagenesis" method that overcomes aforementioned limitations and can therefore be used to create a mutant with any length of insertion or deletion virtually at any position of a DNA fragment (9). To make a larger insertion, an insertion cassette is prepared separately by a PCR using a pair of chimeric primers. The resulting insertion cassette is then recombined to two flanking fragments in the second step of PCR. Another novel feature of this procedure lies in primer design, in that chimeric primers with a fixed length (i.e., 27mer) are used in the procedure, regardless of the insertion size. Chimeric primers are composed of two DNA sequence segments: (1) 18 nt sequence derived from the DNA template and (2) 9 nt sequence derived from the DNA sequences to be added as a "tail" (see Fig. 1). The idea underlying the design of 27mer chimeric primers is twofold. First, 18 nt is the minimal length of the primer for efficient PCR amplification. The 9 nt tail at the 5′ end of chimeric primer makes up the overlap region, which is a hallmark of the overlap extension PCR method (see Fig. 1). Secondly, the 18 nt overlap segment, made up of two 9 nt tails derived from two adjacent PCR fragments, is long enough to facilitate hybrids formation that can be extended and amplified during the second step of PCR (see Fig. 2). Consequently, having 18 nt of homologous region allows both PCR steps to be carried out under an identical annealing temperature (55°C). A similar strategy applies to the design of chimeric primers for a deletion (see Fig. 3). By using this method, we have successfully generated mutants with insertions or deletions of up to a few kb (9–14).

2. Materials

Most reagents available in a standard molecular biology laboratory can be used. Reagents directly related to PCR mutagenesis are listed here.

1. PCR polymerase (see Note 1): *Pfu* or equivalent high-fidelity *Taq* DNA polymerase.
2. DNA template (see Note 2): Plasmid DNA prepared by either PEG precipitation or commercial resins such as QiaQUICK (Quiagen) can be used. DNA is resuspended in Tris–EDTA buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA).

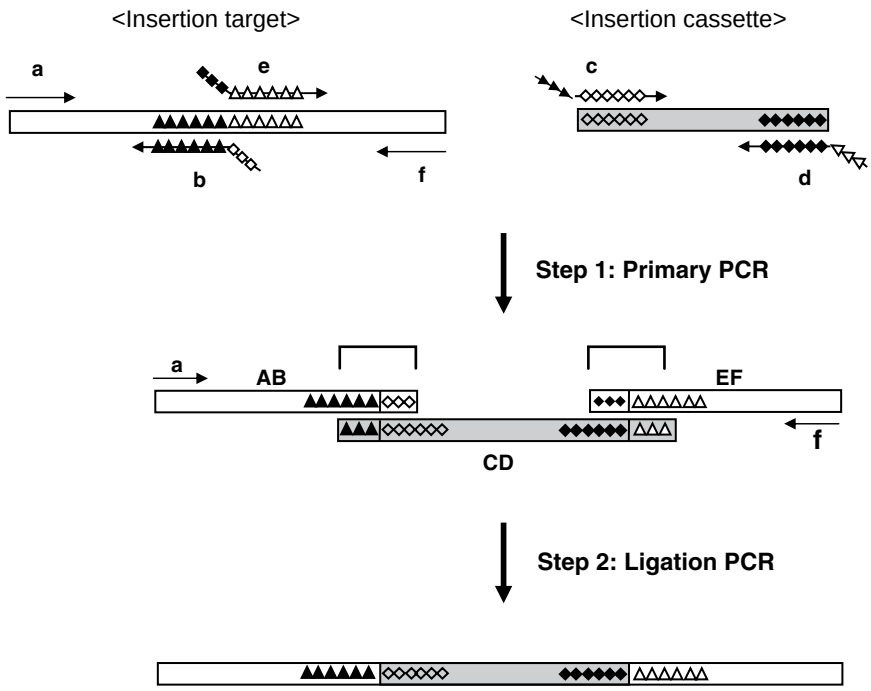


Fig. 1. Schematic diagram of an overlap extension PCR protocol for insertion mutagenesis. The features of an insertion cassette (*gray*) are denoted by diamonds (*open or closed*), whereas those of flanking regions are denoted by triangles (*open or closed*). Chimeric primers of 27 nt (e.g., **b** and **e**) in size are composed of 18 nt derived from the template (*triangles*) and 9 nt from the sequence (*diamonds*) to be added: note that each diamond or triangle represents 3 nt sequence. The 3′ end of the primer is denoted by an arrow. The site where an insertion cassette will be inserted is demarcated by two neighboring triangles with different shading (i.e., *open or closed*). Three PCR products (i.e., AB, CD, and EF) are prepared first by separate reactions with appropriate primer pairs: for instance, the DNA fragment AB is a PCR product by primer **a** and **b**, and so forth. Note that the resulting PCR fragments harbor the 18 nt overlap region (*bracket*) at either one (AB and EF) or both ends (CD). In the second step, PCR is performed with the outermost flanking primer pair (i.e., **a** and **f**) using a mixture of the above three PCR fragments as template. Due to their terminal complementarities in the overlap region, these products will anneal and subsequently be extended during the first cycle of the second step of PCR (see Fig. 2). This second step of PCR is often called “a ligation PCR”, as it essentially recombines DNA fragments.

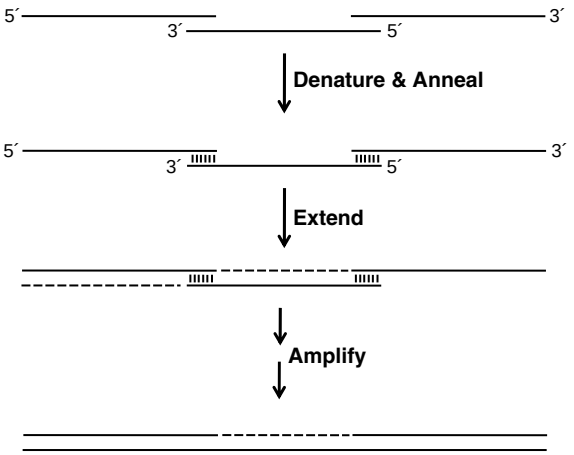


Fig. 2. A diagram illustrating the proposed intermediate steps occurring during the second step of PCR of the insertion mutagenesis. The denatured PCR fragments anneal at the overlap region via terminal complementarities. The resulting hybrid DNA molecules are extended to 3′ by *Taq* DNA polymerase (*dotted line*) to form a ligated duplex. Subsequent PCR cycles amplify the ligated duplex.

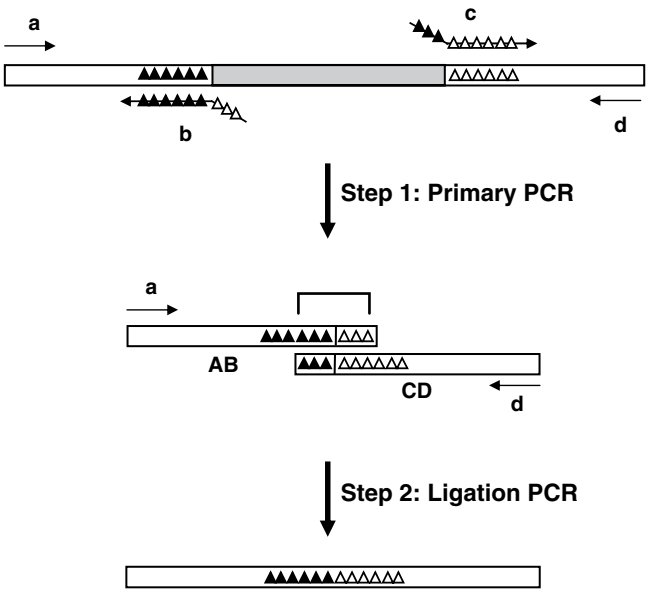


Fig. 3. Schematic diagram of an overlap extension PCR protocol for deletion mutagenesis. Two PCR products representing the flanking regions of the sequence to be deleted are prepared by using one nonchimeric and one chimeric primer: **a** and **b** or **d** and **c**, respectively. In the second step, two PCR products (AB and CD) are used as the template for "a ligation PCR" that contains the outermost primer pair (**a** and **d**). The sequence to be deleted is denoted by *grey*, whereas the flanking regions are denoted by triangles (*open or closed*).

3. 10× PCR reaction buffer: 100 mM Tris–HCl, pH 8.3, 500 mM KCl, 15 mM $MgCl_2$, 0.1% gelatin.
4. dNTPs: (2.5 mM each). Store at –20°C.
5. Primers: Synthetic oligonucleotides are resuspended in water to make up 10 picomoles/μL.
6. Geneclean kit (Bio 101, La Jolla, CA)

3. Methods

3.1. Insertion Mutagenesis Strategy

The strategy employs an established overlap extension PCR protocol (2, 3), but involves three templates including an insert fragment (see Fig. 1). In the first step, three PCR products are prepared separately, including one insert fragment and two flanking fragments using appropriate primer pairs. The two flanking fragments thus generated harbor the overlap region at one end (denoted by a bracket in Fig. 1), whereas the insert fragment contains the overlap region at both ends. In the second step, when these intermediate products are mixed together, melted, and reannealed, three DNA strands form a complex that

can be extended by *Taq* polymerase (see Fig. 1). The resulting extended molecule is amplified in subsequent PCR cycles by using the outermost primer pair.

3.2. Primer Design

1. Six primers are required that include four chimeric primers (i.e., primer **b**, **c**, **d**, and **e**) and two nonchimeric flanking primers (i.e., primer **a** and **f**).
2. Sequences of four chimeric primers are derived partly from the insert and partly from the template. To generate a PCR fragment harboring sequence derived from the insert at its end (i.e., AB and EF), chimeric primers with 27 nt (i.e., primer **b** and **e**) are designed such that 18 nt of sequence are derived from the template and 9 nt are derived from the insertion cassette (see Fig. 1). Care should be taken to design the primer with a correct polarity. The 18 nt sequence should be placed at the 3′ side, while the 9 nt sequence is placed at the 5′ side of the chimeric primer.
3. Likewise, to generate an insert fragment (i.e., CD), two chimeric primers, each 27 nt in size (i.e., primer **c** and **d**) are designed such that 18 nt of sequence is derived from the insertion cassette and 9 nt is derived from either flanking region. This design provides two 9 nt terminal sequences that are complementary and provide overlap with the first PCR product sufficient to make a hybrid template.
4. In addition, the two outermost flanking primers (i.e., primer **a** and **f**) are needed for the second step of PCR. The flanking primers may include restriction sites at the ends to facilitate subsequent cloning of the ligated PCR product.
5. General considerations in primer design also apply, such as avoiding complementarities between and within primers, and so forth.

3.3. First Step of PCR

1. The first step involves three independent PCRs to generate three products that are used as templates for the second step of PCR. Two PCR products from the first step correspond to the flanking regions of an insertion site, whereas the third PCR product corresponds to the desired insertion cassette (see Fig. 1).
2. The flanking fragments (i.e., AB and EF) are prepared by using one nonchimeric primer and one chimeric primer: primer **a** and **b** or primer **e** and **f**, respectively (see Fig. 1). The resulting PCR products AB and EF contain 9 nt derived from an insertion cassette at one end. Likewise, an insert fragment (i.e., CD) is prepared with two chimeric primers (i.e., primer **c** and **d**) (see Fig. 1).
3. Typically, we set up 50 μL reactions with approximately 10–20 ng of template DNA.

Ingredients	Volume
10× PCR buffer	5 µL
dNTPs	4 µL
Primer 1 (10 µM)	2 µL
Primer 2 (10 µM)	2 µL
Template	10–20 ng
Taq polymerase (2.5 U/µL)	0.5 µL
H_2O	to 50 µL

4. Amplify using the following cycle profile. Denature at 98°C for 2 min and perform 25 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 15 s, and extension at 72°C for 1 min (see Note 6) in a thermal cycler. After the last cycle, incubate for a further 10 min at 72°C to allow extension to go to completion.
5. The resulting three PCR fragments are gel-purified and used as template for the second step of PCR.

3.4. Second Step of PCR

1. The second step of PCR is to combine the three PCR fragments prepared above. In this second step, the three products from the first PCR are annealed owing to terminal complementarity (see Fig. 2). The resulting hybrid molecules are extended and amplified during the second PCR. During this overlap extension PCR, the recombined product is amplified with the two outermost primers: primer **a** and **f** (see Fig. 1).
2. Typically, we set up 50 µL PCRs with 10–20 ng of each gel-purified fragment.

Ingredients	Volume
10× PCR buffer	5 µL
dNTPs	4 µL
Primer 1 (10 µM)	2 µL
Primer 2 (10 µM)	2 µL
Template	10–20 ng
H_2O	to 50 µL
Taq polymerase (2.5 U/µL)	0.5 µL

3. Amplify using the following cycle profile. Denature for 2 min 98°C and perform 25 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 15 s, and extension at 72°C for 1 min (see Note 6) in a thermal cycler. After the last cycle, incubate

for a further 10 min at 72°C to allow extension to go to completion.

4. The resulting ligated PCR fragment is gel-purified and then subjected to restriction enzyme digestion and cloning into your choice of plasmid, according to standard recombinant DNA procedures (10).

3.5. Deletion Mutagenesis

3.5.1. Strategy

The strategy for deletion mutagenesis employs a similar overlap extension PCR protocol as above, but involves two PCR fragments, instead of three. The first step involves two independent PCRs to generate two PCR fragments that are used as template for the second step of PCR (see Fig. 3). Two PCR fragments from the first step correspond to the flanking regions of a deletion cassette. In other words, deletion is essentially a ligation or recombination of two separate DNA fragments.

3.5.2. Primer Design

(a) Four primers are required that include two chimeric primers (i.e., primer **b**, and **c**) and two nonchimeric primers (i.e., primer **a** and **d**).

(b) Sequence of two chimeric primers is derived partly from the template to be amplified and partly from the fragment to be recombined following deletion. Specifically, chimeric primers are designed to contain 18 nt derived from the template sequences and 9 nt from the fragment to be recombined (see Fig. 3).

(c) In addition, two outermost flanking primers (i.e., primer **a** and **d**) are needed for the second step of PCR. The flanking primers may include restriction sites at the ends to facilitate subsequent cloning of the ligated PCR product.

(d) General considerations in primer design also apply, such as avoiding complementarity between and within primers, and so forth.

3.5.3. First Step of PCR

1. The first step involves two independent PCR reactions to generate two PCR products that are used as a template for the second step of PCR. Two PCR products from the first step correspond to the flanking region of the deletion cassette.
2. The flanking fragments are generated by using a pair of one nonchimeric primer and one chimeric primer: primer **a** and **b** or primer **d** and **c**, respectively (see Fig. 3). The resulting PCR products AB and CD contain a "tail" of 9 nt that is derived from the fragment to be recombined.
3. Typically, we set up 50 μL reaction with 10~20 ng of template DNA.

Ingredients	volume
10× PCR buffer	5 μL
dNTPs	4 μL
Primer 1 (10 μM)	2 μL
Primer 2 (10 μM)	2 μL
Template	10 ng
Taq polymerase (2.5 U/μL)	0.5 μL
H_2O	to 50 μL

4. Amplify by PCR using the following cycle profile. Denature at 98°C for 2 min and perform 25 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 15 s, and extension at 72°C for 1 min (see Note 6) in a thermal cycler. After the last cycle, incubate for a further 10 min at 72°C to allow extension to go to completion.
5. The resulting two PCR fragments from the first step are gel-purified and used as the template for the second step of PCR.

3.5.4. Second Step of PCR

1. The second step of PCR is to recombine or ligate two PCR fragments prepared above. In the second step, the two PCR products from the first step, owing to terminal complementarity, form an extended DNA product by forming duplexes during the second step of PCR reaction.
2. During this so-called overlap extension, the recombined product is amplified with the two outermost primers: primer **a** and **d**.
3. Typically, we set up 50 μL reaction with 10~20 ng of template DNA.

Ingredients	volume
10× PCR buffer	5 μL
dNTPs	4 μL
Primer 1 (10 μM)	2 μL
Primer 2 (10 μM)	2 μL
Template	10 ng
Taq polymerase (2.5 U/μL)	0.5 μL
H_2O	to 50 μL

4. Amplify by PCR using the following cycle profile. Denature at 98°C for 2 min and perform 25 cycles of denaturation at

98°C for 10 s, annealing at 55°C for 15 s, and extension at 72°C for 1 min (see Note 6) in a thermal cycler. After the lastt cycle, incubate for a further 10 min at 72°C to allow extension to go to completion.

5. The resulting ligated PCR fragments is to be gel-purified and used for cloning following appropriate restriction digest.

4. Notes

1. Care should be taken in the choice of the *Taq* DNA polymerase. *Pfu* (Stratagene) or equivalent high-fidelity *Taq* DNA polymerase (e.g., PrimerSTAR™ HS *Taq* polymerase from TaKaRa) is highly recommended.
2. Care should be taken in DNA template as well. Plasmid DNA that has never been PCR-amplified is highly recommended and this is an important laboratory practice. Otherwise, mutations at one or more sites may be observed. Repeated use of PCR-amplified DNA over the years as a PCR template could unknowingly accumulate errors (mutations) in the DNA template.
3. Regarding the size of PCR products, it is desirable to amplify <2.0 kb PCR products to minimize errors occurring during PCR.
4. Regarding primer quality, desalting by gel filtration over Sephadex column (GE Healthcare) should be fine.
5. To be cost-effective, we routinely use chimeric primers, 27 nt in size. One may use chimeric primers with >27 nt, but not <27 nt.
6. Several precautions should be exercised with respect to PCR. First, PCR should not exceed 25 cycles, since doing so results in extra, unwanted PCR products. Another important variable that could affect the specificity of the PCR is annealing temperature. A general "rule of thumb' is to start with 55°C, when GC content ranges between 40 and 60%. When nonspecific PCR products predominate, the annealing temperature may be increased by 3°C until nonspecific products are reduced to a minimum or eliminated. Conversely, if little PCR product is yielded, the annealing temperature may be decreased by 3°C and so forth. In addition, the minimal amount of DNA template should be used, ranging between 10 and 20 ng/50 μL reaction; otherwise, nonmutated template DNA could cause a high background in subsequent subcloning step.

We have successfully used the overlap PCR method described here to insert an 87 nt sequence harboring a hepatitis delta virus (HDV) antigenomic ribozyme at various positions within a hepatitis B virus (HBV) genomic construct (9, 10). In addition, we have used the above strategy to generate deletion mutations of various length (13, 14).

Acknowledgments

We thank Dr. Byung-Yoon Ahn (Korea University) for critical reading of the manuscript. This work was supported by a grant R01-2006-000-10620-0 from the Basic Research Program of the Korea Science & Engineering Foundation, Republic of Korea.

References

1. Higuchi R, Krummel B, Saiki RK (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res 16:7351–7367
2. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51–59
3. Horton RM (1997) In vitro recombination and mutagenesis of DNA. In: White BA (ed) PCR: cloning protocols: from molecular cloning to genetic engineering, vol 67. Humana, Totowa, NJ, pp 141–149
4. Horton RM, Cai ZL, Ho SN, Pease LR (1990) Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. Biotechniques 8:528–535
5. Horton RM, Ho SN, Pullen JK, Hunt HD, Cai Z, Pease LR (1993) Gene splicing by overlap extension. Methods Enzymol 217: 270–279
6. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene 77:61–68
7. Mikaelian I, Sergean A (1996) Modification of the overlap extension method for extensive mutagenesis on the same template. In: Trower MK (ed) In vitro mutagenesis protocols, vol 57. Humana, Totowa, NJ, pp 193–202
8. Pogulis RJ, Vallejp AN, Pease LR (1996) In vitro recombination and mutagenesis by overlap extension PCR. In: Tower MK (ed) In vitro mutagenesis protocols, vol 57. Humana, Totowa, NJ, pp 167–176
9. Lee J, Lee H-J, Shin M-K, Ryu W-S (2004) Versatile PCR-mediated insertion or deletion mutagenesis. BioTechniques 36:398–400
10. Lee HJ, Lee J, Shin MK, Ryu WS (2008) Polyadenylation is dispensable for encapsidation and reverse transcription of hepatitis B viral pregenomic RNA. Mol Cells 25: 545–552
11. Lee J, Shin M-K, Lee H-J, Yoon G, Ryu W-S (2004) Three novel cis-acting elements required for efficient plus-strand DNA synthesis of the Hepatitis B virus genome. J Virol 78:7455–7464
12. Ryu DK, Kim S, Ryu WS (2008) Hepatitis B virus polymerase suppresses translation of pregenomic RNA via a mechanism involving its interaction with 5′ stem-loop structure. Virology 373:112–123
13. Shin M-K, Lee J, Ryu W-S (2004) A novel cis-acting element facilitates minus-strand DNA synthesis during reverse transcription of the hepatitis B viruses genome. J Virol 78:6252–6262
14. Shin MK, Kim JH, Ryu DK, Ryu WS (2008) Circularization of an RNA template via long-range base pairing is critical for hepadnaviral reverse transcription. Virology 371:362–373

Chapter 11

Targeted Amplification of Mutant Strands for Efficient Site-Directed Mutagenesis and Mutant Screening

Lei Young and Qihan Dong

Abstract

Site-directed mutagenesis is an invaluable tool for functional studies and genetic engineering. However, most current protocols require the target DNA to be cloned into a plasmid vector before mutagenesis can be performed, and none of them are effective for multiple-site mutagenesis. We now describe a method that allows mutagenesis on any DNA template (e.g. cDNA, genomic DNA, and plasmid DNA), and is highly efficient for multiple-site mutagenesis (up to 100%). The technology takes advantage of the requirement that, in order for DNA polymerases to elongate, it is crucial that the 3′ sequences of the primers match the template perfectly. When two outer mutagenic oligonucleotides (oligos) are incorporated together with the desired mutagenic oligos into the newly synthesized mutant strand, they serve as anchors for PCR primers which have 3′ sequences matching the mutated nucleotides, thus amplifying the mutant strand only. The same principle can also be used for mutant screening.

Key words: Mutagenic, 3′ Mismatch, Single-stranded template, Asymmetrical PCR

1. Introduction

Site-directed mutagenesis (SDM) is widely used in molecular biology to study protein structure and functions. It is also used in genetic engineering to optimize enzyme activity and to generate genetically modified species. Various SDM protocols, both PCR and non-PCR based, have been described (1–5). However, none of these protocols is effective for multiple mutagenesis, with reported efficiencies for all the methods being <30% for more than five mutation sites in a single round of reaction. Apart from the PCR-based methods, which can only mutate one site at a time, all other methods require the target DNA to be cloned into a plasmid vector before mutagenesis. Methods that involve

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_11,

virus DNA preparation also require an f1 origin to be present in the plasmid.

In an attempt to overcome the limitations associated with current SDM protocols, we have developed the targeted amplification of mutant strand (TAMS) technology. TAMS technology has proved to be efficient for multiple-site mutagenesis, and can be adapted to directly mutate cDNA, thus eliminating the need for sub-cloning prior to mutagenesis. The protocol takes advantage of the fact that primers with 3′ mismatches usually cannot amplify the target sequence in PCR (6, 7). Therefore, if two mutations that flank the region to be mutated are incorporated into the mutant DNA strand both upstream and downstream of the desired mutations, they can serve as anchors for PCR primers with 3′ sequence matching only the anchor mutations but not the wild-type sequence to preferentially amplify the whole fragment within the anchor mutations. Since the desired mutations are in the mutant strand, and the mutant PCR primers only amplify the mutant strand, the final PCR product will also contain the desired mutations (Fig. 1).

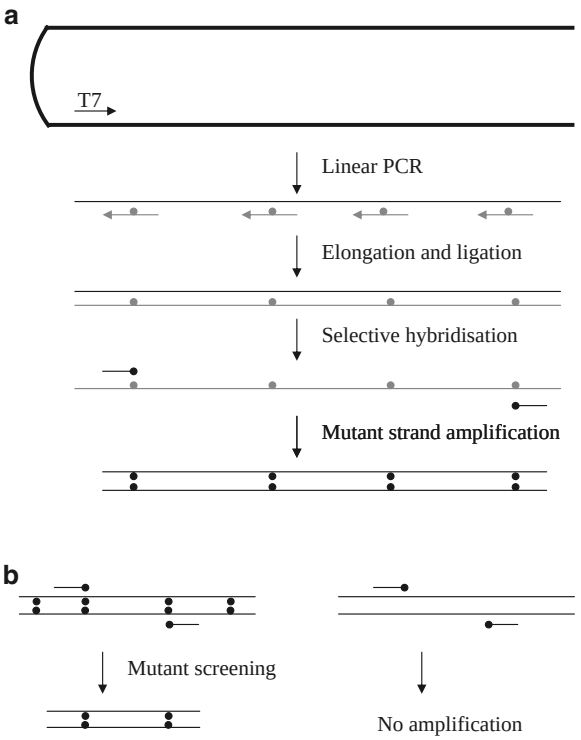


Fig. 1. Schematic diagrams of the TAMS technology. (**a**) ssDNA template was generated by linear PCR on the linearized vector. Approximate positions of the primers are shown with *arrows*. *Blunt arrowheads* mean sequence matching mutant strand. *Large dots* indicate mutated nucleotides. (**b**) The same mechanism can be used to screen mutations from bacteria colonies (Reproduced from NAR with permission).

2. Materials

1. The DNA template can either be a plasmid (preferably having the phage f1 origin of replication), genomic DNA, or first-strand cDNA.
2. Thermostable DNA polymerase (see Note 1), T4 polynucleotide kinase (PNK), T4 DNA polymerase, T4 DNA ligase, and restriction endonucleases: (New England Biolabs; NEB).
3. Superscript II reverse transcription kit and RNase H: (Invitrogen).
4. Primers (Tables 1 and 2) with 5′-OH: (Invitrogen) diluted to a working concentration of 10 µM in sterile water before use.
5. 5× Mutagenesis buffer: 250 mM Tris–HCl (pH 7.5), 50 mM $MgCl_2$, 5 mM ATP, 50 mM DTT, and 1 mM dNTP. This buffer can be made by diluting dNTP solution (Amersham Pharmacia) in the standard 10× T4 DNA ligase buffer.
6. JM109 chemical competent cells and R409 helper phage: (Promega).
7. Reagent and equipment for cloning and transformation into *Escherichia coli* of the mutagenesis product.

Table 1
Primers used for mutagenesis on linear PCR-generated template. Mutated nucleotides are marked in lower case

Name	
Anchor5′	5′-CGAATTCCCGgtaCCGCCA-3′
Anchor3′	5′-TGCAGGCGGCCGCGAgTcgACTAGT-3′
Mut1	5′-TAAGAAGATCaCTATGCCC-3′
Mut2	5′-GCATGCTACTaAGCTTTCA-3′
PCR5′	5′-CCGGCCGCCATGGCGGtac-3′
PCR3′	5′-TGCAGGCGGCCGCGAgTcg-3′
MutPCR1	5′-GGCAACTCAAGGGCATAGt-3′
MutPCR2	5′-GCTCTGAGCATGCTACTa-3′

Table 2
Primers used for mutagenesis on first-strand cDNA. Mutated nucleotides are marked in lower case

Name	
Anchor5′	5′-CGCTCGTCGTCGAtcaCGGCTC-3′
Anchor3′	5′-AAAACCTActgTGCGCA-3′
EcoRV	5′-TGTACGTTGaTATCCAG-3′
XbaI	5′-TCCTGCGTCTaGACCTG-3′
XhoI	5′-GCGCGGCTACAGCT·CGAGCTGCCTGACG-3′
PstI	5′-GCCGACAGGcTGCAGAA-3′
ScaI	5′-AGCGCAAGTAaTCCGTG-3′
ancPCR1	5′-GCGCTCGTCGTCGAtca-3′
ancPCR2	5′-TTGTTTTCTGCGCAcag-3′

3. Methods

3.1. DNA Template Considerations

TAMS technology is very flexible on the template selection. The immediate template for the mutagenic reaction is a single-stranded DNA that can be generated in several ways. If starting with ample amount of double-stranded template, the single-stranded template can be generated by linear PCR with a single primer upstream of the fragment to be mutated. If starting with ample amount of the mRNA of the target gene, the single-stranded template DNA can be generated by first-strand cDNA synthesis, followed by the removal of the RNA strand by RNase H. If starting with a minute amount of DNA, the single-stranded template can be generated by asymmetrical PCR with a larger amount of the forward primer than the reverse primer. Finally, if starting with a plasmid having the phage F1 origin of replication, the circular single-stranded template can be generated by a helper phage such as R408 (8).

3.2. Primer Design

Please make sure the mutagenic oligos are complementary to the template DNA sequence. The mutation(s) should preferably locate in the middle of the mutagenic oligos. It is preferable that at least ten complementary nucleotides are present both upstream and downstream of the mutations, but we found that six nucleotides 3′ and ten nucleotides 5′ to the mutations normally suffice. However, if deletions or insertions are to be made, it is preferable that at least 15 complementary nucleotides are present at each end.

The anchor mutagenic oligos should be designed to incorporate at least three consecutive mutations to increase the discrimination of the mutant strand and the template strand in the selective PCR step. The 3′ nucleotides of the anchor PCR primers should match the mutations incorporated. For example, if the template sequence is 5′-nnnnTCAnnnn-3′, the anchor mutagenic primers may have the sequence 5′-nnnnGACnnnn-3′, and the anchor PCR primers should have the sequence 5′-nnnnGTC-3′.

The efficiency of mutagenesis relies on efficient annealing of the mutagenic oligos to the template. For efficient annealing, the G + C content of the oligos should be 40–50%. If the GC content of the target region is too high, the template DNA tends to self-fold into secondary structures before the oligos can anneal. In such circumstances, longer oligos may increase the mutagenic efficiency.

3.3. Preparation of Single-Stranded Template

1. Preparation by linear PCR.

   To prepare the template by linear PCR, at least 100 ng of the double-stranded template should be used. The only primer used should be 5′ to the 5′ anchor mutagenic oligo. Mix 100 ng of the double-stranded template, 5 µL of 10× PCR buffer, 200 µM of dNTP, and 250 nM of the forward primer in a final volume of 50 µL in a 200-µL PCR tube. Perform thermal cycling with the following conditions: 94°C for 2 min, followed by 30 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 1 min/kb for 30 cycles. An example of mutagenesis with template prepared by linear PCR is shown in Fig. 2.

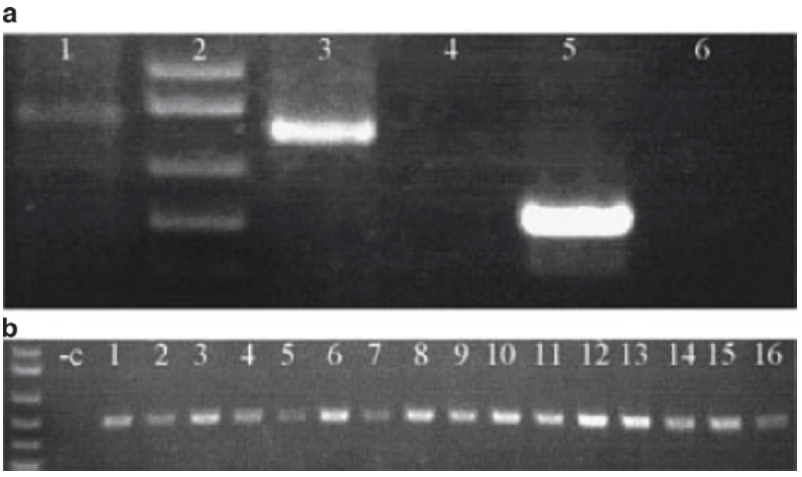


Fig. 2. Mutagenesis on template generated by linear PCR. (**a**) Single-stranded template was amplified by linear PCR (*lane 1*), after mutant strand synthesis, primers PCR5′ and PCR3′ amplified the mutant strand (*lane 3*) rather than the wild-type plasmid (*lane 4*); the same is true with primers MutPCR1 and MutPCR2 (*lanes 5 and 6*). (**b**) PCR product shown in *lane 3* of (**a**) was cloned into pGEM-T Easy and bacteria were transformed. The mutant primers PCRMut1 and PCRMut2 amplified all of the 16 white colonies selected (1–16), but not the wild-type vector (-c).

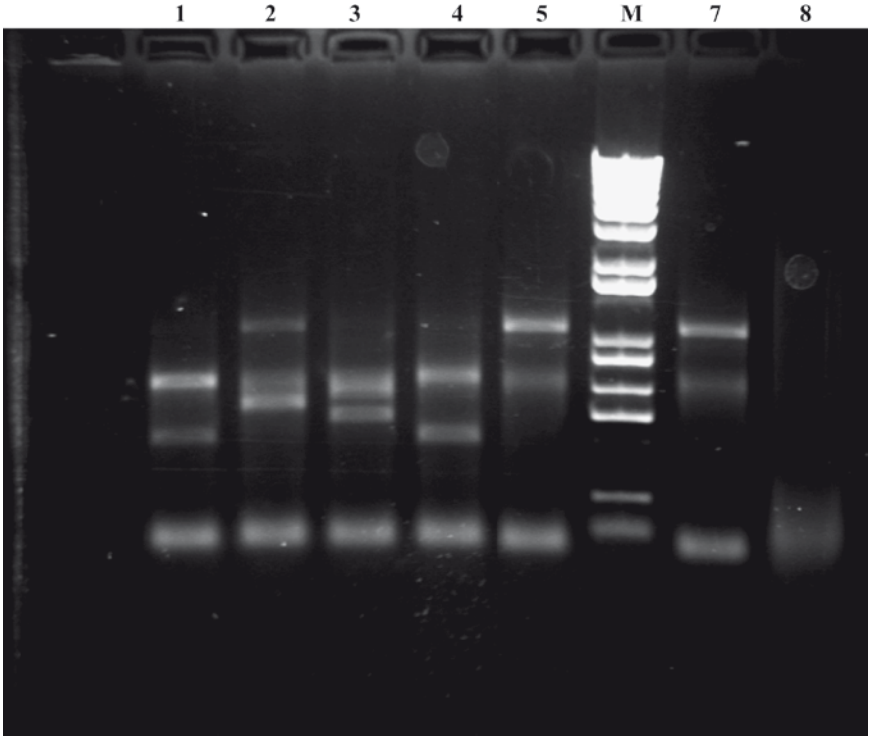


Fig. 3. Mutagenesis on first-strand human β-actin cDNA. The mutagenic primers introduced four restriction sites and deleted one site (ScaI), the mutations include one that deleted 120 bp (XhoI). The anchor PCR primers amplified the mutant strand (*lane 7*), but did not amplify the template cDNA (*lane 8*). The PCR product was subjected to restriction digestion analysis (*lanes 1–5*: EcoRV, XbaI, XhoI, PstI, ScaI). It can be seen that most of the mutations are nearly complete. However, the XbaI mutation is not as efficient due to the high GC content in the surrounding context; this can be optimized by using longer mutagenic primers (data not shown).

2. Preparation by first-strand cDNA synthesis.

   Starting with 5 μg of total RNA, prepare first-strand cDNA in a 20-μL reaction with the Superscript II reverse transcription kit, following the manufacturer's protocol. Add 1 μL of RNase H (2 units) and incubate at 37°C for 20 min. Heat-inactivate the enzymes by incubation at 70°C for 10 min. An example of mutagenesis with first-strand cDNA as template is shown in Fig. 3.

3. Preparation by asymmetrical PCR.

   Asymmetrical PCR should only be used when an insufficient amount of double-stranded template is available. The primers used should be outside of both the 5′ and 3′ anchor mutagenic oligo. Mix 1 μL of the template, 5 μL of 10× PCR buffer, 200 μM of dNTP, 250 nM of the forward primer, and 25 nM of the reverse primer in a final volume of 50 μL in a 200-μL PCR tube. Perform thermal cycling with the conditions: 94°C for 2 min, followed by 30 cycles of 94°C for

30 s, 56°C for 30 s, 72°C for 1 min/kb for 30 cycles, and a final cycle of 72°C for 5 min.

3.4. Phosphorylation of Mutagenic Oligonucleotides

Mix 5 µL of 5× mutagenic buffer, 500 nM each of the inner mutagenic oligos, 100 nM each of the anchor mutagenic oligos, and 1 µL of T4 PNK in a final volume of 22 µL, and incubate the mixture at 37°C for 30 min (see Note 2).

3.5. Mutant Strand Synthesis

1. Anneal the mutagenic oligos by heating 1 L of water in a beaker to 80°C (see Note 3). To the reaction mixture under Subheading 3.4, add 1 µL of the single-stranded DNA template (see Note 4), and place the tube in the beaker. Allow the water to slowly cool to room temperature.
2. Extension and ligation is performed by adding 1 µL of T4 DNA polymerase and 1 µL of T4 DNA ligase to the mixture. Incubate at 37°C for 1 h.

3.6. Mutant Strand Amplification

In this step, the mutant strand but not the wild-type template strand is selectively amplified through anchor PCR primers whose 3′ nucleotides match the mutations in the anchor mutagenic oligos. The anchor mutagenic oligos can be designed to introduce restriction sites, or designed to be outside certain restriction sites that will be used for cloning.

1. Mix 1 µL of the product of Subheading 3.5, step 2, 5 µL of 10× PCR buffer, 200 µM of dNTP, and 250 nM each of the anchor PCR primer in a final volume of 50 µL in a 200-µL PCR tube.
2. Perform thermal cycling with the following conditions: 94°C for 2 min, followed by 30 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 1 min/kb for 30 cycles, and a final cycle of 72°C for 5 min.

3.7. Cloning and Mutant Screening

The PCR amplification product of the mutant strand can be cloned by any established protocols. The same selective PCR strategy used for mutagenesis can also be used for direct colony screening of the mutant clones (see Note 5).

1. Add 20 µL of LB media into each tube, and pick an *E. coli* colony with a pipette tip and mix thoroughly in the media.
2. Mix 1 µL of the mixture in step 1, 5 µL of 10× PCR buffer, 200 µM of dNTP, and 250 nM each of the mutagenic PCR primer (3′ sequence match mutant strand only) in a final volume of 50 µL in a 200-µL PCR tube.
3. Perform thermal cycling with the following conditions: 94°C for 2 min, followed by 30 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 1 min/kb for 30 cycles, and a final cycle of 72°C for 5 min.

4. Analyze the PCR product on 1% agarose gel, and grow the colony broth that produced the detectable band of the right size.

3.8. Brief Description of Results

Mutageneses on both linear PCR-generated template (Fig. 2) and cDNA template (Fig. 3) were carried out. Figure 2 shows the results of simultaneous mutation of four sites (including the anchor mutations). The mutagenic primers only amplified the mutant strands but not the wild-type templates (Fig. 2a). After cloning, all 16 clones picked contained both of the desired mutations, confirmed by both mutagenic primer PCR screening (Fig. 2b) and sequencing (data not shown). Figure 3 shows the results of simultaneous mutation of seven sites (including the anchor mutations). The five inner mutations include both substitutions and deletions, which introduced four restriction sites (EcoRV, XbaI, XhoI, PstI) and eliminated one site (ScaI). Restriction digestions on the PCR product amplified by the anchor mutagenic primers confirmed that all but one (mutation that introduced the XbaI site) mutations are near 100% efficient (cleaved by EcoRV, XbaI, XhoI, PstI, and uncleaved by ScaI). Therefore, when mutagenic oligos are properly designed (see Note 6), the method described here can be 100% efficient for the simultaneous mutation of multiple sites.

4. Notes

1. The preferred thermal stable DNA polymerase in the TAMS technology is one that lacks $3' \rightarrow 5'$ exonuclease activity, such as the Taq DNA polymerase. $3' \rightarrow 5'$ exonuclease activity potentially can digest away the $3'$ non-complementary sequences of the mutagenic PCR primers, thus decreasing mutant strand selection specificity. However, if fidelity is the major concern, a high-fidelity enzyme such as the Phusion DNA polymerase should be used. In such cases, more mutations should be incorporated into the anchor mutagenic oligos to delay the full digestion of the non-complementary sequences. Furthermore, even when the non-complementary sequences are fully digested, the shorter oligos will have lower optimal annealing temperatures. Therefore, the mutant strand will still be preferentially amplified.
2. The phosphorylation of oligonucleotides by T4 PNK is very efficient; therefore, synthesis of $5'$-phosphorylated oligos is not necessary. T4 PNK has low phosphorylation efficiency only on double-stranded polynucleotides.

3. The starting annealing temperature should not exceed 80°C. This temperature is sufficient to melt secondary structures in the single-stranded template. Higher temperature will also melt the double-stranded template. Since annealing of multiple mutagenic oligos to partially denatured double-stranded template is very inefficient, it reduces the overall mutagenic efficiency.
4. We have found that purification of the single-stranded template is not necessary. However, if the DNA polymerase used to generate the template has strand displacement activity, the residual activity can potentially extend from one mutagenic oligo and displace downstream oligos, thereby reducing the mutagenic efficiency. Therefore, phenol–chloroform purification of the template is recommended.
5. As shown in Fig. 2a, lane 6, one nucleotide at the 3′ end of the PCR primer is sufficient to discriminate between the mutant strand and the wild-type strand.
6. When designing mutagenic primers, it is crucial that both the 5′ and 3′ 10–12 nucleotides have a GC content of lower than 50%. High GC content favors internal annealing. If the region to be mutated self-anneals before the mutagenic primers can anneal to it, then the mutagenic efficiency for this site will drop, this is what happened in Fig. 3, lane 2, as the mutagenic primer for this site had a high GC content of 59%. Longer primers can be designed to ensure that the terminal nucleotides at both the 5′ end and the 3′ end have GC contents that are lower than 50%.

References

1. Ishii TM, Zerr P, Xia XM, Bond CT, Maylie J, Adelman JP (1998) Site-directed mutagenesis. Methods Enzymol 293:53–71
2. Ling MM, Robinson BH (1997) Approaches to DNA mutagenesis: an overview. Anal Biochem 254:157–178
3. Braman J (ed) (2002) In vitro mutagenesis protocols, 2nd edn. Humana, Totowa, NJ
4. Kunkel TA, Roberts JD, Zakour RA (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol 154:367–382
5. Sawano A, Miyawaki A (2000) Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. Nucleic Acids Res 28:e78
6. Cha RS, Zarbl H, Keohavong P, Thilly WG (1992) Mismatch amplification mutation assay (MAMA): application to the c-H-ras gene. PCR Methods Appl 2:14–20
7. Huang MM, Arnheim N, Goodman MF (1992) Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. Nucleic Acids Res 20:4567–4573
8. Promega Corporation (1996) Protocols and applications guide, 3rd edn. Promega Corporation, Madison, WI

Chapter 12

A Modified Inverse PCR Procedure for Insertion, Deletion, or Replacement of a DNA Fragment in a Target Sequence and Its Application in the Ligand Interaction Scan Method for Generation of Ligand-Regulated Proteins

Oran Erster and Moti Liscovitch

Abstract

Functional analysis of a protein of interest, by generation of functional alterations in a target protein, often requires the performance of site-directed mutagenesis within the gene sequence. These manipulations are usually performed using "cut and paste" techniques, combined with PCR. Here we describe a simple and general procedure to specifically insert a DNA fragment into any site within a given DNA sequence. We demonstrate this insertional mutagenesis by describing the insertion of a tetracysteine (4C) hexapeptide-encoding sequence into the coding sequence of the antibiotic hydrolyzing enzyme TEM-1 β-lactamase. This procedure could also be applied to insert different DNA sequences or to replace, or delete, existing fragments in a given gene. We have recently used this procedure to develop a general method (ligand interaction scan – LIScan) to generate ligand-regulated proteins.

Key words: Site-directed mutagenesis, Restriction enzymes, Inverse PCR, Insertional mutagenesis, QuickChange, Ligand interaction scan

1. Introduction

The sequenced genomes of a growing number of organisms, from bacteria to humans, provide scientists with an enormous number of newly discovered genes and gene products. Studying their function by domain dissection (1), their expression as recombinant genes in different organisms (2), or their use for applied purposes (3), are standard procedures in basic research and the biotechnological industry. Analysis of a protein's function by insertional mutagenesis can be carried out by taking advantage of a naturally occurring transposing element, or by randomly

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_12, © Springer Science+Business Media, LLC 2010

inserting a foreign DNA into an exon. The first approach utilizes a transposon, which randomly inserts a DNA sequence into the target gene (4). In the second approach, target cells are infected with a retrovirus, which randomly integrates the inserted sequence into the host cell DNA (5). Both approaches enable a random insertion into the target gene/genome, a subsequent identification of the insertion location, and analysis of the mutated protein. However, other applications may require an accurate manipulation of the target gene. The early "genetic engineering" manipulations were done using "cut and paste" techniques, taking advantage of bacterial restriction endonucleases and ligases, to excise and fuse different genetic sequences (6, 7). The number of possible alterations using this technique was dictated by the recognition sites of the endonucleases, and it was therefore limited in its applicability. The invention of the PCR procedure (8) further expanded the range and ease of genetic manipulations, including the incorporation of synthetic sequences into cloned genes. Recently, commercial kits enable site-directed mutagenesis using improved enzymes and optimized protocols. The QuickChange® protocol (Stratagene) and its derivatives are widely used for point mutagenesis and point deletions (9). However, when insertional mutagenesis is desired, this protocol may be limited to short insertions. The use of primers longer than 70 bp may result in the formation of a secondary structure of the primer, forcing a high annealing temperature, and increasing the chances of PCR errors, or, alternatively, rendering the reaction unsuccessful (10). The mutagenesis procedure described below is based on the "inverse" PCR (11) method. It enables the site-directed insertion of any length of foreign DNA into an existing sequence, and its implementation in the procedure described here also enables a rapid verification of the presence of the inserted sequence. This procedure has proved useful for inserting a ligand-binding domain into a target protein (12). Yet, it may readily be applied to insertions of functional domains, cleavage sites, tags, or other desired sequences into a target gene, or adding restriction sites and regulatory elements into a target plasmid.

2. Materials

2.1. Inverse PCR and Plasmid Self-Ligation

All commercial materials are stored according to the manufacturer's instructions.

1. Restriction enzymes: Fermentas, or New-England Biolabs (NEB).
2. T4 DNA ligase: Fermentas or NEB.
3. Phusion DNA polymerase, including reaction buffers HF and GC: Finnzyme.

4. *Pfu* Turbo DNA polymerase: Stratagene.
5. dNTPs (100 mM stock each) and ATP (100 mM stock): Fermentas. Preparation of dNTPs mix: Mix the 100 mM stocks in H_2O to a final concentration of 10 mM each by taking 10 µl from each dNTP stock (40 µl together) and 60 µl H_2O. This 10 mM dNTPs mix will be used further on.
6. T4 Polynucleotide kinase (PNK enzyme and supplemented buffer): Fermentas.
7. PCR ReadyMix: Sigma.
8. Primers: Weizmann Institute or Integrated DNA Technologies (IDT) and are stored at –20°C until use.
9. Gel/PCR DNA extraction kit: Real-Biotech Corporation (RBC).
10. DNA miniprep kit: RBC.
11. 1 kb GeneRule™ DNA size marker: Fermentas.
12. Ethidium bromide, 1% solution (EtBr): stored at room temperature.

2.2. Plasmid Propagation in E. coli Cells

1. Heat-shock – competent XL1B *E. coli* cells (Stratagene, aliquoted and stored at –80°C) are from the laboratory collection.
2. Liquid Lauria-Bertani (LB) broth (10 g tryptone, 10 g NaCl 5 g yeast extract) is supplied from the bacteriology services at the Weizmann Institute and stored at 4°C.

Preparation of LB agar plates: BactoAgar (0.8 g per 100 ml liquid LB) is dissolved and autoclaved at 121°C for 20 min. After autoclaving and cooling down to ~55°C, appropriate antibiotics are added and the mix is poured into 10-cm plates (Miniplast, Ein-Shemer, 20–30 ml for each plate). Plates are dried in a sterile hood and stored at 4°C until use.

2.3. Screening PCR for Insertion-Containing Colonies

1. PCR ReadyMix: Sigma. Reaction was set according to the manufacturer's instructions.

2.4. Preparation of Tris–Acetate EDTA (TAE) Buffer and TAE Agarose gel

Stock solution of 50× TAE for 1 L: Tris base, 242 g, Glacial acetic acid, 57.1 ml, 0.1 M EDTA, 200 ml, H_2O, Upto 1 L.

The stock 50× TAE is diluted 50-fold for use. For gel agarose, 1 g of agarose is added per 100 ml of 1× TAE in a covered glass bottle, and warmed in a microwave oven until the agarose is dissolved. *Caution*! Overheating may cause the solution to boil and burst the bottle or dislodge the cap causing agarose to spill out. Take out the bottle carefully using a heat-resistant glove. The bottle is cooled for at least 10 min at room temperature. Alternatively, after heating, the bottle could be placed in a 55°C heated bath, for 10 min. EtBr (2–3 µl for each gel) is carefully

added before casting the gel. The gel mix is then poured into the gel cast apparatus. The gel is cooled for 10–20 min until it is solidified, before running the samples.

2.5. Specific Equipment

1. PCR thermocycler: Bio-Rad MyCycler™.
2. Dry heat block.
3. Nucleic acid gel camera/imager.
4. Shaker incubator set for 37°C.
5. UV-illuminated table.
6. NanoDrop spectrophotometer (optional).

3. Methods

3.1. Overview

In this protocol, insertion (or replacement) of a foreign DNA fragment into an existing sequence is performed by an inverse PCR reaction (11). The reaction is designed to amplify the entire plasmid, with the addition of the foreign DNA at the desired location dictated by the primers. The resulting product is a linear plasmid, which is self-ligated following the mutagenesis reaction. The template plasmid is then degraded, and the PCR-generated plasmid is transformed into *E. coli* cells. The resulting colonies are then rapidly screened for the presence of the inserted sequence by PCR. Positive colonies are grown, and the plasmid is extracted and analyzed by restriction enzyme analysis to confirm the presence of the insertion. We used this procedure to generate ligand-modulated mutants of a target protein (see Subheading 3.13). Similarly, other insertions or replacements may be performed, introducing or deleting unique restriction sites. The presence or absence of such sites, compared to the template plasmid, can be used to rapidly identify positive mutant plasmids prior to DNA sequencing.

3.2. Primer Design

Appropriate primer design is crucial to obtain successful mutagenesis. Each primer should include two parts:

(a) In the 5′prime primer, the 3′ portion (at least 24 bases) should be homologous to the template, precisely matching the region downstream to the insertion site. The 5′ prime part should contain half of the inserted sequence. In our implementation, it is the sequence encoding the GCC peptide (5′-GGC TGC TGC). Primer length may vary from 35 to 60 or more, depending on the insertion and the complementary region.

(b) The 3′ prime primer should include in its 5′ portion the inverted sequence that encodes the second half of the inserted sequence.

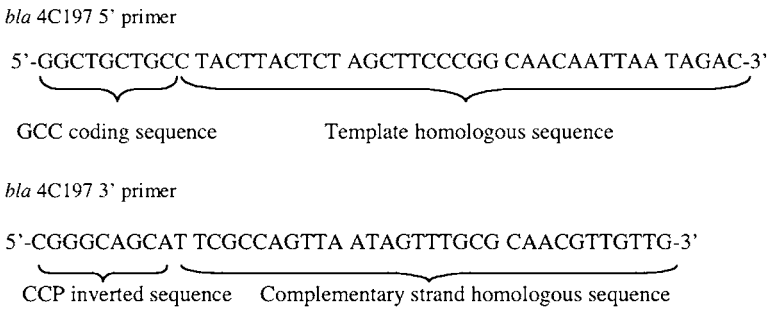


Fig. 1. 4C197 insertion primers. The primers used to insert the 4C sequence downstream of amino acid 197 in the (*beta lactamase*) gene are shown. The 5′ prime end of each primer encodes one half of the CCPGCC peptide. The 3′ prime end of each primer is complementary to the template gene.

In our implementation of the procedure, it is the inversion of the sequence encoding CCP peptide (5′-CGG GCA GCA – inverted sequence of TGC TGC CCG). The 3′ prime end of the primer should contain at least 24 bases homologous to the template strand (corresponding to the 5′ primer's template strand). An example of the primers used for 4C insertion downstream of position 197 in the *bla* gene (encodes TEM-1 β-lactamase) is shown in Fig. 1.

Once the primers are designed, they should be examined according to the following criteria:

1. Minimized non-specific binding on the template plasmid. This may be done using the BLAST website (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi), or any other sequence alignment tool. We used the programs Gap and Bestfit from the Genetics Computer Group software package (GCG software package, http://www-biocomp.doit.wisc.edu/gcg.shtml).
2. Minimal secondary structure formation resulting from self-folding. This test may be done using the program Foldrna from GCG.
3. In-frame translation of the insertion/replacement/deletion within the protein sequence. This may be done with any program that translates nucleotide sequence to peptide sequence. We used DNA Strider (Christian Marck, Service de Biochimie et de Génétique Moléculaire Bat.142 Centre d'Etudes de Saclay, 91191 GIF-SUR-YVETTE CEDEX FRANCE).

3.3. Amplification of the Mutated Plasmid

Generation of a full-length, complete linear plasmid that can subsequently be self-ligated, requires a polymerase enzyme that has a robust activity, enabling it to accurately amplify 6–8 kb, and it should leave blunt ends at the amplified strands (no additional A tail).

We used two such enzymes, *Pfu* Turbo (Stratagene) and Phusion (Finnzyme), both of which worked well. Other enzymes with similar characteristics may also be used (see Notes 1–6).

1. The following mix is prepared in 200 µl PCR tubes (see Notes 1–3 for reaction setup):

For Phusion PCR		**For *Pfu* Turbo PCR**	
Reagent	**Quantity**	**Reagent**	**Quantity**
Template plasmid	20–50 µg	Template plasmid	20–50 µg
DMSO	2.5 µl	DMSO	2.5 µl
Primers	40–55 pmol of each primer	Primers	40–55 pmol from each primer
5× Phusion HF buffer	10 µl	10× *Pfu* buffer	5 µl
dNTPs	1 µl from 10 mM stock mix	dNTPs	1 µl from 10 mM stock mix
Phusion DNA polymerase	0.5 µl	*Pfu* Turbo DNA polymerase	1 µl
PCR grade H_20	To a final volume of 50 µl	PCR grade H_2O	To a final volume of 50 µl

2. The thermocycler is set according to the following parameters (see Notes 4 and 5 for parameters considerations):

	For Phusion PCR				**For *Pfu* Turbo PCR**		
	Step	**Temperature**	**Time (min)**		**Step**	**Temperature**	**Time (min)**
	1.	98°C	2		1.	95°C	2
24 X	2.	98°C	40	24 X	2.	95°C	1
	3.	58°C	1		3.	55°C	1
	4.	72°C	1'40"		4.	68°C	6
	5.	72°	10		5.	72°	12
	6.	10°C	Hold		6.	10°C	Hold

3. The PCR tubes are placed in the thermocycler and the PCR program is immediately run. The PCR will generate a linear plasmid, with each end containing an addition of half of the 4C inserted coding sequence. For schematic description of the

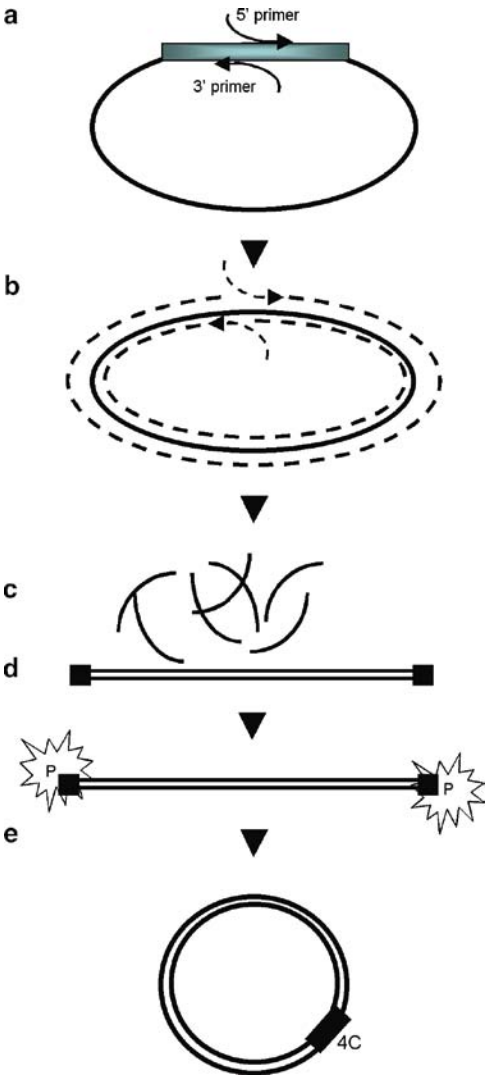


Fig. 2. Insertion of the tetracysteine (4C) encoding sequence using site-directed insertion mutagenesis. (**a**) Primers facing opposite directions anneal to the template plasmid. Each primer contains half of the insertion at the 5′prime end and site complementary sequence at the 3′ prime end. (**b**) Amplification reaction yields a complete, linear plasmid, containing half of the inserted sequence at each end. (**c**) *DpnI* digests the template plasmid, leaving only the 4C-containing PCR product. (**d**) PCR product is phosphorylated with T4 PNK. (**e**) PCR product is self-ligated with T4 DNA ligase. 4C – The complete 4C sequence after plasmid self-ligation.

procedure, see Fig. 2. An example of inverse PCR products is shown in Fig. 3.

4. After the amplification reaction is complete, a sample (2–5 µl from the reaction tube) of the PCR is run in EtBr-containing 1% TAE agarose gel to visualize the product. A full-length, linear plasmid should be seen. If there is no PCR product, see Notes 6 and 7.

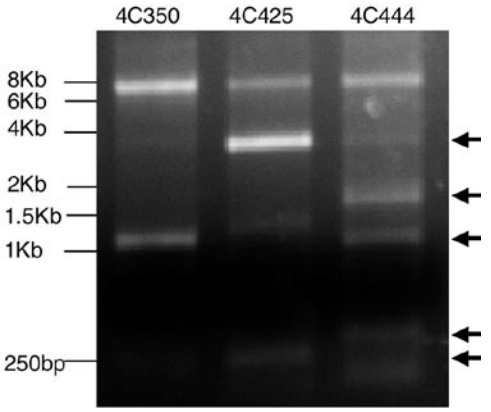


Fig. 3. Gel electrophoresis of Inverse PCR product. Amplification of three different mutants containing 4C-coding sequence, using a 6.33-kb plasmid as template. Numbers above each lane indicate the location of the insertion. Arrows indicate non-specific products.

3.4. Template Plasmid Digestion

1. *DpnI* (2U) is added directly to the PCR tube and the tube is incubated for 2–4 h at 37°C. As *DpnI* digests only methylated DNA, the PCR product will not be digested (Fig. 2).
2. The PCR product is purified by using a PCR purification column (RBC, see Subheading 2.1). The PCR product is eluted with 40 µl of either H_2O or the manufacturer's elution buffer.

3.5. Gel Purification of PCR Product (Optional)

Some amplification reactions may yield non-specific products, in addition to the desired one (see Fig. 3). In order to avoid the presence of the non-specific products in subsequent steps, the PCR products may be separated using TAE agarose gel, according to the following steps:

1. Samples are run in EtBr-free gel, using 90 V, until sufficient separation is obtained (this is dependent on the size difference of the desired vs. non-specific products).
2. The gel is stained with 1% EtBr-containing TAE buffer for 30 min.
3. The gel is examined under UV illumination to identify the band corresponding to the desired PCR product.
4. Using an ultraviolet-illuminated table, the desired fragment is excised from the gel using a scalpel and collected into a 2-ml tube.
5. The desired product is purified from the gel fragment using a gel purification kit.

3.6. 5′ End Phosphorylation of the PCR Product

Unless pre-phosphorylated primers are used, this step is necessary to enable ligation of the PCR product. Phosphorylation of the PCR product is achieved using T4 Polynucleotide Kinase (PNK, Fermentas, see Subheading 2.1).

1. The following mix is prepared in a 1.5-ml tube:

Reagent	Quantity
Eluted PCR products	40 µl
10× PNK buffer	5 µl
ATP from 10 mM stock	4 µl
T4 Polynucleotide kinase (PNK)	1 µl

2. The tube is incubated at 37°C for 60 min.
3. The phosphorylated product is purified as in step 2 of Subheading 3.4.
4. A sample (3–5 µl) of the eluted product is separated in an EtBr-containing 1% TAE agarose gel to verify integrity and quantity. The concentration of the product is quantified, either by spectrophotometry, or by comparing it to a known amount of a DNA size marker that is approximately the same size as the PCR product. For spectroscopy, we used NanoDrop (Thermo Scientific) that requires 1–2 µl of the sample.

3.7. Plasmid Self-Ligation

1. The following reaction is set up in a 200-µl tube:

Reagent	Quantity
10× T4 ligase buffer	2 µl
50–150 ng DNA product	Xµl
ATP from 10 mM stock	1 µl
PEG4000 (supplied with T4 DNA ligase)	1 µl
T4 DNA ligase	1 µl
H_2O	To a final volume of 20 µl

2. The reaction is incubated at 16°C for 16–24 h, or, alternatively, at room temperature for 2 h.

3.8. Second DpnI Digestion (Optional)

To further decrease the chances of obtaining colonies containing the template plasmid, a second *DpnI* digestion may be performed, by adding 1 µl of *DpnI* to the ligation tube, and incubating it at 37°C for 60 min. *DpnI* can then be inactivated by incubating the tube at 80°C for 20 min.

3.9. Heat-Shock Transformation

1. An aliquot tube of XL1B competent cells (see Subheading 2.2) from–70°C stock is placed on ice for 10 min.
2. An aliquot of the ligation reaction (5–8 µl) is inserted into the cell tube, mixed by pipetting and placed on ice for another 20 min.

3. The tube is placed in a 42°C pre-heated heat-block for 45 s and is immediately put back on ice for 1 min. This step should be performed with maximum accuracy. Higher temperature or longer incubation at 42°C may reduce transformation efficiency sharply.
4. LB (1 ml) is added (without antibiotics) to the cells and the tube is incubated for 1–4 h in 37°C for recovery from heat-shock.
5. Cells are centrifuged in 12,000 × *g* for 30 s. The excess LB is discarded, and the cells are re-suspended in 200 µl LB.
6. Cells are plated on LB agar containing the appropriate selection antibiotic and incubated overnight at 37°C.

After 24 h, colonies should be visible to the naked eye. The number of colonies may vary from one to a few hundred. Longer incubation is not recommended, as it may cause the growth of satellite colonies, and may impair the screening for desired colonies. If no colonies are visible after 24 h, see Note 8.

When colonies are formed, it is possible to make starter cultures by choosing colonies randomly, or perform a PCR screening reaction to screen for positive colonies. In our hands, 50–95% of the colonies usually contain the plasmid with the desired insertion. However, we recommend screening the colonies according to the following protocol, to reduce the rate of false positive errors.

3.10. Mutagenesis Screening PCR

This procedure is aimed to detect the mutated colonies, before culturing them for plasmid extraction. In the case of a deletion, screening may be done using one primer complementary to either the 5′ prime or the 3′ prime end of the mutated gene, and a second primer designed to anneal to the deleted fragment. In that case, a product will only be generated from colonies that contain the deleted sequence (i.e., contain the template, rather than the newly synthesized, mutated plasmid). Colonies that test negative in this screen, i.e., *not* containing the template for the deleted primer (therefore containing the mutated plasmid), are used. An insertion or a replacement mutagenesis will require the use of one primer complementary to either the 5′ prime or the 3′ prime end of the mutated gene, and a second primer complementary to the inserted or replaced sequence. In our implementation of the screening PCR, we used one primer either for the 3′ prime or the 5′ prime end of the gene, and a second primer complementary to the 4C encoding sequence, as a marker for the presence of the insertion (see example in Fig. 4a). In our experience, when generating a large number of 4C mutants, this preliminary screen saves a significant amount of effort.

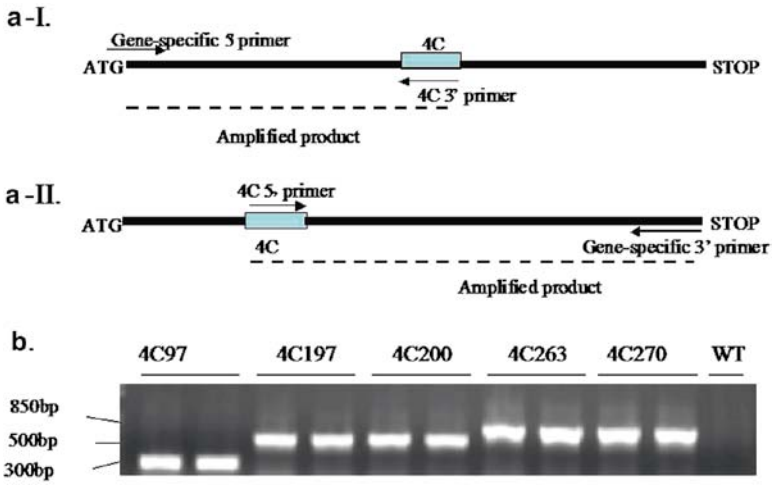


Fig. 4. PCR screen for transformed colonies containing 4C-coding sequence. (**a**) Two screening options are shown: (**a**-I) Using a 5′prime gene-specific primer with a 3′ prime 4C primer. This is recommended when the 4C insertion is at least 250 bp downstream to the 5′ prime primer annealing sequence. (**a**-II) Using 5′ prime 4C primer with a 3′ prime gene-specific primer. This is recommended when the 4C insertion is at least 250 bp upstream to the 3′ prime primer annealing sequence. (**b**) Amplification analysis after colony growth for β-lactamase mutants. PCR was done with 5′ prime primer complementary to the 5′ prime end of the *bla* gene, and 3′ prime primer complementary to the 4C sequence. Mutants' names denote the location in which the 4C was inserted (for example, 4C97 represents insertion downstream to amino acid in position 97).

1. Numbered 1.5-ml tubes containing PCR-grade 20 µl ddH_20 (does not have to be sterile) and numbered PCR 200-µl tubes are prepared.
2. Using a sterile 10 µl tip, a single colony is picked from the agar plate, pipetted up and down in the labeled 1.5 tube, and 2 µl of the suspended colony is transferred to the corresponding PCR tube (i.e., from 1.5-ml tube labeled "1", to the PCR tube labeled "1"). This step is repeated for all the colonies to be screened. Finally, a set of 1.5-ml tubes containing the selected colonies, with corresponding PCR tubes, each containing a 2 µl sample of each colony, is obtained. The 1.5-ml tubes are stored on ice or at 4°C.
3. The PCR Ready Mix is prepared according to the manufacturer's instructions, and a 14 µl reaction per tube is set up. The mix should contain two primers, one complementary to either the 5′ prime or 3′ prime end of the mutated gene, and the other complementary to the mutated sequence (inserted, replaced, or deleted). For considerations regarding the screening PCR primers and procedure, see Notes 9–12.

Mutagenesis screening PCR mix:

Reagent	Quantity
2× ReadyMix solution	7 µl
Template colony	2 µl
Primers	5–10 pmol
H_2O	To a final volume of 14 µl

4. The thermocycler is set according to the following parameters and the PCR program is run.

   Mutagenesis screening PCR program

	Step	Temperature	Time (min)
	1.	95°C	2
X 30	2.	95°C	1
	3.	66°C	1
	4.	72°C	1
	5.	72°	10
	6.	10°C	Hold

5. When the PCR is completed, the samples are separated in EtBr-containing 1% TAE agarose gel, and positive colonies are detected. The desired number of positive colonies is picked from the 4°C-stored tubes, according to the PCR results, for propagation and subsequent plasmid extraction (see Note 12). An example for such screening PCR is shown in Fig. 4b. If no positive colonies are identified, see Note 13.

3.11. Restriction Analysis for Mutated Plasmids

Alterations in the original sequence of the mutated gene may generate or delete restriction sites. Once a mutated plasmid is extracted, it could be analyzed using restriction analysis, to verify the presence and location of the mutated sequence. In our implementation of the mutagenesis procedure, introducing the 4C peptide Cys-Cys-Pro-Gly-Cys-Cys was carried out using the nucleotide sequence TGC TGC CCG GGC TGC TGC, generating an *SmaI* restriction site (CCCGGG). This allowed for a rapid identification of 4C-containing plasmids, by digesting the plasmid with *SmaI* (Example is shown in Fig. 5). Even if a *SmaI* site is present in the parental plasmid (that contains the WT gene), the addition of the new site will alter the restriction pattern and confirm, or exclude, the presence of the 4C encoding sequence. The only limitation of this analysis is when the 4C sequence is inserted very close to an original *SmaI* site. In this case, the digestion will

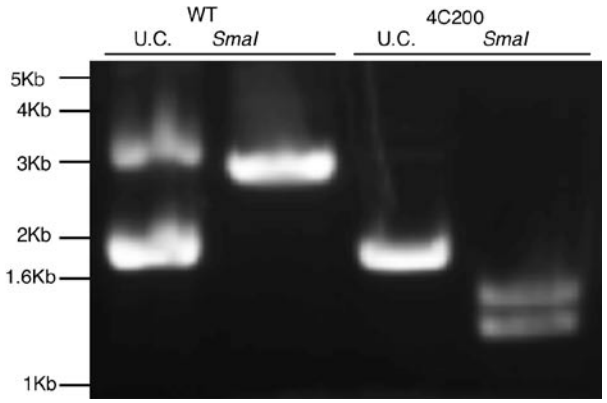


Fig. 5. *SmaI* restriction analysis. pUC18 plasmids harboring either WT or mutant 4C200 β-lactamase genes were extracted by miniprep and digested with *SmaI*. The WT plasmid is linearized upon *SmaI* digestion, while the 4C200 plasmid is digested into two fragments. *UC*: uncut plasmid.

not be informative. It is important to include a parental plasmid (containing the unmutated gene) in the analysis, to compare the restriction pattern of the 4C plasmid to the WT plasmid (Fig. 5).

1. The digestion mix is prepared as follows:

Reagent	Quantity
Plasmid DNA	200–1,000 ng
10× restriction enzyme buffer	2 μl
Restriction enzyme	5–10 U
H_2O	To a final volume of 20 μl

2. Samples are incubated at the appropriate temperature (depending on the restriction enzyme used) for 1 h and the products are separated in an EtBr-containing 1% TAE agarose gel. The mutated plasmid should show a digestion pattern that is distinct from that of the wild-type plasmid (WT). An example of *SmaI* digestion of WT vs. 4C-containing β-lactamase genes in pUC18 is shown in Fig. 3. If a colony that tested positive in the PCR screen is tested negative in the restriction analysis, see Note 14. Sequencing of the mutated plasmid is strongly recommended, even if the PCR screen and the restriction analysis are both as expected (see Note 15).

3.12. Expected Results and Time Schedule

After calibration and optimization of the PCR conditions, it is possible to obtain a purified plasmid containing the mutated gene within 3–4 working days. The mutagenesis PCR, *DpnI* digestion, 5′ phosphorylation, ligation (2 h), and transformation can all be done in one day. The PCR screen for the colonies could be done during the following day, and minipreps could be produced on

the third working day. If a prolonged (overnight) ligation is preferred, another day may be added. In our hands, with three different plasmids and genes, generation of 25–30 mutants may take 15–20 working days. The largest plasmid to which we successfully applied this method was approximately 10 kb, and we assume that it should work with larger vectors as well. We occasionally encountered several PCR sequence errors that were not detectable by the PCR screen, or by the restriction analysis. As much as the Phusion and *Pfu* Turbo polymerases are considered to be practically error-free, compared to other polymerases, elongation of a 6–9-kb plasmid can result in amplification errors (See Note 15 for details). We nevertheless found this method fast and reproducible, enabling the production of a large number of site-directed mutants simultaneously.

3.13. Application of Insertional Mutagenesis PCR in the Ligand Interaction Scan Method

Using the modified inverse PCR procedure described herein, we recently developed a novel, general, and simple procedure for engineering small-molecule ligand-regulated forms of any protein (12). The Ligand Interaction Scan (LIScan) method involves insertion of a chemical-genetic "switch," comprising a genetically encoded peptide module (a tetracysteine-containing hexapeptide) that binds with high affinity to a small-molecule ligand (the cell-permeable biarsenical fluorescein derivative FlAsH, (13)) into a given protein. The insertion position(s) are selected empirically to confer ligand-dependent modulation of activity. Ligand-regulated mutants may then be expressed in cells or inserted genomically, wherein they can be regulated by ligand administration (12).

We assessed the feasibility of the LIScan method by applying it to the TEM-1 β-lactamase antibiotic resistance gene, generating a set of 4C mutants, each containing the insertion in a different site. TEM-1 4C mutants that can be either inhibited or stimulated by FlAsH were engineered. Our results demonstrate that drug-sensitive alleles of TEM-1 can be generated by the LIScan method and suggest that the method may be applied to any protein given an appropriate activity assay. Because of its simplicity and generality, the LIScan method may complement other genetic and chemical genetic methods for analysis of protein function and drug target validation (12).

3.14. Concluding Remarks

Several approaches are currently used in order to mutate a target gene, randomly, serially, or in a rationally designed, site-directed manner. Due to its relative simplicity, the procedure described above may complement other insertional mutagenesis protocols as a standard procedure. Generation of mutated genes using this procedure is more time-consuming than the QuickChange® method, due to the additional phosphorylation and ligation steps, but it can be extended to longer insertions/deletions, and, in our

hands, it is reproducible and reliable. The example presented in subheading 3.13, of inserting a ligand-binding domain into a target protein, demonstrates the usefulness of this procedure as a general method, showing how one can gain control on both the location and the composition of the inserted fragment. This procedure is especially useful for inserting or replacing a desired sequence at a desired location, in contrast to transposon-based methods (14), in which a random sequence is inserted randomly. It also differs from random digestion-based insertions (15) because it enables control over the exact location of the insertion. The ability to perform precise manipulations in any plasmid DNA sequence, and the relative technical simplicity of this procedure, may render it the method of choice for a wide variety of in vitro mutagenesis applications.

4. Notes

1. For the Phusion polymerase, it is possible to use either HF buffer or GC buffer. We used HF buffer successfully, but the optimal conditions for each reaction may depend on the plasmid, the template sequence, and the primers.
2. This procedure was also done with *Pfu* Turbo DNA polymerase (Stratagene). If *Pfu* Turbo is used, the enzymes and buffer volumes should be 1 μl enzyme and 5 μl buffer per 50 μl reaction, respectively.
3. Due to 3′ to 5′ exonuclease activity, the enzyme is added just before starting the reaction. It is recommended to avoid prolonged incubation of the mix with the enzyme prior to starting the reaction.
4. Elongation time should be 15–30 s per 1 kb for Phusion PCR, and 1.5–2 min/kb for *Pfu* Turbo.
5. The above reaction conditions were successfully used in our mutagenesis procedure. The exact conditions for each reaction may depend on the template plasmid and the primers used.
6. If the reaction does not seem to work and no product is generated, the conditions of the reaction should be tuned, as in a standard PCR. In some cases, elevation of the Tm by 2°C can solve the problem. If the Phusion polymerase is used, the 5× GC buffer may be used instead of the 5× HF buffer, which is the standard choice. If the change of parameters does not help, designing new primer(s) may be considered. Usually, elongating one of the two primers is sufficient to get better results.

7. We used the above conditions to generate three sets of insertional mutants of three different genes. Usually, once the conditions for a certain template gene (and plasmid) are established, most of the primer pairs for that gene work well under the same conditions. However, the annealing temperature may be increased or decreased, depending on the primers, even when the same template is used, which means different reactions for different members of the same mutants set.
8. If no colonies grow following transformation (and the control transformation works), repeat the ligation with increased amount of plasmid. Incubate the reaction for 12–18 h at 16°C.
9. Primers considerations for mutagenesis PCR screen: The primers in the PCR mix should include one primer flanking the mutated gene (5′ prime or 3′ prime end), and a second primer complementary to the deletion/replacement/insertion (either in 5′ prime or 3′ prime orientation, depending on the gene-specific primer that is used). This should allow amplification only of a template sequence that contains insertion/replacement, or those which do not contain the deletion. The size of the resulting product should correspond to the distance of the insertion, either from the 5′ prime or from the 3′ prime end of the mutated gene, depending on the primers in use (Fig. 4). When analyzing the results, it is important to note that bands smaller than 250 bp are sometimes difficult to detect in 1% agarose minigels. Therefore, the use of a 5′ prime gene-specific primer with a 3′ prime mutation primer, or the opposite, should be determined according to the distance of the 4C insertion from either the 5′ prime or the 3′ prime end of the mutated gene. An example for such a screen is presented in Fig. 4. Another alternative, if the PCR product is less than 250 bp, is to separate it by using 1.5% or 2% agarose gel.
10. Annealing temperature for the screening PCR: It is important to maintain a high annealing temperature in order to avoid non-specific annealing. We typically use 66°C, and 68°C was also successfully examined.
11. It is recommended to perform a control reaction with the WT gene, to identify non-specific products that may form.
12. The PCR is extremely sensitive and a few bacterial cells can give a very clear signal. It is therefore important to pick a single colony for each PCR reaction, to avoid false positive signals on the one hand and contamination of a positive colony with a negative one on the other.
13. In some amplification reactions, non-specific products may occur (Fig. 3) and the amount of the desired product is

relatively low. The presence of an undigested template plasmid in the transformation may thus result in an increased number of colonies containing the template plasmid, rather than the desired, mutated plasmid. If all colonies in the screen are tested negative, it means that the template plasmid was not properly digested. *DpnI* digestion may then be repeated for a prolonged time (at least 4 h at 37°C). As a control for the entire procedure, 10–40 ng of methylated plasmid may be digested with *DpnI*, followed by a control transformation, together with the mutagenesis transformation. If the digestion was efficient, very few colonies should grow on the control plate.

14. If colonies testing positive in the screening PCR give a WT (unmutated) plasmid, the Tm in the mutagenesis PCR should be elevated, to decrease the chance of non-specific product formation.
15. Even if the restriction analysis is positive, as amplification and cloning errors may occur, the correct insertion/deletion/replacement must be verified by DNA sequencing.

Acknowledgments

This work is dedicated to the memory of Professor Mordechai (Moti) Liscovitch, who passed away during the preparation of this chapter.

In the past few years, Moti became interested in exploring the potential use of small molecules to affect the activity of proteins. He then formulated the idea of the ligand interaction scan (LIScan), a general method that enables the engineering of "regulatable" proteins, using insertion mutagenesis. His goal was to provide the scientific community a novel tool to investigate the function of novel and uncharacterized proteins.

It is our hope that in spite of his untimely death, his ideas and his work will continue to promote research and scientific doing.

We thank all members of our group for help and support. We are grateful to Mina Marmor, Dana Reichmann, Yoav Shaul, Edith Kario, and Dana Chuderland for help with methods. This work was supported by the J&R Center for Scientific Research, the Willner Center for Vascular Biology and La Fondation Raphael et Regina Levy. M.L. was the incumbent of the Harold L. Korda Professorial Chair in Biology and O.E. was a recipient of The Dr. Anna Goldfeder Scholarship.

References

1. Dugan GE, Hewitt EW (2008) Structural and functional dissection of the human cytomegalovirus immune evasion protein US6. J Virol 82(7):3271–3282
2. Ludin B, Doll T, Meili R, Kaech S, Matus A (1996) Application of novel vectors for GFP-tagging of proteins to study microtubule-associated proteins. Gene 173(1 Spec No): 107–111
3. Bartonova V, Kral V, Sieglova I et al (2008) Potent inhibition of drug-resistant HIV protease variants by monoclonal antibodies. Antiviral Res 78(3):275–277
4. Hallet B, Sherratt DJ, Hayes F (1997) Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein. Nucleic Acids Res 25(9):1866–1867
5. Telmer CA, Berget PB, Ballou B, Murphy RF, Jarvik JW (2002) Epitope tagging genomic DNA using a CD-tagging Tn10 minitransposon. Biotechniques 32(2):422–424, 426, 428–430
6. Hamer DH, Thomas CA Jr (1976) Molecular cloning of DNA fragments produced by restriction endonucleases Sa1I and BamI. Proc Natl Acad Sci U S A 73(5):1537–1541
7. Roberts RJ (1976) Restriction endonucleases. CRC Crit Rev Biochem 4(2):123–164
8. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H, Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol 51(Part 1):263–273
9. Wang W, Malcolm BA (1999) Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. Biotechniques 26(4):680–682
10. Zheng L, Baumann U, Reymond JL (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. Nucleic Acids Res 32(14):e115
11. Hemsley A, Arnheim N, Toney MD, Cortopassi G, Galas DJ (1989) A simple method for site-directed mutagenesis using the polymerase chain reaction. Nucleic Acids Res 17(16):6545–6551
12. Erster O, Eisenstein M, Liscovitch M (2007) Ligand interaction scan: a general method for engineering ligand-sensitive protein alleles. Nat Methods 4(5):393–395
13. Adams SR, Campbell RE, Gross LA et al (2002) New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. J Am Chem Soc 124(21):6063–6076
14. Hayes F, Hallet B (2000) Pentapeptide scanning mutagenesis: encouraging old proteins to execute unusual tricks. Trends Microbiol 8(12):571–577
15. Guntas G, Ostermeier M (2004) Creation of an allosteric enzyme by domain insertion. J Mol Biol 336(1):263–273

Chapter 13

Amplification of Orthologous Genes Using Degenerate Primers

Samya Chakravorty and Jim O. Vigoreaux

Abstract

This chapter describes the use of degenerate primers for PCR amplification of orthologous DNA from related species. While several methods for designing degenerate primers have been described, an important consideration is to base the design on a short region of highly conserved amino acids. Here, we present the use of a degenerate primer design strategy called Consensus-degenerate hybrid oligonucleotide primer (CODEHOP). This strategy is very useful when there is low conservation in the multiple alignments of sequences across species. We demonstrate the use of CODEHOP to amplify the entire coding region of the flightin gene from multiple *Drosophila* species.

Key words: Degenerate primers, Evolutionary genetics, CODEHOP, Coding sequence, Flightin, Drosophila, Muscle, Phylogeny

1. Introduction

Research on *Drosophila melanogaster*, spanning nearly 100 years, has provided innumerable insights into the biology of multicellular organisms. Fundamental tenets of classical and molecular genetics emerged from studies of this laboratory-friendly organism. *D. melanogaster* has proven equally valuable for studies in evolutionary biology, ecology, and behavior, to name a few (1). Given this prominence, *D. melanogaster* was among the first multicellular organisms chosen to have its genome fully sequenced (2). The genome sequence information, together with our intimate knowledge of *D. melanogaster* biology and the increasing interest in many other species in the *Drosophila* genus, provided the impetus for the sequencing of 11 additional *Drosophila* genomes (3). These landmark achievements have placed the *Drosophila* genus front and center in comparative evolutionary genomics.

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_13, © Springer Science+Business Media, LLC 2010

Furthermore, information from the 12 sequenced genomes has provided the seed for more expansive phylogenetic studies and for research on the evolution of specific genes (4). After all, DNA sequence information is available for only 12 of the more than 3,500 species in the drosophilid family. The availability of hundreds of drosophilid species from stock centers and the ease with which they can be cultured in the lab provide a great advantage for the use of flies in evolutionary biology studies.

In this chapter, we describe a strategy for isolating an expressed gene of interest using sequence information from orthologous genes of related species. Design of degenerate primers is based on the "Consensus-degenerate hybrid oligonucleotide primer" (CODEHOP) strategy (5). We applied this technique to RT-PCR-generated templates to study the rate of evolution of flightin, an essential flight muscle protein in *Drosophila* that appears to be evolving fast under positive selection (6).

2. Materials

2.1. Fly Stocks

Various *Drosophila* species were obtained from the Drosophila Species Stock Center, San Diego (http://stockcenter.ucsd.edu/info/welcome.php). We selected species whose genomes have been sequenced (*D. melanogaster* and *Drosophila mojavensis*) and species with unknown sequence (*Drosophila repleta*, *Drosophila hydei*, *Drosophila funebris*, *and Drosophila buzatti*).

2.2. Media

1. Different species have different food requirements. The following recipes were used: Corn Meal, Wheeler Clayton and the Banana Opuntia media. We used Corn Meal fly food for *D. melanogaster*, Banana–Opuntia fly food for *D. mojavensis*, *D. buzatti*, and *D. repleta*, and the Wheeler Clayton fly food for *D. funebris* and *D. hydei*.
2. The recipes for the three fly foods used can be found in the following websites:
   - Standard Corn Meal recipe: http://stockcenter.ucsd.edu/info/food_cornmeal.php,
   - Banana–Opuntia recipe: http://stockcenter.ucsd.edu/info/food_banana_Opuntia.php,
   - Wheeler Clayton recipe: http://stockcenter.ucsd.edu/info/food_Wheeler-Clayton.php.

The following standard ingredients were obtained from local supermarket or health food stores: oatmeal, bananas, malt, corn syrup, high protein baby formula, Special K cereal, Product 19 cereal, and wheat germ.

2.3. Other

1. 25 × 95 mm Polypropylene Fly vials (Applied Scientific Drosophila Products, Fisher Scientific).
2. Six oz. Square Bottom Polypropylene Bottles (Genesse Scientific).
3. Eppendorf tubes (Fisherbrand Microcentrifuge Tubes, 1.5–1.7mL).
4. 1.5 Pellet Pestle, Blue (S. Jersey Precision Tool and Mold Inc.).
5. Mineral Oil (Sigma).
6. Agar and yeast were obtained from SciMart.
7. Methylparaben (mold inhibitor) from Fisher.
8. Opuntia cactus powder from the University of Arizona.

2.4. RT-PCR and Agarose Gel Electrophoresis

2.4.1. Kits and Reagents

1. Qiagen RNeasy Protect Minikit (to extract total RNA).
2. Pure Taq-Ready-To-Go PCR beads (GE Healthcare).
3. Buffer RLT (supplied in the Qiagen RNeasy Protect Kit).
4. QIAshredder spin column placed in a 2 mL collection tube (supplied in the Qiagen RNeasy Protect Kit).
5. 70% Ethanol.
6. Buffer RW1 (supplied in the Qiagen RNeasy Protect Kit).
7. Buffer RPE (supplied in the Qiagen RNeasy Protect Kit).
8. RNase-free water (supplied in the Qiagen RNeasy Protect Kit).
9. dNTP mix (10 mM each of dATP, dGTP, dCTP and dTTP at neutral pH).
10. Oligo $dT_{12–18}$ primers (0.5 µg/µL) from Invitrogen.
11. 5× RT Buffer (Reverse Transcriptase Buffer) from Invitrogen.
12. 0.1 M Dithiothreitol (DTT) from Invitrogen.
13. RNaseOUT™ Recombinant RNase Inhibitor from Invitrogen.
14. Superscript III Reverse Transcriptase (Invitrogen).
15. Pure Taq-Ready-To-Go PCR beads (Each PCR bead contains approximately 2.5 units of PuReTaq DNA polymerase, 10 mM Tris–HCl, pH 9.0 at room temperature, 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM of each dNTP, stabilizers, and BSA in a 25 µL reaction volume) from GE Healthcare.
16. Agarose Powder (US Biologicals).
17. 6× Agarose Gel Loading Dye (30% Glycerol, 0.15% Xylene Cyanol, 0.15% Bromophenol)
18. SYBR Safe DNA gel stain (10,000× concentrate in Dimethyl Sulfoxide, Invitrogen)
19. DNA marker – Quickload 100 bp DNA ladder (Invitrogen).

2.4.2. Degenerate Primers

1. The following primers were synthesized by Invitrogen: (see Note 1)

   FlnDgF: 5′-ATGGSRGACGARGARGATCCWTGG-3′ T_m = 70.52°C at 50 mM Na^+

   FlnDgR: 5′-GCACGGAGGCGTACTTCTGRTTNATRTA-3′ T_m = 68.80°C at 50 mM Na^+

2.4.3. Nondegenerate Primers

1. The following primers were synthesized by Invitrogen:

   D.M.FlnCdsF: 5′-ATGGCAGACGAAGAAGATCC-3′ T_m = 59.25°C at 50 mM Na^+

   D.M.FlnCdsR: 5′-CTAAAGGACACTGGCATACC-3′ T_m = 54.83°C at 50 mM Na^+

2.4.4. Preparing the Agarose Gel

To make the gel, mix 1.5 gm of agarose powder in 100 mL of TAE (40 mM Tris, 20 mM Acetate, 1 mM EDTA, pH 8.4) buffer in an Erlenmeyer flask and use a sharpie to mark the top level of the liquid. Dissolve by gentle heating and swirling in a microwave oven and add deionized water back up to the mark when the agarose is completely dissolved. Add 2 µL of SYBR Safe DNA gel stain for staining.

3. Methods

3.1. Maintaining Fly Stocks

Flies are maintained in vials containing the indicated food (see Subheading 2.2) and transferred to fresh food vials every 3 weeks. All the species are kept in a constant 22°C room in 65% humidity.

3.2. Isolation of Total RNA

1. Isolation of total RNA is done using the Qiagen RNeasy Protect Kit following the kit protocol unless as indicated below. (For detailed description and reference see RNeasy Mini Handbook in http://www1.qiagen.com/literature/handbooks/literature.aspx?id=1000297).
   (a) Place flies in 1.5 mL Eppendorf tube and grind them with pestle in Buffer RLT (see Note 2).
   (b) Pipette the lysate directly into a QIAshredder spin column placed in a 2 mL collection tube, and centrifuge for 2 min at full speed.
   (c) Add 1 volume of 70% ethanol to the homogenized lysate and mix well by pipetting. Do not centrifuge.
   (d) Transfer up to 700 µL of the sample, including any precipitate that may have formed, to an RNeasy spin column placed in a 2 mL collection tube. Close the lid

gently and centrifuge for 15 s at 12,000 rpm (8373 g). Discard the flow-through.

(e) Add 700 µL of Buffer RW1 to the RNeasy spin column. Incubate for 5 min at room temperature to eliminate DNase. Close the lid gently and centrifuge for 15 s at 12,000 rpm to wash the spin column membrane. Discard the flow-through.

(f) Add 500 µL Buffer RPE to the RNeasy spin column. Close the lid gently and centrifuge for 15 s at 12,000 rpm to wash the spin column membrane. Discard the flow-through.

(g) Add 500 µL Buffer RPE again to the RNeasy spin column. Close the lid gently and centrifuge for 2 min at 12,000 rpm to wash the spin column membrane.

(h) Place the RNeasy spin column in a 2 mL collection tube and discard the old collection tube containing the flow-through. Close the lid gently, and centrifuge at full speed 14,000 rpm (11400 g) for 1 min.

(i) Place the RNeasy column in a new 1.5 mL collection tube. Add 35 µL RNase-free water directly to the spin column membrane. Close the lid gently and centrifuge for 1 min at 12,000 rpm to elute the RNA.

(j) To improve the yield, repeat the previous step (**i**) by adding the 35 µL of eluate to the column again and repeating the centrifugation step.

2. Determine the RNA concentration and purity (A260/280) using a Nanodrop Spectrophotometer (Nanodrop Technologies, Inc., Wilmington, DE).

3.3. RT-PCR Reaction

For each RNA extraction combine:

3.3.1. Reverse Transcription (RT) Reaction

1. 1 µL Oligo $dT_{12–18}$ primers (0.5 µg/µL).
2. 1 µL 10 mM dNTP mix.
3. 100 ng of total RNA (see Note 3).
4. Add RNase-free water to 13 µL.

Incubate at 65°C for 5 min and then transfer to ice for at least 1 min. For each reaction above, add the following components:

1. 4 µL 5× RT (for First-Strand synthesis) buffer.
2. 1 µL 0.1 M DTT.
3. 1 µL RNaseOUT™ Recombinant RNase Inhibitor.
4. Superscript III Reverse Transcriptase.

Incubate the reaction at 50°C for 50–60 min. Inactivate the reaction by heating to 70°C for 15 min, then transfer to ice for

1 min followed by a brief centrifugation to collect all the liquid. Set up the Polymerase Chain Reaction (PCR) reaction.

3.3.2. Polymerase Chain Reaction

PCR is done using Pure Taq-Ready-To-Go PCR beads. A quantity of 200 ng of the template was used. Final concentration of each primer in the reaction was 0.2 µM. Program the thermal cycler to the following protocol:

1. 96°C for 3 min
2. Thirty-five cycles of
   - 96°C for 30 s
   - 53°C for 1 min
   - 72°C for 1 min
3. 72°C for 10 min

3.4. Primer Design

Several methods have been described for designing degenerate primers (7–10). The design of degenerate primers is facilitated by the availability of sequenced genomes from closely related species (see Note 4). To illustrate this, we present the following example using the flightin gene from *D. melanogaster*.

1. We started with the known coding sequence (CDS) of the *D. melanogaster* flightin gene (11), also obtained from Flybase (http://www.flybase.org). Then, we retrieved the CDS of flightin from the other 11 sequenced *Drosophila* species using the "Orthologs" ∀ *D. melanogaster* flightin CDS (http://flybase.org/reports/FBgn0005633.html) (Table 1).
2. We used ClustalW to align the 12 flightin CDS (see Note 5). The N-terminal end of flightin is well conserved, and thus we could design N-terminal degenerate primers manually (Fig. 1).
3. The C-terminal end is less conserved than the N-terminal end and degenerate primers could not be designed from a visual alignment. For designing the degenerate C-terminal primers, we used the CODEHOP strategy (see Note 6 and ref. (5)). The program is available at http://blocks.fhcrc.org/codehip.html.
4. The C-terminal ends of the 12 *Drosophila* flightins were aligned with ClustalW (Fig. 2). Instead of using the multiple alignment as input, the alignment is broken into Blocks with the help of the Block Maker software present in the CODEHOP website (see Note 7). Blocks are pasted automatically onto the "Paste your block(s) below" field in the CODEHOP program site by clicking on "CODEHOP" button from the BlockMaker site after the BlockMaker run. The Blocks are then used to design the degenerate primers. For the flightin multiple sequence alignment, there were four Blocks. The CODEHOP program designs possible

Table 1
Sequences used for multiple sequence alignments

Species name	Sequence used (Flybase ID)
D. melanogaster	Flightin full-length CDS (FBgn0005633)
D. mojavensis	Orthologous sequence to flightin CDS in *D. melanogaster*, (FBgn0136135)
D. pseudoobscura	Orthologous sequence to flightin CDS in *D. melanogaster* (FBgn0244340)
D. willistoni	Orthologous sequence to flightin CDS in *D. melanogaster* (FBgn0220979)
D. erecta	Orthologous sequence to flightin CDS in *D. melanogaster* (FBgn0105625)
D. annanaseae	Orthologous sequence to flightin CDS in *D. melanogaster* (FBgn0087873)
D. simulans	Orthologous sequence to flightin CDS in *D. melanogaster* (FBgn0183970)
D. sechellia	Orthologous sequence to flightin CDS in *D. melanogaster* (FBgn0169780)
D. persimilis	Orthologous sequence to flightin CDS in *D. melanogaster* (FBgn0162637)
D. virilis	Orthologous sequence to flightin CDS in *D. melanogaster* (FBgn0198760)
D. grimshawi	Orthologous sequence to flightin CDS in *D. melanogaster* (FBgn0122202)
D. yakuba	Flightin full-length CDS, (FBgn0067972)

```
D.melano        ATGGCAGACGAAGAAGATCCATGGGGTTTCGACGACGGCGGC------GAG---GAG---
D.ana           ATGGCAGACGAAGAAGATCCATGGGGTTTCGACGATGGCGGC------GAG---GAG---
D.grim          ATGGGGGACGAAGAAGATCCTTGGGGTTTTGATGATGAAGGT------GAATCTGAT---
D.moja          ATGGGGGACGAAGAAGATCCTTGGGGTTTCGATGATGGCGGT------GATGCTGAG---
D.per           ATGGCAGACGAAGAAGATCCATGGGGAGACGACGCCGGTGGT------GATACTGAG---
D.pseudo        ------------------CCATGGGGAGACGACGCCGGTGGT------GATACTGAG---
D.sechellia     ATGGCAGACGAAGAAGATCCATGGGGTTTCGACGACGGCGGC------GAG---GAG---
D.simulans      ATGGCAGACGAAGAAGATCCATGGGGTTTCGACGACGGCGGC------GAG---GAG---
D.virilis       ATGGCGGACGAAGAAGATCCTTGGGGTTTTGATGAAGGTGATACCGTTGAGTCCGAT---
D.willistoni    ATGGGAGACGAGGAGGATCCTTGGGGTTTTGATGATGGCGGC------GATGCTGAGCCA
D.yakuba        ATGGCAGACGAAGAGGATCCATGGGGTTTCGACGACGGCGGC------GAG---GAG---
D.erecta        ATGGCAGACGAAGAAGATCCATGGGGTTTCGACGACGGCGGC------GAG---GAG---
                                ** *****    ** *  *  *        **    **
```

Fig. 1. Multiple DNA sequence alignment of the 5′ CDS of the flightin gene from the 12 *Drosophila* species with sequenced genomes. Alignments were generated using ClustalW (1.83). The *asterisks* indicate fully conserved positions in the alignment.

```
D.anna     KKQTGVSREIPRPQTWAERVLRTRNISGSGIDSYAPSAKRDKQLIQTLAASIRTYNYHTKAYINQRYASVL
D.ere      KKQTGVAREIPRPQTWAERVLRTRNISVGDIDSYAP-AKRDKQLIQTLAASIRTYNYHTKAYINQRYASVL
D.grim     KKQVGVSREIPRPQTWAERVLRTRDINGNGIDNYAQSTKRDKHLIQTLAASIRTYNYHTKAYINQKYAGVI
D.melano   KKQTGVAREIPRPQTWAERVLRTRNISGSDIDSYAP-AKRDKQLIQTLAASIRTYNYHTKAYINQRYASVL
D.moja     KKQVGVARDIPRPQTWAERVLRTRDINAGGIDNYSQSTKRDKHLIQTLAASIRTYNYHTKAYINQKYASVL
D.per      KKQVGVARDIPRPQTWAERVLRTRNVSGSGIDSFEPSAKRDKQLTQTLAASIRTYNYHTKAYMNQKYGSVL
D.pseudo   KKQVGVARDIPRPQTWAERVLRTRNVSGSGIDSFEPSAKRDKQLTQTLAASIRTYNYHTKAYMNQKYGSVL
D.sec      KKQTGVAREIPRPQTWAERVLRTRNISGSDIDSYAP-AKRDKQLIQTLAASIRTYNYHTKAYINQRYASVL
D.sim      KKQTGVAREIPRPQTWAERVLRTRNISGSDIDSYAP-AKRDKQLIQTLAASIRTYNYHTKAYINQRYASVL
D.vir      KKQVGVTRDIPRPQTWAERVLRTRDINASGIDHINLSTKRDKQLVQTLAASIRTYNYHTKAYINQKYANVL
D.wili     KKQVGEARDIPRPQTWAERVLRTRNISGSGIDSFAPSTKRDKQLIQTLAASIRTYNYHTKAYINQKYASVL
D.yak      KKQTGVAREIPRPQTWAERVLRTRNISGSDIDSYAP-AKRDKQLIQTLAASIRTYNYHTKAYINQRYASVL
           ***.* :*:***************:: ..**     :****:* *****************:**:*..*:
```

Fig. 2. Multiple amino acid sequence alignment of the flightin C-terminus from the 12 *Drosophila* species with sequenced genomes. Alignments were generated using ClustalW (1.83). The *asterisks* indicate fully conserved positions in the alignment.

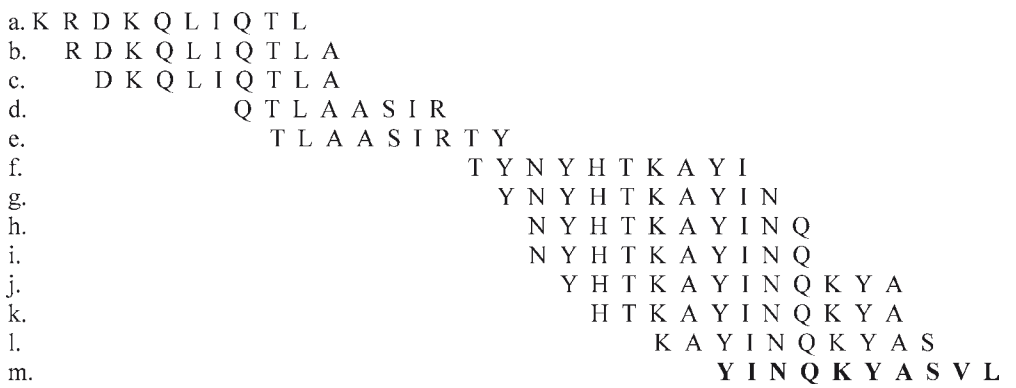


Fig. 3. The last block (Block flnaaseqoD) obtained from BlockMaker for flightin sequences. This last block was chosen in order to design reverse primers from the extreme C-terminal end of the protein. In this figure, the program has divided the last block into segments from which possible CODEHOPs can be made. The blocks made by BlockMaker indicate the most common amino acid at each position in the multiple alignments. Note that among the last five amino acids, two (Y and V at positions –5 and –2 from the C-terminal end, respectively) are strictly conserved (see Fig. 2).

degenerate primers at different positions or segments within each block (see Note 8).

5. The parameters in the CODEHOP program site were set as default, namely:
   Degeneracy – 128, strictness – 0, temperature – 60°C, nucleotide runs in the clamp region – 5 in a row, $(K^+) = 50$ mM.
6. We selected the "Standard" genetic code and the *D. melanogaster* codon usage table.
7. In order to amplify the full CDS, we restricted our interest to segment M of the last block (the C-terminal end block; Fig. 3). The reverse degenerate primer shown below corresponds to the complement of this segment.

   Primer oligo: 5′-GCACGGAGGCGTACTTCtgrttnatrta-3′
   degeneracy = 16, temp = 62.0°C.

3.5. Agarose Gel Electrophoresis

1. Analysis of the reaction products was done by mixing 10 µL of the RT-PCR products with 1.5 µL of 6× Agarose Gel Loading Dye and loading the sample in a 1.5% agarose gel.
2. Electrophoresis was carried out at 100 V until the tracking dye reached the middle of the lower half of the gel. Gel was visualized under UV light using a High Performance UV Transilluminator (Ultra Violet Products).
3. Figure 4a shows that amplification with degenerate primers gave rise to a single band of ~550 bp for all species tested. Based on the known sequence, the expected sizes for *D. melanogaster* (lane 1) and *D. mojavensis* (lane 2) are 549 bp and 552 bp, respectively. Note that a similar size band is obtained for all the other species for which the sequence is not known.

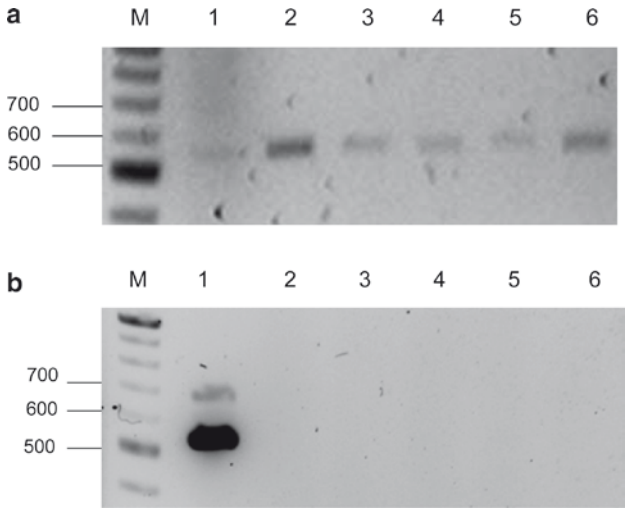


Fig. 4. (**a**) Agarose gel showing PCR products generated with degenerate primers. Each lane was loaded with 10 µL of the PCR reaction. Lane M is the marker lane. Lanes 1–6 are the PCR products of *Drosophila melanogaster*, *Drosophila mojavensis*, *Drosophila repleta*, *Drosophila hydei*, *Drosophila funebris*, and *Drosophila buzatti*, respectively. Only bands of the expected size (~550 bp) are detected. (**b**) Agarose gel showing PCR products generated with nondegenerate primers. The reactions were carried out with primers specific to the *D. melanogaster* flightin CDS. Each lane was loaded with 10 µL of the PCR reaction. Lane M is the marker lane. Lanes 1–6 are the PCR products of *D. melanogaster*, *D. mojavensis*, *D. repleta*, *D. hydei*, *D. funebris*, and *D. buzatti*, respectively. Only the CDS of the gene from the targeted species (lane 1) was amplified. The band at ~549 bp corresponds to the full-length CDS; a second, fainter band at ~700 bp corresponds to a spliced intermediate (see Note 9).

4. In contrast to the exact PCR products obtained with degenerate primers, the use of species-specific nondegenerate primers did not amplify orthologous genes (Fig. 4b). Using primers designed for the *D. melanogaster* CDS resulted in amplification of the *D. melanogaster* gene only (see Note 9).

4. Notes

1. The IUBPAC codes for nucleotide degeneracies are used:

   N – A, C, G, T

   S – G, C

   R – A, G

   W – A, T

   Most melting temperature (Tm) calculations are done based on Na^+ as the primary cation. Increasing the salt concentration increases Tm as higher ionic concentration will tend to stabilize duplex DNA (i.e., cations have stronger affinity

for double-stranded DNA than for single-stranded DNA). Invitrogen supplies Tm for degenerate primers at 50 mM Na^+ concentration that should be taken into account for estimating the annealing temperature for PCR. This Tm is based on salt concentration and primer sequence (G-C rich sequences have higher Tm than A-T rich sequences). However, primer concentration must also be considered in selecting an annealing temperature as high primer concentration increases Tm by virtue of its effect on duplex formation. We calculated the Tm using the IDT OligoAnalyzer 3.1 in the following website keeping the parameters set as default: http://www.idtdna.com/analyzer/applications/oligoanalyzer/.

2. The number of flies to grind for RNA extraction depends on the size of the species of Drosophila used. More than 30 mg of flies or tissues should not be used as this can clog the RNeasy column.
3. The amount of total RNA can be as low as 10 pg or as high as 5 µg, depending on the concentration.
4. The most important thing in designing degenerate primers is to start with a highly conserved area of the protein of at least 3–4 amino acids.
5. ClustalW parameters: Gap open penalty- 10, Gap extension penalty- 0.5, Weight matrix- BLOSUM.
6. The primers designed by the CODEHOP strategy consist of a short 3′ degenerate core region and a longer 5′ consensus clamp region (5). According to the CODEHOP program, only 3–4 highly conserved amino acid residues are required for designing the degenerate core which gets stabilized by the consensus clamp region during the annealing reaction to the template. Afterwards, during the later rounds of amplification, the consensus clamp gives stabilization to the annealing reaction.
7. Blocks are multiply aligned ungapped segments corresponding to the most highly conserved regions of proteins.
8. For detailed description of the steps in the primer design process and how the CODEHOP program works, see ref. (5) and the following website: http://bioinformatics.weizmann.ac.il/blocks/help/CODEHOP/CODEHOP_program.html.
9. In Fig. 4b, we used nondegenerate primers specific for *D. melanogaster* flightin. The primers amplified the CDS of the *D. melanogaster* gene but not of the other five species. In *D. melanogaster*, the flightin gene has three introns of 514 bp, 66 bp and 62 bp respectively. In Fig. 4b, the dark band in the *D. melanogaster* lane (lane 1) corresponds to the open reading frame (~550 bp) while the faint, slower

migrating band corresponds to an incompletely spliced product that includes the second and the third introns (66 bp and 62 bp, respectively). Thus, species-specific nondegenerate primers were not successful in amplifying genes from other species. In Fig. 4a, we used degenerate primers using the CODEHOP strategy and were successful in amplifying the gene CDS in all the six species. The intensities of the bands vary due to varying concentrations of starting RNA.

References

1. Rubin GM, Lewis EB (2000) A brief history of Drosophila's contributions to genome research. Science 287(5461):2216–2218
2. Adams MD, Celniker SE, Holt RA et al (2000) The genome sequence of Drosophila melanogaster. Science 287(5461):2185–2195
3. Birney E (2007) Evolutionary genomics: come fly with us. Nature 450(7167):184–185
4. Clark AG, Eisen MB, Smith DR et al (2007) Evolution of genes and genomes on the Drosophila phylogeny. Nature 450(7167):203–218
5. Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res 26(7):1628–1635
6. Reedy MC, Bullard B, Vigoreaux JO (2000) Flightin is essential for thick filament assembly and sarcomere stability in Drosophila flight muscles. J Cell Biol 151:1483–1499
7. Linhart C, Shamir R (2007) Degenerate primer design: theoretical analysis and the HYDEN program. Methods Mol Biol 402: 221–244
8. Souvenir R, Buhler J, Stormo G, Zhang W (2007) An iterative method for selecting degenerate multiplex PCR primers. Methods Mol Biol 402:245–268
9. Najafabadi HS, Saberi A, Torabi N, Chamankhah M (2008) MAD-DPD: designing highly degenerate primers with maximum amplification specificity. Biotechniques 44(4):519–520, 522, 524–526
10. Preston GM (1997) Cloning gene family members using the polymerase chain reaction with degenerate oligonucleotide primers. Methods Mol Biol 69:97–113
11. Vigoreaux JO, Saide JD, Valgeirsdottir K, Pardue ML (1993) Flightin, a novel myofibrillar protein of Drosophila stretch-activated muscles. J Cell Biol 121(3):587–598

Part III

Reviews

Chapter 14

Computational Evaluation of Protein Stability Change upon Mutations

Shuangye Yin, Feng Ding, and Nikolay V. Dokholyan

Abstract

When designing a mutagenesis experiment, it is often crucial to estimate the stability change of proteins induced by mutations ($\Delta\Delta G$). Despite the recent advances in computational methods, it is still challenging to estimate $\Delta\Delta G$ quickly and accurately. We recently developed the Eris protocols for *in silico* evaluation of the $\Delta\Delta G$. Starting from the tertiary structure of the wide-type protein, the Eris protocols can model the structure of the mutant protein and estimate $\Delta\Delta G$ using the structure models. The Eris protocols not only efficiently optimize the side chains conformations, taking advantage of a fast rotamer-based searching algorithm, but also allow protein backbone flexibility during the modeling. As a result, the Eris protocols effectively resolve steric clashes induced by certain mutations and have more accurate $\Delta\Delta G$ predictions than a fixed-backbone approach. We discuss the general aspects of computational $\Delta\Delta G$ estimations and discuss in detail the principles and methodologies of the Eris protocols.

Key words: Protein stability, Comparative modeling, Protein design, Eris

1. Introduction

The thermodynamic stability of a protein (ΔG) is measured by the free energy difference between the folded state and the unfolded state ($\Delta G = G_{\text{fold}} - G_{\text{unfold}}$). The value of ΔG determines the fraction of a protein in the folded state, thereby having a profound effect on protein activities. Experiments have shown that natural proteins are only marginally stable (1), and the measured ΔG values of most proteins fall in the range of −3 to −15 kcal/mol (2). The large energetic contributions that favor folding, including hydrophobic packing, hydrogen bonding, and electrostatic interactions, are subtly balanced by the entropic penalization of folding. As a result, a single mutation may shift this balance and significantly affects the stability of the whole protein. To accurately estimate

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_14,

the protein stability changes induced by mutation (measured as $\Delta\Delta G = \Delta G^{Mut} - \Delta G^{WT}$) remains a significant challenge in protein engineering.

1.1. Experimental Measurements of $\Delta\Delta G$

Experimentally, $\Delta\Delta G$ can be determined from ΔG measurements for both wild-type and mutant proteins using denaturing experiments (3–6). In a denaturing experiment, the protein unfolds by adding denaturing agents such as urea or guanidinium HCl (GdHCl) or by increasing temperature. The fraction of the denatured protein can be estimated by monitoring certain measurable parameters such as circular dichroism (CD), which feature distinct properties in the folded and unfolded states. Therefore, the free energy difference can be determined as $\Delta G = -RT\ln((y_f - y)/(y - y_u))$, where R is the ideal gas constant, T is the absolute experiment temperature, y is the measured property (such as CD), y_f and y_u would be the intrinsic CD properties of the protein in the folded and unfolded states, for example.

In a chemical denaturing experiment, it is experimentally observed (4) that ΔG has a linear dependence on denaturant concentration $[D]$ as $\Delta G = \Delta G[H_2O] - m[D]$, where $\Delta G[H_2O]$ is the solution stability without denaturants, and m is a constant that measures the dependence of ΔG on denaturant concentration. The parameter m has been found to be proportional to the solvent accessible area. Experimentally, ΔG is measured for a number of denaturant concentrations, and the $\Delta G[H_2O]$ value is obtained from linear extrapolation. The $\Delta\Delta G$ values thus determined are denoted as $\Delta\Delta G[H_2O] = \Delta G^{Mut}[H_2O] - \Delta G^{WT}[H_2O]$. It has also been pointed out that the error during extrapolation to obtain $\Delta G[H_2O]$ is large (7), so that $\Delta\Delta G[H_2O]$ can have a large uncertainty on the order of 0.5 kcal/mol. In practice, the solution ΔG values are often obtained at a specific denaturant concentration $[D]^{50\%}$, where 50% of the protein is denatured. Thus, the solution ΔG can be estimated as $m[D]^{50\%}$ and the $\Delta\Delta G^{50\%} = m([D]^{50\%Mut} - [D]^{50\%WT})$ are experimentally reproducible within 0.1 kcal/mol.

In a thermal denaturing experiment, ΔG is measured as a function of temperature near the unfolding transition, from which the melting temperature (T_m) can be determined. The ΔG at the physiological temperature is calculated using the Gibbs–Helmholtz equation as (3), $\Delta G(T) = \Delta H(T_m)(1 - T/T_m) - \Delta C_p(T_m - T + T\ln(T/T_m))$, where $\Delta H(T_m)$ is unfolding enthalpy at T_m, and ΔC_p is the heat capacity change during unfolding. Both $\Delta H(T_m)$ and ΔC_p values can be obtained from calorimetric measurements (6).

1.2. Computational Method for $\Delta\Delta G$ Evaluation

Theoretically, given the physical model, free energy can be obtained from statistical mechanics using the partition function, Z, as $G = -RT\ln Z$. However, analytical calculations of partition functions require integration over all degrees of freedom in the protein's conformational space, which is practically impossible

except for the simplest cases. Advances in computational biology have made direct calculation of $\Delta\Delta G$ possible using molecular dynamics (MD) or Monte Carlo (MC) simulations (8–11), a comprehensive review of which can be found in ref. (12). Using MD simulations, $\Delta\Delta G$ value has been calculated for T157V mutant of T4 lysozyme, and the estimation is in close agreement with experimental measurements (9). Alternatively, simplified models have been adopted to estimate the $\Delta\Delta G$ without intensive computation (13–23). Some approaches use statistical or empirical functions to describe interactions within proteins (13, 15, 17–23). Other approaches use machine learning techniques, and the predictor programs are trained using a large set of experimental $\Delta\Delta G$ data (14, 16). Reasonable accuracy has been achieved in both cases. However, regression-based methods have to rely on known $\Delta\Delta G$ data to calibrate the parameters, and are thereby susceptible to any bias in training. Statistical analysis of known $\Delta\Delta G$ data reveals that most of the mutations are from larger residues to smaller residues, as obtained from alanine scan or glycine scan experiments (24). Therefore, an unbiased prediction method that can be applied on mutations from small residues to larger residues at reasonable accuracy is desirable.

1.3. Eris Method $\Delta\Delta G$ Evaluation

Eris is a unique computational approach that combines physical and statistical force fields with a fast side-chain packing algorithm. The protocol is not limited to large-to-small mutations since all the parameters are obtained independent of experimental $\Delta\Delta G$ data. We have also validated the protocol using a large dataset of 595 mutations from five protein families (24). The mutation sites are located on all types of secondary structures and have various solvent accessibilities. Unlike other methods which often assume that the protein backbones are fixed upon mutation, Eris can explore backbone conformations near the original structure to find the lowest-energy structure. The flexible backbone protocol has been shown to be successful in removing atomic clashes introduced by some small-to-large mutations (25). All the functionalities of the Eris protocols are freely accessible from a web interface at http://eris.dokhlab.org/.

2. Materials

2.1. Protein Stability Data Collection

In order to validate Eris protocols, we tested all protocols in a large set of experimental $\Delta\Delta G$ data. Some of the data are from validation sets of previously published $\Delta\Delta G$ prediction methods (18, 20). Other data are from the ProTherm database (26), a free online source of manually collected $\Delta\Delta G$ data from published literature. All $\Delta\Delta G$ experimental measurements and annotations,

including the protein name, mutation type, experimental conditions, and reference literature are searchable from the web interface of ProTherm. To ensure the reliability of the experimental data, we only select proteins that have a majority of experimental $\Delta\Delta G$ data at physiological pH ($\Delta\Delta G(H_2O)$). In addition, for each protein, we select the measurements that mostly appear in the same publications to ensure the best consistency within the data. At the end, we obtain a collection of experimental $\Delta\Delta G$ data for 595 mutations from five different protein families, including FK506 binding protein, staphylococcal nuclease, spectrin, myoglobin, and serine protease inhibitor CI-2.

2.2. Obtaining Protein 3D Structure

The Eris protocols use protein tertiary structure to calculate protein stabilities. The most common source for the structure is the Protein DataBank (27) (PDB), which contains tens of thousands of user-deposited protein structures determined experimentally by NMR spectroscopy or X-ray crystallography techniques. X-ray structures usually have higher resolution, meaning that the positioning of atoms is more accurate than that determined by NMR. Such atomic accuracy is very important for structural computational protocols such as Eris, which relies on the physical interactions between atoms and is sensitive to incorrect positioning of the atoms.

2.3. Preprocessing of Protein 3D Structures

2.3.1. Missing Heavy Atoms

Some PDB files may be inappropriate for structure-based energy evaluations directly due to various issues. One of the common problems includes missing (unresolved) heavy atoms in one or more residues. If the missing atoms are in the amino acid side chains of the protein, Eris will regenerate the side chains and optimize the conformation according to the residue topology, backbone conformation, and neighboring atomic interactions. If the missing atoms are in the main chain or the whole residue is missing, the computer program will not be able to determine the coordinates of these atoms. Evaluation of proteins with such absent constituents can be potentially problematic if those missing atoms are near the mutation sites of interest, since some important interactions will be overlooked.

2.3.2. Alternative Conformations

Sometimes a PDB file contains several possible coordinates for certain atoms because the position of these atoms cannot be determined unambiguously. Such ambiguity of the position of atoms may arise from the limited resolution of the experimental techniques, or reflect the intrinsic flexibility of these atoms. The alternative conformations are often related to the ambiguity of determining the conformation of certain amino acid side chains. In Eris protocols, we only use the first conformation listed in the PDB file, as it is usually energetically more favorable. Considering only one conformation is sufficient even if the side chain in the

actual structure is in an alternative conformation since all side chain conformations will be explored in Eris protocols.

2.3.3. Separate the Domains of Interest

In some other cases, the PDB structure contains a protein complex with multiple domains. It may be advantageous to isolate the domain where mutations are located, given that the mutated residues are not interacting with residues in other domains. Protein structure visualization programs such as PyMOL (28) are often helpful for locating the mutation sites and editing the structures. Removing remote domains reduces the size of the conformational space and improves the sampling efficiency of the program. It is also potentially beneficial to remove part of the structure that either has low resolution or has missing/incorrect structure, if the removed part is not in the vicinity of the mutation sites.

2.3.4. Relaxation of the Input Structures

The accuracy of the atomic coordinates in protein structures is limited by experimental techniques. This limited resolution can potentially affect the energetic evaluations. In general, NMR-determined structures have lower resolution than X-ray determined structures. Therefore, Eris provides a tool to refine the input structure before $\Delta\Delta G$ estimations. This preprocessing is especially helpful if the only available structures are from NMR spectroscopy or from computational modeling. The prerelaxation step optimizes all the side chains and backbone dihedral angles of the input wild-type protein. Other molecular modeling tools, such as Amber (29) and CHARMM (30), may be used to optimize bond length and angles, which are sometimes incorrectly assigned by homology modeling tools.

2.3.5. Protonation

Protonation of the PDB is not required before using Eris. Position of hydrogen atoms usually cannot be resolved by X-ray. Therefore, Eris automatically generates and determines the position of polar hydrogen atoms for each residue. In Eris protocols, all nonpolar hydrogen atoms are ignored, and their effects are included implicitly in the corresponding bonded heavy atoms (united-atom model). All polar hydrogen atoms are considered explicitly in order to calculate the hydrogen-bonding energy. The polar hydrogen atoms are generated based on the topology of the amino acids, and the positions are optimized to minimize the total energy of the protein.

3. Methods

3.1. The Medusa Force Field

We use the Medusa force field (31) to describe the atomic interactions within a protein. The nonbonded interactions are described by a linear combination of van der Waals, solvation and

hydrogen-bonding energies. The internal interactions of each residue are described by statistical backbone-dependent rotamer energies. In addition, the free energy of the unfolded state is described by a linear sum of reference energies over all amino acids (see Note 1). Specifically, the folding free energy of a protein is described using:

$$\begin{aligned}\Delta G = {} & W_{\text{vdw\_attr}} E_{\text{vdw\_attr}} + W_{\text{vdw\_rep}} E_{\text{vdw\_rep}} + W_{\text{solv}} E_{\text{solv}} \\ & + W_{\text{bb\_hbond}} E_{\text{bb\_hbond}} + W_{\text{sc\_hbond}} E_{\text{sc\_hbond}} + W_{\text{bb\_sc\_hbond}} E_{\text{bb\_sc\_hbond}} \\ & + W_{\text{aa}|\varphi,\psi} E_{\text{aa}|\varphi,\psi} + W_{\text{rot}|\varphi,\psi,\text{aa}} E_{\text{rot}|\varphi,\psi,\text{aa}} - E_{\text{ref}}\end{aligned} \quad (1)$$

Here, $E_{\text{vdw\_attr}}$ and $E_{\text{vdw\_rep}}$ are the attractive and repulsive parts of the van der Waals energy, respectively; E_{solv} is the solvation free energy; $E_{\text{bb\_hbond}}$, $E_{\text{sc\_hbond}}$, and $E_{\text{bb\_sc\_hbond}}$ are the hydrogen-bonding energies between backbone atoms, side chain atoms, and between backbone and side chain atoms; $E_{\text{aa}|\phi,\psi}$ is the statistical energy of the amino acid for any given backbone dihedral angles ϕ and ψ; $E_{\text{rot}|\phi,\psi,\text{aa}}$ is the statistical energy for the side chain rotamer for any given amino acid and backbone dihedral angles; E_{ref} is the sum of reference energies for all amino acids; $W_{\text{vdw\_attr}}$, $W_{\text{vdw\_rep}}$, W_{solv}, $W_{\text{bb\_hbond}}$, $W_{\text{sc\_hbond}}$, $W_{\text{bb\_sc\_hbond}}$, $W_{\text{aa}|\phi,\psi}$, and $W_{\text{rot}|\phi,\psi,\text{aa}}$ are factors to weight the different energy terms.

The parameters of the van der Waals interaction are adapted from CHARMM19 force field (30). The solvation energy is modeled using the EEF1 implicit solvent model by Lazaridis and Karplus (32). The hydrogen-bonding interaction is calculated using a statistical model developed by Kortemme and Baker (33). The rotamer energies are derived from the Dunbrack backbone-dependent rotamer library (34). Since various energy contributions are taken from difference sources, weighting factors are used to rescale and balance those terms of energies. Those weighting factors as well as the 20 reference energies for the 20 amino acid types have been trained using 34 high-resolution protein structures to recapitulate the native amino acid sequences (31).

3.2. Algorithms

3.2.1. $\Delta\Delta G$ Calculation

The $\Delta\Delta G$ is calculated as the energy difference between the wild-type protein and mutant (Fig. 1). The protocol requires an input structure of the wild-type protein, from which the mutant protein structure will be derived. In both wild-type and mutant proteins, all the side chains are repacked to minimize the total Medusa energy of the protein. A Monte Carlo (MC) simulated annealing algorithm is used for finding the optimal conformation of side chains. To reduce the statistical uncertainty resulting from the MC simulation (see Note 4), the side chain packing is performed 20 times to generate 20 output structures, from which the average ΔG will be evaluated. If the MC simulation fails to generate reasonable structures, errors will be reported (see Note 2).

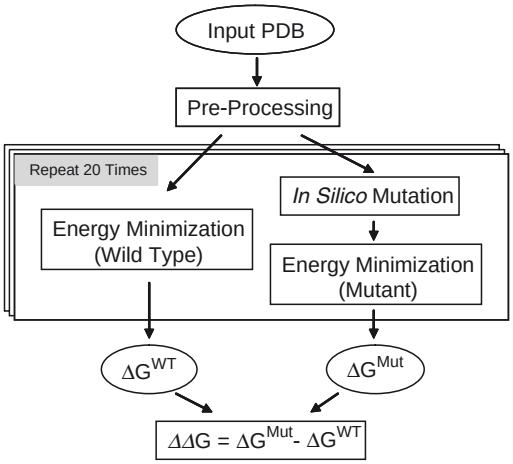


Fig. 1. Schematic flowchart of the Eris $\Delta\Delta G$ estimation protocols. The input wild-type protein structure is preprocessed before applying *in silico* mutation and energy minimization. To reduce the statistical uncertainty, the energy minimization is repeated 20 times, and the $\Delta\Delta G$ evaluation is averaged over 20 runs.

3.2.2. Backbone Prerelaxation Protocol

Accurate estimation of the protein stability relies on the details of the atomic contacts at high resolution. In some cases, however, only low-resolution structures are available, either from NMR spectroscopy or from computational protein modeling. At low resolution, the side chains of the protein core are sometimes not well packed, which reduces $\Delta\Delta G$ estimation accuracy. We use a backbone prerelaxation protocol to improve the side chain packing quality of the low-resolution protein structures. The prerelaxation is performed by applying the flexible-backbone algorithm on the input structure for 20 repetitions. The output structure with the lowest energy is selected as the prerelaxed structure for $\Delta\Delta G$ evaluation. We have demonstrated that using this prerelaxation technique can significantly improve the $\Delta\Delta G$ prediction accuracy for a chicken spectrin protein (25).

3.2.3. Generation of Allowed Rotamer List

Starting from the input structure, the program generates a list of available rotamers for each residue site. For the given backbone dihedral angles, only rotamers that have a probability of greater than 0.01 are added in the rotamer lists. The native rotamer state is also added to the list if no other rotamers are available for this residue. For the mutation sites, only the rotamers of the target amino acid are added into the list, and the target amino acid is set to a randomly assigned state before MC simulations.

3.2.4. MC-Simulated Annealing

After the lists of available rotamers have been formed for each site, the Eris program uses an MC-simulated annealing algorithm (Fig. 2) to find the best packing of the side chains by minimizing the total Medusa energy of the protein (Eq. 1). In the Eris protocol,

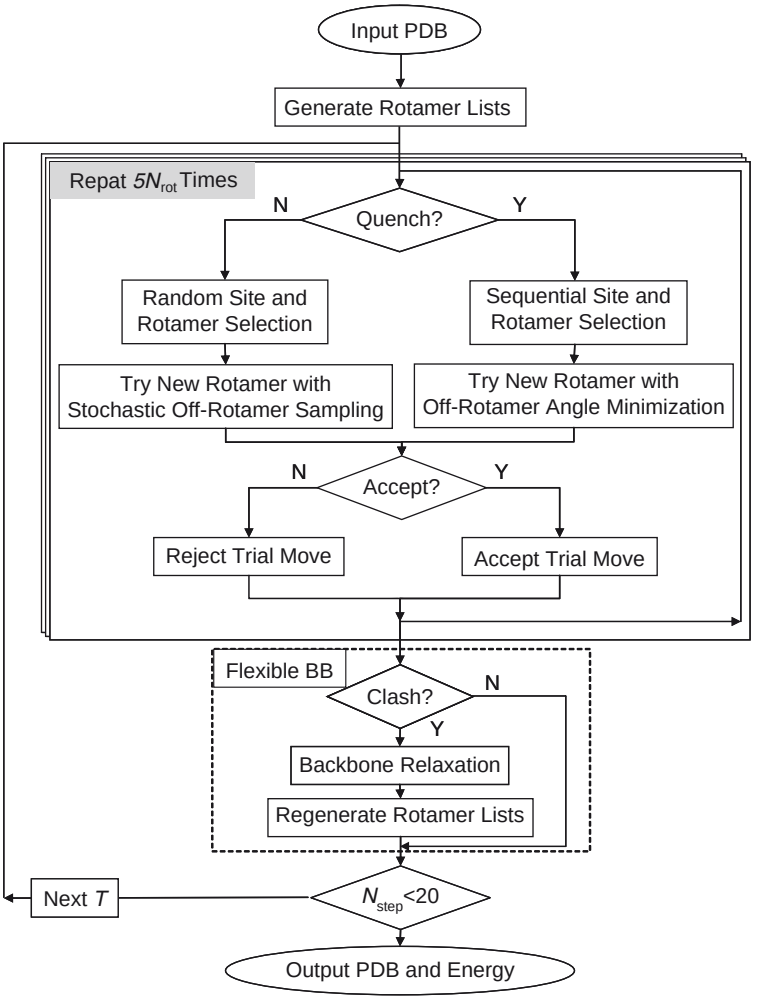


Fig. 2. Schematic illustration of the Eris energy minimization procedure using Monte Carlo simulated annealing.

the simulated annealing is performed in 20 stages. The starting and final MC temperatures are 10 kcal/R and 0.1 kcal/R, respectively, where R is the ideal gas constant. The MC temperature decays at a rate of $1/e$, where e is the base of the natural logarithm. The total number of MC iterations in each annealing stage is $5N_{rot}$, where N_{rot} is the total number of available rotamer states. In each MC iteration, a residue is randomly chosen to adopt a randomly selected rotameric state from the precalculated list, and this trial move of the rotamer state is accepted or rejected according to the Metropolis criterion (see Note 4).

3.2.5. Off-Rotamer Sampling

In Eris, we allow the rotamer dihedral angles to deviate from the average values as reported in the rotamer libraries. The maximum allowed deviation is one standard deviation for each dihedral angle.

During every MC iteration, the off-rotamer dihedral angles are sampled stochastically with a normal distribution. The center and width of the normal distribution are taken according to the average and standard deviation of dihedral angles for the rotamer as it is listed in the rotamer library (31).

3.2.6. Quenching

In the last MC-annealing stage, a quenching technique is applied. At this stage, the amino acid sites and rotamer states are selected sequentially and only the new trial step that will lower the total Medusa energy will be accepted. During quenching, the off-rotamer dihedral angles for each rotamer are also minimized during each trial move. The minimization step, which quickly searches the local energy minimum for the side chain dihedral angles, is performed using a conjugate gradient minimization algorithm.

3.2.7. Flexible-Backbone Protocol

The backbone dihedral angles are adjusted if severe clashes are detected during MC simulations. More specifically, the backbone relaxation is performed at the end of a specific MC stage when the acceptance rate during that stage is less than 0.05 and MC temperature is less than 0.12 kcal/R. Such a low acceptance rate usually indicates a large number of unresolved atomic clashes. If the backbone relaxation is considered necessary, the backbone dihedral angles (including ϕ, ψ, and ω) are optimized by a conjugate gradient minimization method at the end of the MC stage. Comparison of the energies shows that the flexible backbone protocol usually further lowers the total energy by approximately 10–20 kcal/mol when compared with a fixed backbone protocol (see Note 5).

3.2.8. Computational Techniques for Fast Energy Evaluation

To accelerate the computational speed of energy evaluations, Eris generates cubic grids to trace the coordinates of the atoms. The size of the grid lattices is set to be the maximum nonbonded interaction distance, so that only neighbor lattices need to be searched in order to calculate the interaction of an atom. Two sets of grids are used in Eris: a grid with lattice size of 9Å is used to evaluate van der Waals and solvation energies, and a fine grid with lattice size of 4Å is used to evaluate hydrogen-bonding interactions. To further facilitate the energy evaluations, the total energy of the protein is decomposed as pairwise interresidue contributions. After each MC iteration, only the pairwise interaction energies between that moved residue and other residues need to be updated. To prevent overcounting of the physical interactions, the calculation of nonbonded interaction (van der Waals, solvation and hydrogen bonding) does not include atomic pairs that are within the same residue and the neighboring peptide planes the residue connects to. Such interactions are presumably included in the backbone-dependent rotamer energies.

4. Notes

1. In the denatured state, the protein chain is in an extended configuration. As a result, each amino acid can adopt a large number of conformations. The number of possible conformations is significantly reduced in the more compact folded state. In other words, entropy is lost due to folding. Such entropic penalization is not considered explicitly in Eris protocols; however, it is partly counted by the reference energies (Eq. 1). In this simplest consideration, each type of amino acid has a reference energy, which represents a fixed amount of a free energy loss upon folding. In Eris, the reference energy has two sets of parameters. The original parameter set is determined by training over 34 high-resolution protein structures, so that the native amino sequences tend to have the lowest energy. However, parameters obtained in this way not only include the entropic penalty discussed above, but also incorporate additional factors such as the natural abundance of the amino acids. We separate out these contributions by fitting the reference energies over the $\Delta\Delta G$ experimental data.
2. In a protein structure, repulsive interactions prevent atoms from getting too close together. However, Eris may not be able to find an optimal configuration that can resolve all clashes due to the limited sampling of conformational space or due to inaccuracy in the backbone conformations. In this case, the output energy will be dominated by the unrealistically large van der Waals repulsion energy, and the calculated $\Delta\Delta G$ estimation will not correlate with the actual stability changes. We deal with the unresolved clashes using two approaches: First, in addition to the side chain degrees of freedom, we sample extra backbone conformations in the flexible backbone protocol. Preliminary tests have shown that the flexible backbone protocol can resolve some atomic clashes introduced by mutations involving the replacement of buried small residues by larger residues. Second, we monitor the total van der Waals repulsion energy terms from the $\Delta\Delta G$ calculation. If the repulsion energy is found to be larger than a predetermined cutoff value, the predicted $\Delta\Delta G$ will be considered unreliable and will not be reported by Eris. Based on our experiments of the protocols on multiple tests, we empirically set the van der Waals repulsion cutoff value to be 10 kcal/mol.
3. Eris does not model the interaction between protein and PDB-denoted heterogeneous atoms, including metal ions and ligands. Therefore, if the protein is stabilized by metals

or other ligands, the $\Delta\Delta G$ estimation may not be correct. In addition, Eris currently cannot model modified amino acids. The prediction accuracy may be affected if such modifications affect the stability of the protein.

4. Finding the optimal side chain packing is a considerable computational challenge, even with the simplifications of the nearly fixed backbone and discrete rotamer conformations. For the MC-simulated annealing algorithm used in Eris, we found that the sampling is sufficient for the proteins we tested, with lengths of up to 153 residues. Within this size range, the statistical uncertainty resulting from the insufficient sampling is less than 1 kcal/mol after averaging over the 20 MC runs. This statistical uncertainty increases as the protein size becomes larger, and roughly scales with $\sqrt{N}$, where N is the number of residues. Therefore, for a larger protein with approximately 1,000 residues, the statistical uncertainty after averaging over 20 runs can be as large as 2.6 kcal/mol. For larger proteins, the MC sampling can be trapped in a local minimum with a higher probability, thereby causing systematic error for $\Delta\Delta G$ evaluation. One possible solution to this convergence problem is to increase the number of MC runs to reduce the statistical uncertainty. In addition, we may also limit the sampling to residues that are in the proximity of the mutation sites, assuming that only those neighboring residues will be affected by the mutation. This can be implemented by fixing side chain rotamers of remote residues beyond a cutoff distance, or by editing the input protein structure and taking the fraction of the protein surrounding the mutation sites. From our experience, most $\Delta\Delta G$ estimation results will not change significantly when the cutoff distance increases to greater than 10Å.
5. The advantage of the flexible backbone protocol is that it provides extended sampling of backbone conformations. Although the backbone conformation sampling is still limited to small perturbations near the native structure, it can successfully remove some of the atomic clashes as indicated in benchmark results, and improves $\Delta\Delta G$ prediction accuracy in such cases. However, the extended sampling comes at an increase of nearly six times computational cost. Extended sampling also introduces additional errors from inappropriate backbone motions if the protein has missing backbone atoms. additional backbone motion also generates noise in $\Delta\Delta G$ estimations, even though there are no clashes introduced by mutations. Therefore, it is wise to apply the relatively faster fixed-backbone protocol first and apply flexible-backbone protocol only if there are clashes that cannot be solved with the backbone fixed.

6. Besides the web server, we also provide the Eris protocols in a stand-alone software package. The Eris stand-alone package includes the following functionalities:
   - *ΔΔG estimation*. Estimate the changes in protein stability upon point mutation, using either fixed or flexible backbone method.
   - *Fixed backbone design*. For a given protein backbone conformation, identify an amino sequence that minimizes energy of this conformation.
   - *Flexible backbone design*. For a given protein backbone conformation, identify an amino sequence that minimizes energy of this conformation with the ability to locally perturb the backbone dihedral angles.
   - *Scan*. For a given protein conformation and given amino acid positions, identify substitute amino acids that minimize this conformation.

Acknowledgments

We thank Elizabeth Proctor for reading the manuscript. The Eris stand-alone package is distributed by Molecules in Action LLC.

References

1. Taverna DM, Goldstein RA (2002) Why are proteins marginally stable? Proteins 46:105–109
2. Pakula AA, Sauer RT (1989) Genetic-analysis of protein stability and function. Annu Rev Genet 23:289–310
3. Becktel WJ, Schellman JA (1987) Protein stability curves. Biopolymers 26:1859–1877
4. Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor-2.1. Evidence for A 2-state transition. Biochemistry 30:10428–10435
5. Makhatadze GI, Privalov PL (1992) Protein interactions with urea and guanidinium chloride. A calorimetric study. J Mol Biol 226:491–505
6. Privalov PL (1979) Stability of proteins: small globular proteins. Adv Protein Chem 33:167–241
7. Jackson SE, Moracci M, elMasry N, Johnson CM, Fersht AR (1993) Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. Biochemistry 32:11259–11269
8. Bash PA, Singh UC, Langridge R, Kollman PA (1987) Free-energy calculations by computer-simulation. science 236:564–568
9. Dang LX, Merz KM, Kollman PA (1989) Free-energy calculations on protein stability – Thr-157-) Val-157 mutation of T4 lysozyme. J Am Chem Soc 111:8505–8508
10. Duan Y, Kollman PA (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 282:740–744
11. Vorobjev YN, Hermans J (1999) ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model. Biophys Chem 78:195–205
12. Kollman P (1993) Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev 93:2395–2417
13. Bordner AJ, Abagyan RA (2004) Large-scale prediction of protein geometry and stability changes for arbitrary single point mutations. Proteins 57:400–413
14. Capriotti E, Fariselli P, Calabrese R, Casadio R (2005) Predicting protein stability changes from sequences using support vector machines. Bioinformatics 21:54–58

15. Carter CW, LeFebvre BC, Cammer SA, Tropsha A, Edgell MH (2001) Four-body potentials reveal protein-specific correlations to stability changes caused by hydrophobic core mutations. J Mol Biol 311:625–638
16. Cheng JL, Randall A, Baldi P (2006) Prediction of protein stability changes for single-site mutations using support vector machines. Proteins 62:1125–1132
17. Gilis D, Rooman M (1996) Stability changes upon mutation of solvent-accessible residues in proteins evaluated by database-derived potentials. J Mol Biol 257:1112–1126
18. Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320:369–387
19. Hoppe C, Schomburg D (2005) Prediction of protein thermostability with a direction- and distance-dependent knowledge-based potential. Protein Sci 14:2682–2692
20. Khatun J, Khare SD, Dokholyan NV (2004) Can contact potentials reliably predict stability of proteins? J Mol Biol 336:1223–1238
21. Ota M, Isogai Y, Nishikawa K (2001) Knowledge-based potential defined for a rotamer library to design protein sequences. Protein Eng 14:557–564
22. Saraboji K, Gromiha MM, Ponnuswamy MN (2006) Average assignment method for predicting the stability of protein mutants. Biopolymers 82:80–92
23. Zhou HY, Zhou YQ (2003) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction (vol 11, pg 2714, 2002). Protein Sci 12:2121
24. Yin S, Ding F, Dokholyan NV (2007) Eris: an automated estimator of protein stability. Nat Methods 4:466
25. Yin S, Ding F, Dokholyan NV (2007) Modeling backbone flexibility improves protein stability estimation. Structure 15:1567–1576
26. Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A (2004) ProTherm, version 4.0: thermodynamic database for proteins and mutants. Nucleic Acids Res 32:D120–D121
27. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat T et al (2002) The protein data bank. Acta Cryst D58:899–907
28. DeLano WL (2002) The PyMOL molecular graphics system. Ref Type: Computer Program
29. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An all atom force-field for simulations of proteins and nucleic-acids. J Comput Chem 7:230–252
30. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616
31. Ding F, Dokholyan NV (2006) Emergence of protein fold families through rational design. PLoS Comput Biol 2:e85
32. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. Proteins 35:133–152
33. Kortemme T, Morozov AV, Baker D (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes. J Mol Biol 326:1239–1259
34. Dunbrack RL, Cohen FE (1997) Bayesian statistical analysis of protein side-chain rotamer Preferences. Protein Sci 6:1661–1681

Chapter 15

Approaches for Using Animal Models to Identify Loci That Genetically Interact with Human Disease-Causing Point Mutations

Josef D. Franke

Abstract

The complexity of human illnesses often extends beyond a single mutation in one gene. Mutations at other loci may act synergistically to affect the penetrance and severity of the associated clinical manifestations. Discovering the additional loci that contribute to an illness is a challenging problem. Animal models for disease, based on engineered point mutations in a homologous gene, have proven invaluable to better understand the mechanism(s) which give(s) rise to the observed physiological effects. Importantly, these animals can also function as the basis for genetic modifier screens to discover other loci which contribute to an illness. This chapter discusses the theory, considerations, and methodology for performing genetic modifier screens in animal models for human disease.

Key words: Modifier screen, Genetic interaction, Model organism, *MYH9*-related disorders

1. Introduction

Human diseases can rarely be attributed to a single mutation in a particular gene. While certain well-documented symptoms may be attributed to an individual point mutation, the molecular basis underlying the penetrance and expressivity of the associated clinical manifestations is often far more complex. Many of these variations are believed to result from an individual's genetic background. Polymorphisms at unlinked loci, which do not result in observable symptoms on their own, can interact with a mutant protein to influence the range and severity of the symptoms. Despite the complexity of this problem, the

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_15,

identification of individual point mutation(s) responsible for at least some of the clinical manifestations can facilitate more detailed investigations about the molecular basis of the disease. These point mutations can serve as extremely powerful genetic tools in second-site modifier screens (described later in more detail) enabling the discovery of those loci that act synergistically to cause an illness.

MYH9-related disorders provide an example of the potential for using point mutations to identify unlinked synergistic loci. *MYH9* encodes one of three human nonmuscle myosin II heavy chains, and particular point mutations in the *MYH9* gene result in a spectrum of autosomal dominant, hematologic illnesses (collectively termed *MYH9*-related disorders) (1–3). Most individuals have enlarged platelets (macrocytosis), a reduction in platelet number (thrombocytopenia) and leukocyte inclusions. Alone, these hematologic manifestations are generally mild and are often only discovered when routine blood smears are performed due to symptoms of an unrelated illness. Interestingly, despite having an identical *MYH9* mutation, some individuals exhibit more severe nonhematologic clinical manifestations (high-tone, sensorineural deafness, cataracts, and nephritis). The severity and occurrence of these nonhematologic manifestations vary between individuals and can be life-threatening. It has been hypothesized that the variability and severity of these more severe manifestations is most likely due to genetic interactions of the mutant *MYH9* locus with unlinked polymorphic loci (3, 4). Identifying these loci is an extremely difficult undertaking in humans.

The development of disease models in genetically tractable animal models offers several advantages to studies in humans, including facile genetic screening approaches; easy identification of interacting loci; fast generation time; and the potential for initial drug and/or therapeutic tests. Establishing disease models can be a difficult and timely pursuit as homologous, or orthologous, mutations in a different organism can result in very subtle or unexpected phenotypes. Even if the phenotypic effects are subtle, the generation of a disease model can be very informative as the underlying genetic causes are often conserved. Second-site modifier screens often enable one to understand the inherent genetic complexity of a particular disease and provide candidate loci for further investigations. This chapter outlines experimental approaches to help researchers elucidate whether particular point mutations can serve as the basis for genetic screens and how such screens can be performed. Our work on the development of a model for *MYH9*-related disorders in the fruit fly, *Drosophila melanogaster*, is given as an example throughout the chapter.

2. Materials

2.1. Determining the Residues to Mutate and Making Site-Directed Mutants

The first step in generating an animal model for disease is to identify disease-causing mutations in the gene of interest. For many diseases, the OMIM database (*O*nline *M*endelian *I*nheritance in *M*an; http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM) is an excellent, fully referenced resource that provides background information about gene function, phenotypes and the different allelic variants of the desired loci. The next step is to identify the homologous or analogous amino acid residue(s) to mutate in the animal model of choice. Common alignment programs (e.g., MegAlign by DNAStar) are an excellent tool for determining if the amino acid residues in question are well conserved between humans and other animals. When several human disease-causing mutations have been identified, it is best to choose those amino acids which are most conserved between humans and the model animal of choice. Ideally, these amino acids will sit within a region of high homology between the two proteins. In our study, alignment of the human *MYH9* gene with the only *D. melanogaster* nonmuscle myosin II heavy chain encoded by the *zipper* locus (5) clearly identified analogous amino acid residues for each of the four residues of interest (6, see Fig. 1 for examples of alignments around each chosen amino acid residue).

Several excellent sources discuss the molecular details of performing site-directed mutagenesis and so we do not discuss the topic. However, we do want to stress that it is essential to sequence the entire gene completely after the mutagenesis is performed to ensure that no unintended mutations arose from the mutagenesis procedure.

2.2. Selecting the Expression System

It is difficult to predict the physiological effects of introducing an analogous, disease-causing mutation into a different animal, so care must be taken when selecting an expression system for the mutant gene. Constitutive promoters (heat-shock or ubiquitous), therefore, may not be the best choice when more specific, conditional expression systems (such as tissue-specific promoters) are available. As generating transgenic animals is a significant time investment, it is important to consider which type of expression method is ideal for the intended study. Generally, selecting the expression system providing the greatest control over expression conditions is preferable. The most common conditional expression systems in model animals are the Gal4–UAS system in *D. melanogaster* (7) and Cre-lox in the mouse, *Mus musculus* (8). In the worm *Caenorhabditis elegans*, transgenes are most often generated by placing known tissue-specific promoters upstream

of the gene of interest, which are then inserted into the genome. To date, conditional expression is infrequently used in zebrafish (*Danio rerio*) as detailed knowledge of tissue-specific promoters is still under investigation (9).

Ideally, a mutant form of the gene under a conditional promoter would be inserted by homologous recombination to replace the wild-type gene at the endogenous locus; unfortunately, this approach is currently only tractable in mice. In other model animals, mutant genes are inserted randomly into the genome as stable transgenes. Due to this random insertion, transgenes are subject to position effect variegations. This can affect the expression level of identical transgenes in two different transgenic animals if the transgenes have different insertion sites in the genome. As a consequence, when random transgene insertions are employed, one must be careful to identify a transgenic line in which expression closely matches wild-type levels.

When we studied *MYH9*-like mutations in the *D. melanogaster zipper* gene we utilized the Gal4–UAS transgene expression system which provides temporal and tissue-specific control of transgene expression. This system takes advantage of a yeast transcription factor (Gal4) and its DNA recognition sequence (UAS repeats). A transgenic fly line which places an endogenous fruit fly promoter in front of the Gal4 coding sequence is generated. A separate transgenic line is generated in which a gene of interest is placed downstream of UAS repeats. As the Gal4 protein and UAS repeats are not endogenous to fruit flies, neither transgenic line suffers any adverse affects. In one line, the Gal4 protein is conditionally expressed by the selected fruit fly promoter but is innocuous as binding sites are lacking. In the other line, the gene of interest is not expressed as there is no transcription factor capable of binding the upstream UAS repeats. Only when flies from each line are mated do the offspring contain both the promoter-Gal4 transgene and the UAS-transgene, thereby resulting in the conditional expression of a gene of interest.

2.3. Comparing the In Vivo Characteristics of Mutant Protein(s) to Wild Type

It is important to characterize the effect of the introduced mutation(s) on the protein once transgenic animals have been generated. Any change in the amino acid sequence of a protein can affect its stability and/or subcellular distribution. As such, it is necessary to quantitate and observe the localization of the mutant protein(s) and compare it to a similarly expressed wild-type protein. The most common approach is to engineer a tag, either a small epitope (e.g., myc or HA) or a larger tag, like GFP, to either the N- or C-terminus of each protein. Such tags can be used to assay protein expression levels and cellular distribution provided the tag itself does not affect the protein.

Western blotting is the most straightforward means to quantitate expression levels. Mutant protein expression levels can differ significantly from wild type, so it may be necessary to screen through multiple, independent insertion lines to find one that is most similar to wild type. If multiple mutant lines consistently show reduced expression, then the mutation most likely affects protein stability and any associated phenotypes could simply be the result of reduced protein levels. For mutations which do not affect stability, the next step is to compare the subcellular distribution of the mutant to wild-type proteins.

Immunofluorescence on fixed samples and live imaging of GFP-fusion proteins are the most straightforward approaches. Three major outcomes are possible: (1) if the mutation affects protein folding, then aggregation will often result in bright foci within cells; (2) if the mutation affects targeting, then a clear difference in subcellular distribution between the mutant protein(s) and wild type will occur; and (3) if neither folding nor targeting is affected, then mutant and wild-type distribution should be identical. If the mutant protein(s) do not affect protein stability or subcellular distribution, then the next step is to begin a genetic characterization.

Mutations that affect stability, folding or distribution can be used in genetic studies, but analysis is more complicated. For example, a point mutation that results in folding defects will likely recruit chaperone proteins. In a genetic screen, this mutation would likely pull out chaperones and/or genes responsible for chaperone regulation, but these loci may, or may not, be linked to the disease in question. Additional experimental evidence will be necessary in order to show a *direct* genetic link with candidate loci. As a result, mutations that affect stability, folding or subcellular distribution will often result in an increased rate of false-positives.

In examining specific point mutations in the *zipper* gene, we engineered an N-terminal GFP tag into each of our transgenes (wild type and four point mutants). We selected this tag and its N-terminal location as previous studies demonstrated that an N-terminal GFP does not affect nonmuscle myosin II function. To quantitate expression levels of each mutant protein to wild type, each was expressed ubiquitously. Whole cell lysates were generated and then probed with an anti-GFP antibody. As each protein was a GFP-fusion, video time-lapse microscopy was performed to examine the subcellular distribution of each. We selected a 3-h interval of embryonic development in which the distribution of nonmuscle myosin II has been well characterized. No adverse effects in expression or distribution were observed for any of the mutations.

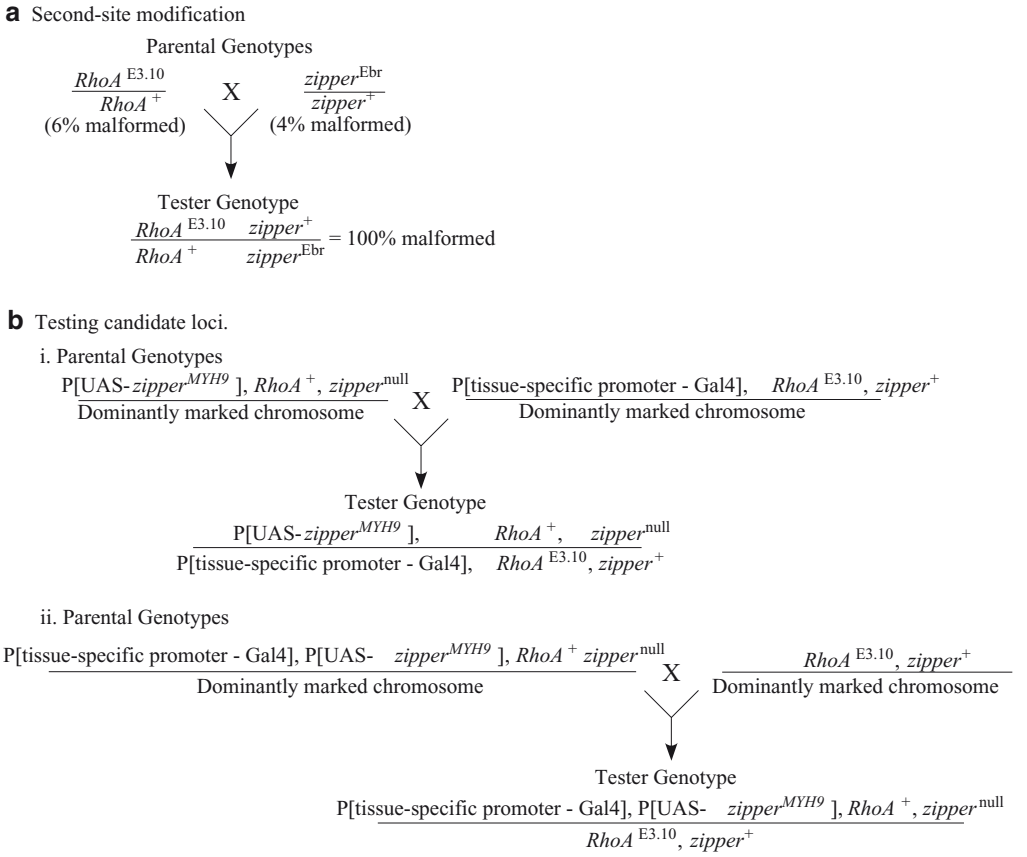


Fig. 1. (**a**) Example of second-site modification. On their own, *RhoA* heterozygous animals (*RhoA*$^{E3.10}$/*RhoA*$^{+}$) and *zipper* heterozygotous animals (*zipper*^{Ebr}/*zipper*$^{+}$) are predominantly wild type in appearance with a small percentage of having malformations. When these animals are mated, all of the doubly heterozygous progeny (tester genotype; *RhoA* heterozygous and *zipper* heterozygous) display a malformed phenotype (11). (**b**) Two examples of crosses to determine if a site-directed mutation in one gene (*zipper*) sensitizes the background such that a mutant allele of a candidate gene (*RhoA*) acts as a second-site modifier to produce novel phenotypes: (*i*) A site-directed mutation in the *zipper* gene (P[UAS-*zipper*^{MYH9}]) was recombined into a chromosome with a *zipper* null allele and a wild-type *RhoA* allele. These animals were then mated to animals having a chromosome with a mutant *RhoA* allele (*RhoA*$^{E3.10}$), a wild-type *zipper* allele and a transgene with a tissue-specific promoter causing expression of the site-directed mutant protein in the tester progeny; (*ii*) The P[UAS-*zipper*^{MYH9}] transgene was recombined into a chromosome with a *zipper* null allele, a wild-type *RhoA* allele and a tissue-specific promoter to drive expression of the site mutant. These animals were mated to animals with a mutant *RhoA* allele (*RhoA*$^{E3.10}$) and a wild-type *zipper* allele. In both schemes, dominantly marked chromosomes were used in each parent in order to unambiguously distinguish the tester animals from other progeny.

3. Methods

3.1. The Basics of Second-Site Modifier Screens

Once stable transgenic lines have been generated and certain characteristics of the mutant protein(s) have been studied (e.g., stability and subcellular distribution), the next step is to determine whether a mutation can function as a unique allele for genetic interaction studies. These studies are often referred to

as second-site modifier screens. The first "site" being a known gene of interest with a well-characterized mutant allele or, for the purposes of this chapter, a desired mutation generated by site-directed mutagenesis. Animals heterozygous for the mutation are most commonly studied as they likely show no phenotypic abnormalities unless the locus displays haploinsufficiency or the particular allele has some dominant effects. A screen is then performed such that matings generate tester animals that are heterozygous at the first locus and have an additional mutation at another, unknown locus. Each tester animal is therefore a complex heterozygote (Fig. 1a) and is examined for phenotypes. A positive hit in the screen occurs when either a new phenotype is observed or there is an increase in the expressivity or penetrance of an existing phenotype. The cause for the phenotypic difference is the modified genetic background that now includes a mutation at second locus – hence the term second-site modifier. Using molecular and genetic tools, the locus and mutation of a second-site modifier can be identified.

Perhaps the most crucial element to any second-site modifier screen is the nature of the mutation at the first locus. This mutation must "sensitize" the genetic background such that it is susceptible to producing discernible phenotypic effects when crossed into different mutant backgrounds. Unfortunately, not every allele of a locus will result in a sensitized genetic background. For example, Halsell et al. demonstrated that different alleles of the *zipper* locus sensitized the background to different extents such that some alleles were much better suited for second-site modifier screens (10). Even more unfortunate is that it is not possible to predict which allele(s) will work best; therefore, it is recommended that more than one site-directed mutant be generated when possible. In trying to establish an *MYH9*-related disorder model in *D. melanogaster*, we identified four different well-conserved amino acid residues between the human *MYH9* and the *D. melanogaster zipper* gene. As it was not possible to predict the physiological effects of any of these mutations, we therefore generated and tested each separately.

3.2. Generating Parental Animals with the Necessary Genetic Backgrounds

It is important to mimic a traditional heterozygous genetic background as close as possible for second-site modifier screens employing a site-directed mutant located on a transgene. The best approach is to generate a chromosome that contains both the transgene and a null allele of that gene at the endogenous locus. The method for generating such a chromosome can vary depending on the animal model, but is generally very straightforward.

The use of a conditional, tissue-specific promoter allows internal negative control cells (those cells that do not express the transgene) within the same animal (Fig. 2a, b). These internal controls simplify the phenotypic examination as both the

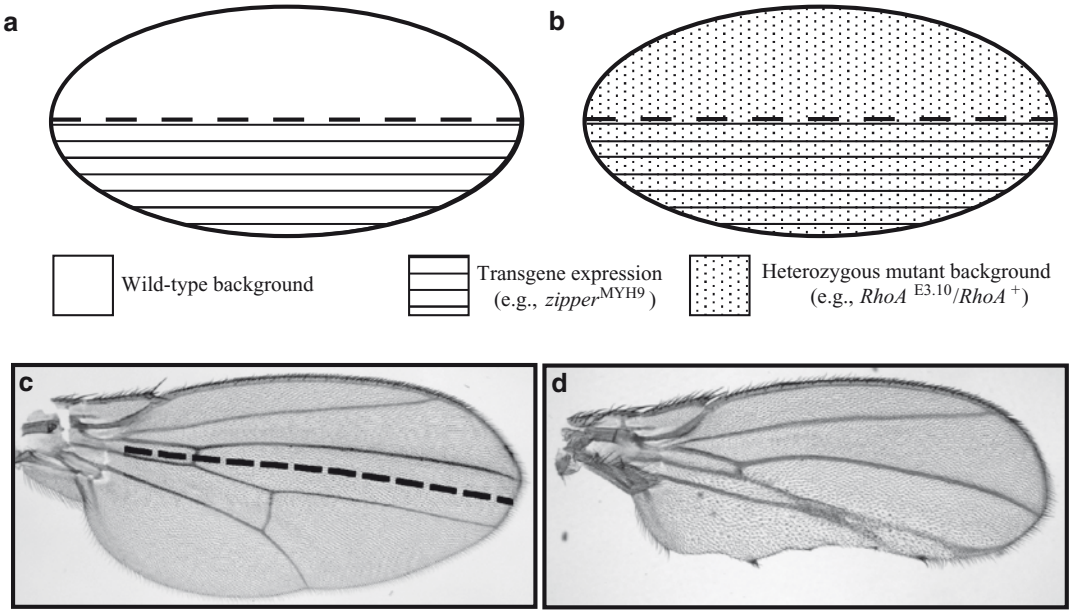


Fig. 2. The use of tissue-specific expression of transgenes as internal controls determines the specificity of phenotypes. (**a**) Examining the effect of expressing the site-directed mutant in a wild-type background. The *oval shape* represents a hypothetical wild-type tissue in which a tissue-specific promoter has expressed a transgene encoding a desired point-mutation (*horizontal lines*). By comparing the lower half with the upper half of the tissue, one can determine if the expression of the mutant protein causes altered phenotypes. (**b**) Examining the effect of expressing a site-directed mutant in a potential second-site modifier background. The *oval shape* represents a hypothetical tissue in which all cells are heterozygous for potential second-site modifier. A tissue-specific promoter was used to express a transgene encoding a desired point mutation. By comparing the lower half with the upper half of the tissue, one can determine if observed phenotype(s) are general to the heterozygous background (*top*) or the result of a genetic interaction due to the expression of the mutant protein (*bottom*). (**c**) Image of a wild-type *D. melanogaster* wing in which a point mutation in the *zipper* gene (UAS-*zipper*^{MYH9}) was expressed under the *engrailed* promoter (*engrailed*-Gal4; below the *dashed line*). The anterior region (above the *dotted line*) of the wing functions as an internal control. (**d**) Example of a positive second-site modifier interaction. The entire wing is heterozygous at the RhoA locus (*RhoA*$^{E3.10}$/*RhoA*$^{+}$). A point mutation in the *zipper* gene (UAS-*zipper*^{MYH9}) was expressed under the *engrailed* promoter (*engrailed*-Gal4; below the *dashed line*). As phenotypes are confined to the region where the transgene was expressed, one can conclude that these phenotypes result from a genetic interaction between the *RhoA* and *zipper* locuses.

experimental and control cells can be viewed simultaneously. If expression is not limited to a subset of cells, then animals with slightly different genetic backgrounds must be used as control subjects, which increases the possibility of both missed interactions and false-positive interactions.

In our *D. melanogaster* studies, we used standard fruit fly recombination matings to generate a chromosome containing a null *zipper* allele and a desired transgene (P[UAS-*zipper*^{MYH9}]; Fig. 1b; parental genotype). As *zipper* is a ubiquitously expressed gene, we selected a well-characterized Gal4 promoter line (*engrailed*-Gal4) to drive transgene expression in a subset of cells. In particular, the *engrailed* promoter has a very specific expression pattern in certain adult fly appendages (e.g., the posterior but not

the anterior part of the wing), allowing easy determination of control and experimental regions.

3.3. Testing the Ability of Site-Directed Mutations to Function in Second-Site Modifier Screens

Before embarking on a large-scale genetic screen, we strongly recommend examining the mutant phenotype when the site-directed mutant is expressed in known effector, or candidate, genetic backgrounds. Such candidate genes could be identified by several means: including; (1) previous genetic interaction studies linking the two loci; (2) biochemical studies showing a physical interaction between the protein of interest and another protein; or (3) if the protein of interest is a known participant in some pathway (e.g., a signal transduction cascade), then other members of the pathway are candidate genes.

Once candidate genes have been identified, test matings with mutant alleles of these genes should be set up to directly examine each for an interaction with the site-directed mutant protein (Fig. 1b). A genetic interaction should not be expected with every mutant allele of each candidate. Rather, any observable genetic interaction supports the use of the mutation in the second-site modifier screen. The goal of these candidate crosses is twofold. First, one must identify whether or not the site-directed mutant(s) sensitizes the genetic background such that second-site modifiers can be identified. Second, one must document and become familiar with the type and penetrance of phenotypes that may result from genetic interaction studies. Phenotypes can be subtle and may only be present in a percentage of the tester animals; therefore these candidate crosses are important for training the eyes.

We examined the ability of each *MYH9*-like mutation introduced into the *zipper* gene to sensitize the genetic background. Each protein (four mutants and the wild type) were individually expressed in a genetic background previously shown to display a second-site modifier interaction with certain *zipper* alleles. We observed that each *MYH9*-like mutation in *zipper* was capable of sensitizing the background such that in a specific genetic background, phenotypes were observed in tissues expressing the mutant proteins (Fig. 2d). We also found that they did so to different extents, such that certain *MYH9*-like mutations would likely function better in a second-site modifier screen.

3.4. Large-Scale Genetic Screens

Having confirmed that a site-directed mutant can sensitize the genetic background to identify second-site modifiers, the next step is to scale up and perform a genome-wide search for loci that show a genetic interaction. The exact crossing scheme used in a large-scale screen will depend on many factors, including the animal model and the method of transgene expression. The researcher will need to decide how to tailor the screen to suit their needs (see Note 1). The parental genotypes in the

screen will be identical to the candidate loci example (see Fig. 1b) except that the alleles to be tested will no longer be candidates. Newly mutagenized chromosomes (e.g., those generated by ethane methyl sulfonate, EMS) are often a common starting point in screens, but it is important to remember that identifying the molecular lesion responsible for the genetic interaction on a newly mutagenized chromosome is a significant time investment (see Note 1).

Alternatively, many organisms have stock centers where animals with characterized mutant alleles can be ordered and directly tested. In addition, some animals, including *D. melanogaster*, have deficiency lines that together span most of the genome. Animals in each deficiency line are missing a known segment of a chromosome which can span anywhere from ten to more than a hundred genes. Thus, deficiency lines can be used to simultaneously screen for a genetic interaction with many loci (10). Additionally, lack of interaction with a deficiency line can eliminate a large number of genes in a single cross. If an interaction is observed within a deficiency region, then mutant alleles for genes within the deficiency can be tested.

3.5. Screening for Disease-Contributing Nongenetic Factors

It is easy to forget that factors affecting the spectrum and severity of human pathologies are not always genetic in nature. It is important to broaden the scope of candidate factors to include environmental interactions and stresses (e.g., temperature, circadian cycles, caloric intake, physical exertion, diet, and other environmental factors). These factors could act in concert with known disease-causing mutations resulting in additional or more severe clinical manifestations. Studying the effect(s) of environmental factors in a laboratory setting can be straightforward and should not be ignored, especially once the time and effort have been invested to develop a disease model. It is important to consider any or all factors that might influence the range and severity of symptoms associated with different pathologies to increase our limited knowledge of factors contributing to the range and severity of disease symptoms.

4. Note

1. Designing a genetic screen and then mapping an unknown mutation, generated via random mutagenesis, to a gene is a time-intensive process. Depending on the organism, different approaches are available. Included here are sources which describe in detail the preferred methodologies available for each organism: for *C. elegans*, *C. elegans: a practical approach*,

by Ian A. Pope (editor); for *D. melanogaster*, *Fly pushing: the theory and practice of drosophila genetics*, by Ralph Greenspan; for *D. rerio*, *Zebrafish*, edited by Christiane Nusslein-Volhard and Ralf Dahm; for *Mus musculus*, *The laboratory mouse*, edited by Hans Hedrich.

Acknowledgments

The author would like to thank Jill Blankenship, Jaclyn Novatt, Jody Plank, and Brent Berwin for critically reading and providing valuable suggestions. This work was supported by NIH F32GM082029.

References

1. Kunishima S, Kojima T, Matsushita T, Tanaka T, Tsurusawa M, Furukawa Y, Nakamura Y, Okamura T, Amemiya N, Nakayama T, Kamiya T, Saito H (2001) Mutations in the NMMHC-A gene cause autosomal dominant macrothrombocytopenia with leukocyte inclusions (May-Hegglin anomaly/Sebastian syndrome). Blood 97(4):1147–1149
2. Kelley MJ, Jawien W, Ortel TL, Korczak JF (2000) Mutation of MYH9, encoding non-muscle myosin heavy chain A, in May-Hegglin anomaly. Nat Genet 26(1):106–108
3. Seri M, Pecci A, Di Bari F, Cusano R, Savino M, Panza E, Nigro A, Noris P, Gangarossa S, Rocca B, Gresele P, Bizzaro N, Malatesta P, Koivisto PA, Longo I, Musso R, Pecoraro C, Iolascon A, Magrini U, Rodriguez Soriano J, Renieri A, Ghiggeri GM, Ravazzolo R, Balduini CL, Savoia A (2003) MYH9-related disease: May-Hegglin anomaly, Sebastian syndrome, Fechtner syndrome, and Epstein syndrome are not distinct entities but represent a variable expression of a single illness. Medicine (Baltimore) 82(3):203–215
4. Balduini CL, Iolascon A, Savoia A (2002) Inherited thrombocytopenias: from genes to therapy. Haematologica 87(8):860–880
5. Young PE, Richman AM, Ketchum AS, Kiehart DP (1993) Morphogenesis in *Drosophila* requires nonmuscle myosin heavy chain function. Genes Dev 7:29–41
6. Franke JD, Montague RA, Rickoll WL, Kiehart DP (2007) An MYH9 human disease model in flies: site-directed mutagenesis of the Drosophila non-muscle myosin II results in hypomorphic alleles with dominant character. Hum Mol Genet 16(24):3160–3173
7. Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118:401–415
8. Sauer B (1998) Inducible gene targeting in mice using the Cre/lox system. Methods 14(4):381–392
9. Deiters A, Yoder JA (2006) Conditional transgene and gene targeting methodologies in zebrafish. Zebrafish 3(4):415–429
10. Halsell SR, Kiehart DP (1998) Second-site noncomplementation identifies genomic regions required for *Drosophila* nonmuscle myosin function during morphogenesis. Genetics 148(4):1845–1863
11. Halsell SR, Chu B, Kiehart DP (2000) Genetic analysis demonstrates a direct link between Rho signaling and nonmuscle myosin function during *Drosophila* morphogenesis. Genetics 155:1253–1265

Part IV

Protein Evolution Mutagenesis

Chapter 16

Using Peptide Loop Insertion Mutagenesis for the Evolution of Proteins

Christian Heinis and Kai Johnsson

Abstract

The insertion of peptide loops into the polypeptide chain of proteins at surface-exposed regions is an attractive avenue to modify the protein's properties or to evolve new functionalities. The strategy of peptide loop insertion has, for example, been used to create new binding sites in protein scaffolds and has led to the isolation of proteins with excellent binding affinities for various biological structures. Peptide loops have also been inserted into enzymes to modulate their catalytic properties. We recently used loop insertion mutagenesis to evolve a mutant of O^6-alkylguanine-DNA-alkyltransferase (AGT) that reacts with the nonnatural substrate O^6-propargylguanine for applications in molecular imaging. In this chapter, we describe in detail the protocols that we have applied (1) to identify sites in AGT that are permissive to loop insertion, (2) to manipulate DNA to create large loop insertion libraries, and (3) to identify mutants with desired properties. The experimental procedures are general and can easily be adapted for the insertion of peptide loops into other classes of enzymes or into any protein of interest.

Key words: Loop insertion mutagenesis, Directed evolution, Enzyme, Substrate specificity, O^6-alkylguanine-DNA-alkyltransferase

1. Introduction

The inherent stability of native proteins and their flexibility for sequence variations allows the insertion of supplementary peptide loops while preserving their overall structure and function. The inserted peptides can alter the properties of the proteins or exert new functionalities. The strategy of peptide loop insertion is, for example, used for the generation of proteins with new binding specificities. Towards this end, either rational or directed evolution approaches are followed. In a rational approach, peptide motifs with known binding specificities are grafted into

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_16,

a protein. Examples are the creation of metalloproteins through the insertion of metal ion-binding peptides (reviewed in (1)) or the grafting of linear epitopes of antigens (2, 3) into unrelated protein scaffolds. In a directed evolution approach, large libraries of proteins that display on their surface one or multiple random peptide loops are generated and variants that bind to a protein of interest are identified in screening or selection procedures. Examples for proteins that were used as supporting frameworks for the display of peptide loops in directed evolution approaches are the tenth fibronectin type III domain of human fibronectin (FNfn10) (4–6), the lipocalin (7), the squash trypsin inhibitor EETI-II (8), the cytotoxic T lymphocyte-associated antigen 4 (CTLA-4) (9), and the tendamistat (10). The strategy of peptide loop insertion is also used to modulate the catalytic properties of enzymes. Additional amino acids were, for example, inserted into the polypeptide chain of the TEM 1 β-galactosidase to study the effect of amino acid insertions (11), to modulate the activity (12, 13), or to obtain mutants with activities towards new β-lactam antibiotics (14, 15). We recently inserted peptide loops into the active site region of O^6-alkylguanine-DNA-alkyltransferase (AGT) to alter its substrate specificity and to use it in molecular imaging applications (16).

Different methodological aspects need to be considered when a peptide loop insertion mutagenesis strategy is planned. These considerations include the site of loop insertion, the number of inserted loops, the length of the loops, and the mutagenesis method. Potentially attractive regions for peptide insertions are surface-exposed loops that link stable structural elements like α-helices or β-sheets. It is recommended to assess in a preliminary experiment whether a specific site is permissive to loop insertions. This is done by inserting constant or random peptide sequences at the sites of interest and testing the stability and functionality of the modified proteins (4). Alternatively, suitable sites are identified through the insertion of peptide loops at completely random positions and by analysis of the integrity of these mutants (17). For the creation of proteins with new binding sites, multiple peptide loops (typically two to three) are inserted in positions of a protein that are in close proximity to generate a contiguous binding surface. For the modulation of the catalytic properties in enzymes, peptide loops are preferentially inserted close to the active site region of the catalyst. The length of the inserted peptides can be chosen arbitrarily or can be set to be random. Good results were obtained in approaches where five to ten additional amino acids were inserted into the protein scaffolds. Methods for the generation of DNA with nucleotide insertions are manifold. Well-established methodologies are the Kunkel mutagenesis (18) and overlap extension PCR mutagenesis (19).

2. Materials

2.1. Identification of Suitable Sites for the Insertion of Peptide Loops

1. ^{M}AGT is an engineered AGT mutant (20) with improved expression properties, with reduced affinity for DNA and with increased activity towards O^6-benzylguanine (BG) substrates. The protein contains 182 amino acids.
2. The following DNA primers are used: p1 (5′-CGAAA TGGATCCATGGACAAGG ATTGTGAAATG-3′), p2 (5′-CCACGGCTCCATTGATATTG-3′), p3 (5′-CAAT AT CAATGGAGCCGTGGGCGGTTACNNKNNKNNKNN KNNKNNKNNKGGACTGGCCGTGAAGGAATG-3′), and p4 (5′-GCCTTTGAATTCCTTAATGATGGTG ATGATGG TGTCCCAAGCCTGGCTTCCC-3′). N represents adenine, thymine, guanine, or cytosine nucleotides, and K represents a thymine or a guanine nucleotide. The restriction sites of *BamH*I and *EcoR*I are underlined.
3. PCR buffer (10×): 100 mM Tris–HCl, pH 8.5, 500 mM KCl, 15 mM MgCl.
4. DNA is extracted from agarose gels with a gel extraction kit (Qiagen).
5. Buffer for the digestion of DNA with *BamH*I and *EcoR*I (final concentration): 50 mM, Tris–HCl, pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, 0.1 mg/mL BSA.
6. DNA treated with enzymes is purified with a PCR purification kit (Qiagen).
7. DNA ligation buffer (final concentration): 50 mM Tris–HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 10 mM dithiothreitol.
8. The vector pGEX-2T (GE Healthcare Life Sciences) is used for the expression of AGT mutants as fusion proteins with GST (AGT is linked to the C-terminal end of GST).
9. The *E. coli* strain XL1-Blue (Stratagene) is used as an expression host.
10. 2YT medium: 16 g bactotryptone, 5 g yeast extract, 5 g NaCl, 1 L H_2O.
11. SB medium: 35 g bactotryptone, 20 g yeast extract, 5 g NaCl, 1 L H_2O, adjust pH to 7.5 with NaOH.
12. Lysis buffer: 50 mM Tris–HCl, pH 8, 100 mM NaCl, 5 mM EDTA, and 1 mg/mL lysozyme.
13. Dithiothreitol (DTT) stock solution: 100 mM DTT in H_2O.
14. The AGT substrate BGFL is a conjugate of O^6-benzylguanine and fluorescein (21). A tenfold stock solution of BGFL: Dilute BGFL (2 mM, in DMSO) in H_2O to a final concentration of 20 μM.

15. Benzylguanine (BG; Sigma-Aldrich): 5 mM stock: 6 mg BG in 1 mL DMSO.
16. Protein loading buffer (reducing, 5×): 240 mM Tris–HCl, pH 6.8, 6% (w/v) SDS, 30% glycerol (w/v), 1.2 M β-mercaptoethanol, 0.3% bromophenol blue (w/v) in H_2O.
17. Running buffer (10×): 0.25 M Tris–HCl, pH 8.3, 2 M glycine, 1% SDS.
18. Fluorescence imager (Image Station 440, Kodak).

2.2. Cloning of Peptide Loop Insertion Libraries and Display on Phage

1. Phagemid vector pAK100 (a 6.4-kb plasmid; obtained from the research group of Professor A. Plückthun, University of Zurich) is used for the display of AGT on phage (22). The vector contains a chloramphenicol resistance gene and the genes for the expression of the gene-3-protein and a leader sequence for the export of the gene-3-protein to the periplasmic space. A tetracycline resistance gene (2.1 kb) between the leader sequence and the gene 3 is flanked by two *Sfi*I restriction sites and can be replaced by a gene of interest. The two *Sfi*I sites have two different nonpalindromic sequences that assure the insertion of genes in the correct orientation.
2. The primers p1, p2 (p1 and p2 are described in Subheading 2.1), p5 (5′-CGATCCTTAGACCTGAACGCAGGTTTCCCGACTGGAAAG-3′), and p6 (5′-GCGTCAGGGTTACAAGTTCATGGTTTACCAGCGCCAAAG-3′) are used.
3. The PCR buffer and the ligation buffer described in Subheading 2.1 are used.
4. Buffer for digestion with *Sfi*I (final concentration): 10 mM, Tris–HCl pH 7.9, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, BSA 0.1 mg/mL.
5. The *E. coli* strain JM101 is used as a host for the production of phage (22).
6. Helper phage M13KO7 (New England Biolabs) is an M13 derivative, which carries a kanamycin resistance gene. M13KO7 is able to replicate in the absence of phagemid DNA. In the presence of a phagemid bearing a wild-type M13 or f1 origin (as pAK100), single-stranded phagemid is packaged preferentially and secreted into the culture medium.
7. The 2YT media described in Subheading 2.1 is used.
8. PEG buffer: 20% (w/v) PEG 8000, 2.5 M NaCl.

2.3. Phage Selection to Identify Loop Insertion Variants with Desired Activities

1. Phosphate-buffered saline (PBS): 137 mM NaCl, 4.3 mM Na_2HPO_4, 1.7 mM KH_2PO_4, 2.7 mM KCl, pH 7.4.
2. PBS/milk: 4% (w/v) skimmed milk to PBS (137 mM NaCl, 4.3 mM Na_2HPO_4, 1.7 mM KH_2PO_4, 2.7 mM KCl, pH 7.4).

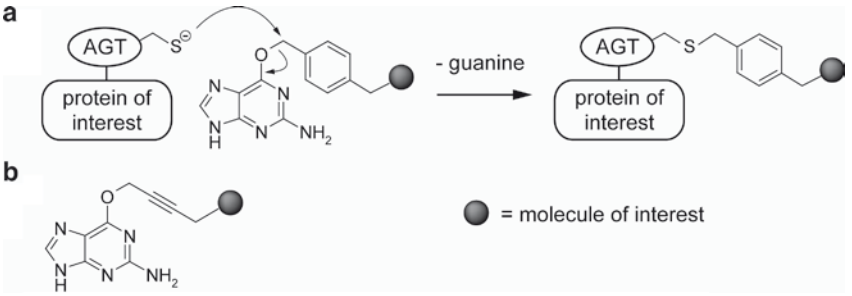


Fig. 1. Labeling of AGT fusion proteins with chemical probes. (**a**) General mechanism of the reaction of AGT with O^6-benzylguanine (BG) substrates. A label that is linked to the benzyl group of BG is transferred to the active cysteine residue of AGT. (**b**) O^6-propargylguanine (PG), a poor substrate of wild-type like AGT (^{M}AGT). An AGT mutant that reacts efficiently with the nonnatural substrate was engineered using peptide loop insertion mutagenesis.

3. O^6-propargylguanine-digoxigenin (PGDIG; see Fig. 1b) stock solution: 2 mM in DMSO.
4. The PEG buffer described in Subheading 2.2 is used.
5. Magnetic beads coated with antidigoxigenin antibody (Roche Applied Science) are blocked by incubation of 50 μL bead suspension in 0.5 mL PBS/milk for 30 min at room temperature. Magnetic antidigoxigenin beads can also be produced by reacting antidigoxigenin antibody (Roche Applied Science) with tosyl-activated magnetic beads (Dynal M-280, Invitrogen) according to the supplier's instructions.
6. PBS/Tween-20: Tween-20 (0.01% (v/v)) in PBS.
7. Acidic elution buffer: 100 mM glycine, pH 2.2.
8. Neutralization buffer: 1 M Tris–HCl, pH 8.

2.4. Characterization of Isolated Loop Insertion Mutants

1. Buffers as PBS and PBS/milk are described in Subheading 2.3.
2. The primers p1 and p4, the vector pGEX-2T and the reagents for the PCR reaction, the restriction of DNA, and the ligation are described in Subheading 2.1.
3. The 2YT and SB media are described in Subheading 2.1.
4. IMAC lysis buffer: 50 mM Tris–HCl, pH 8, 300 mM NaCl, 10 mM imidazole, and 1 mg/mL lysozyme.
5. IMAC washing buffer: 50 mM Tris–HCl, pH 8, 300 mM NaCl, 10 mM imidazole.
6. Ni-NTA slurry (Qiagen) is washed with IMAC washing buffer as follows. 0.5 mL of Ni-NTA resin (about 1 mL slurry) is resuspended in 5 mL IMAC washing buffer and spun at 750 rpm for 5 min. The supernatant is discarded and the resin is again resuspended in 5 mL IMAC washing buffer and spun

at 750 rpm for 5 min. The buffer is discarded and the resin is resuspended in 1 mL of IMAC washing buffer.

7. IMAC elution buffer: 50 mM Tris–HCl, pH 8, 300 mM NaCl, 250 mM imidazole.
8. Empty poly-prep chromatography columns from BioRad.
9. Reducing HEPES buffer contains 50 mM 4-(2-hydroxyethyl)-1-piperazineethane-sulfonic acid (HEPES) and 1 mM DTT, pH 7.4.
10. 96-well dot-blot system (Minifold I, Schleicher & Schuell Biosciences).
11. PBS/BSA: 4% (w/v) BSA in PBS, pH 7.4.
12. Preblocked antidigoxigenin-HRP antibody conjugate (Roche Applied Science): diluted 1:1,000 in PBS/BSA described in 2.4.11 and slowly rotated on a wheel for 30 min at room temperature.
13. Chemiluminescence imager (Image Station 440, Kodak).

3. Methods

AGT is a DNA repair protein that transfers irreversibly alkyl groups from O^6-alkylated guanine in DNA to its active cysteine. The enzyme has previously been exploited by our group as a tag that can be covalently labeled with chemical probes in living cells (Fig. 1a). Preferred substrates of AGT are O^6-benzylguanine or derivatives thereof that carry probes at the 4-position of the benzyl ring. Reaction of AGT with these substrates leads to a specific transfer of the probes to the enzyme.

We recently attempted to reprogram the substrate specificity of AGT to create new protein tags with orthogonal substrate specificities for the labeling of different AGT fusion proteins with different probes. Orthogonal AGT mutants with excellent activities towards the nonnatural substrate O^2-benzylcytosine (BC) could be generated using saturation and random mutagenesis approaches (23). In contrast, mutants that react with another nonnatural substrate, O^6-propargylguanine (PG; Fig. 1b), could not be generated using these traditional mutagenesis techniques. Consequently, we followed an approach that is based on peptide loop insertion to evolve mutants that react with PG. We speculated that additional amino acid sequences on the surface of the protein close to the active site could provide the enzyme with new substrate-binding elements. The protocols that led to a semiorthogonal AGT (although the AGT mutant reacts with PG it retains its activity towards BG substrates) are described in the following sections (see Notes 1 and 2).

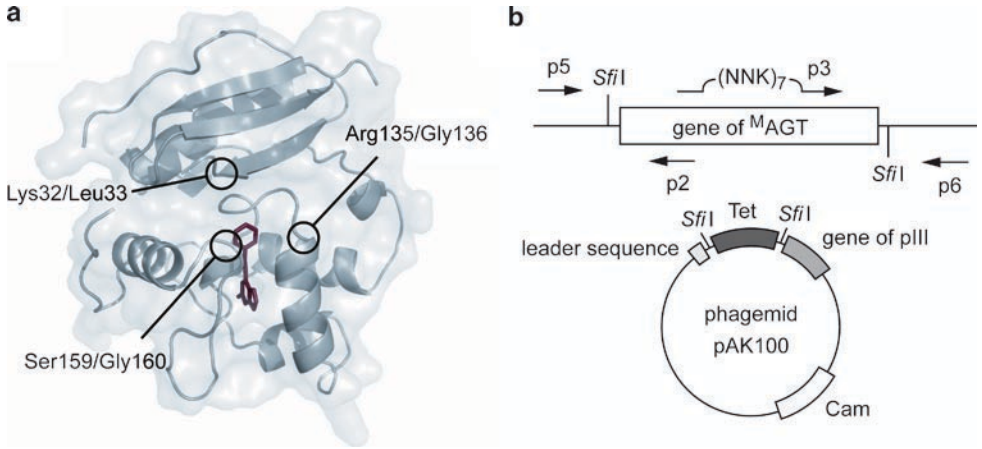


Fig. 2. Insertion of peptide loops into AGT. (**a**) Three potential sites in the polypeptide chain of AGT are chosen for the insertion of random peptide loops. The sites are *encircled* and the flanking amino acid positions are indicated. (**b**) Genes encoding MAGT mutants with five additional random amino acids and randomized amino acids at the sites of the loop insertion are generated by overlap extension PCR. The genes are ligated into a phagemid vector (pAK100) using the two *Sfi*I sites for the display of AGT loop insertion mutants on phage.

3.1. Identification of Suitable Sites for the Insertion of Peptide Loops

1. Three positions in the polypeptide chain of AGT are tested for their tolerance for peptide loop insertions. Towards this end, random peptide loops of five amino acids are inserted at these potential insertion sites and the functionality of 20 mutants (per insertion site) is tested. Insertion sites are chosen that are at the surface of AGT and in close proximity to the active site of the enzyme. The following positions are tested for the tolerance of peptide loop insertions (also shown in Fig. 2a): The turn between the short β-strand 7 and α-helix 8 (insertion between amino acid Glu159 and Gly160), the terminus of helix 6 (insertion between amino acid Ser135 and Gly136), and the terminus of the third β-strand (insertion between amino acid Lys32 and Leu33). The mutants with the additional peptide loops are also randomized at the junctions of the insertion sites because the amino acids at these positions may require a different conformation to accommodate the inserted peptides.
2. Genes of MAGT (MAGT is a mutant of human AGT with improved properties) with insertions of 15 nucleotides (the codons for five additional random amino acids) are generated by overlap extension PCR. The cloning strategy for the generation of mutants with loop insertions between Glu159 and Gly160 is described in this protocol as an example. The DNA encoding the amino acids 1–158 of MAGT is amplified with the two DNA primers p1 and p2. The primer p1 adds a *BamH*I restriction site to one end of the PCR product. The DNA encoding the amino acids 161–182 are amplified in a separate PCR using the primers p3 and p4.

The DNA primer p3 adds seven random codons (randomized as NNK; N represents an adenine, thymine, guanine, or cytosine nucleotide, and K represents a thymine or a guanine nucleotide) for the five additional amino acids, the amino acids in the positions 159 and 160 and an appendage of 20 nucleotides to the PCR product. The 20 nucleotide appendage overlaps with the product of the first PCR and is used for the assembly of the two DNA fragments. The primer p4 adds nucleotides encoding a polyhistidine sequence and an *EcoR*I site at one end of the PCR product. 50 μL reactions are prepared using standard PCR conditions (0.1 μg pAK100 vector containing the gene of ^{M}AGT, PCR reaction buffer, 50 μM dNTP, 0.5 μM primer, and 2 units Taq polymerase; PCR time and temperature: 30 s 95°C, 30 s 55°C, 1 min 72°C, 30 cycles, additional elongation in the last cycle for 3 min at 72°C). The PCR products are separated on a 1% agarose gel. The PCR products with the correct size are cut out with a razor blade and the DNA is extracted.

3. The two PCR products are assembled in an overlap extension PCR reaction in a 50 μL volume. The two PCR products (2 μg of each) are first assembled in a reaction without DNA primers (PCR buffer, 50 μM dNTP, and 2 units Taq polymerase; PCR time and temperature: 1 min 95°C, 1 min 55°C, 3 min 72°C, 5 cycles). The primers p1 and p4 are then added to the reaction (to a final concentration of 0.5 μM) and the assembled product is PCR-amplified (30 s 95°C, 30 s 55°C, 1 min 72°C, 30 cycles, additional elongation in the last cycle for 3 min at 72°C). The PCR products are separated on a 1% agarose gel. The PCR products with the correct size (about 580 base pairs) are cut out with a razor blade and the DNA is extracted.
4. The genes are ligated into the vector pGEX-2T for the expression of GST-AGT fusion proteins. The assembled DNA (5 μg) and the vector pGEX-2T (20 μg) are cleaved with *BamH*I and *EcoR*I in 50 and 100 μL reactions containing digestion buffer and the two enzymes (2 units of enzyme per 1 μg of DNA). The digested PCR product is purified with an extraction kit. The products of the vector cleavage reaction are separated on a 0.8% agarose gel, and the vector is cut out of the gel and extracted. 0.5 μg of the PCR product and 2 μg of the vector are ligated overnight at 16°C in a 50 μL reaction containing ligation buffer and T4 ligase (100 units). The ligation product is purified with a purification kit.
5. 0.5 μg of the ligated plasmid is electroporated into electrocompetent *E. coli* XL1-Blue cells (50 μL). The electroporated cells are mixed with 2YT media (1 mL), plated (50 μL) in different dilutions (no dilution, 1:10, 1:100, 1:1,000) on agar

plates containing 100 μg/mL ampicillin and incubated at 37°C overnight.

6. Individual colonies are picked and grown overnight in 3 mL 2YT media containing ampicillin (100 μg/mL). Cells from these cultures are diluted in 20 mL of SB medium containing ampicillin (100 μg/mL) to an OD at 600 nm of 0.1 and grown at 37°C. Protein expression is induced when the OD_{600} reaches 0.6 by addition of 1 mM IPTG. After expression at 24°C for 4 h, the cells are harvested by centrifugation at 5,000 rpm for 30 min at 4°C.
7. The cells are resuspended in lysis buffer (0.5 mL) and sonicated (50 W, 20 kHz) for 10 s. The cell debris is pelleted by centrifugation for 30 min at 13,000 rpm at 4°C. The soluble GST-AGT fusion protein in the supernatant is reduced by addition of 5 μL of 100 mM DTT (to a final concentration of 1 mM). A quantity of 5 μL of BGFL (20 μM) is added to 45 μL of the reduced fusion protein (to obtain a final BGFL concentration of 2 μM). The reactions are incubated for 30 min at room temperature and quenched by addition of BG (1 μL of 5 mM BG to obtain a final concentration of 100 μM).
8. The reaction products are separated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE). A 10% tris-glycine gel is installed in a vertical gel apparatus and 500 mL running buffer is added. Loading buffer (10 μL) is added to the reactions (20 μL) and the samples are heated at 70°C for 10 min. The cooled samples are loaded into the wells of the gel and electrophoresed for 60 min at 100 V and 50 mA.
9. The fluorescein-labeled protein in the polyacrylamide gel is visualized on an imager by excitation of fluorescein with UV light and detection of the emission above 500 nm. Expressed AGT mutants that are functional are labeled with the fluorescent probe and become visible at a molecular weight that corresponds to 48 kDa. The more mutants are labeled, the more the site in AGT is permissive to loop insertions. The position Ser159/Gly160 in AGT turned out to be most suitable for the insertion of peptide loops.

3.2. Cloning of Peptide Loop Insertion Libraries and Display on Phage

1. DNA fragments that encode loop insertion mutants of ^{M}AGT are prepared using overlap extension PCR (the cloning strategy is depicted in Fig. 2b). The 15 additional nucleotides are inserted between Glu159 and Gly160 since this position is most permissive to peptide loop insertions. The DNA encoding the amino acids 1–158 of ^{M}AGT (and a stretch of DNA upstream of the AGT gene that also contains a *Sfi*I restriction site) is amplified with the two DNA primers p5 and p2. In a separate PCR, the DNA encoding the amino acids 161–182 (and a stretch of DNA downstream of the

AGT gene that also contains a *Sfi*I restriction site) is amplified using the primers p3 and p6. The mutagenesis primer p3 adds seven random codons (NNK) and an appendage of 20 nucleotides to the PCR product. The 21 random nucleotides encode the amino acids at positions 159, 160, and of the inserted peptide. The appendage of 20 nucleotides is required for the assembly of the two DNA fragments in an overlap extension PCR. Standard 50 μL PCR conditions are used for the two reactions (0.1 μg plasmid pAK100 containing the gene of ^{M}AGT, PCR reaction buffer, 50 μM dNTP, 0.5 μM primer, and 2 units Taq polymerase; reaction time and temperature: 30 s 95°C, 30 s 55°C, 1 min 72°C, 30 cycles, additional elongation in the last cycle for 3 min at 72°C). The PCR products are separated on a 1% agarose gel. The PCR products with the correct size are cut out with a razor blade and the DNA is extracted.

2. The two PCR products are assembled in an overlap extension PCR reaction. The two PCR products (1 μg of each) are first assembled in a 50 μL reaction without DNA primers (PCR reaction buffer, 50 μM dNTP, and 2 units Taq polymerase; reaction time and temperature: 1 min 95°C, 1 min 55°C, 3 min 72°C, 5 cycles). The primers p5 and p6 are then added to the reaction (0.5 μM final concentration) and the assembled product is PCR amplified (30 s 95°C, 30 s 55°C, 1 min 72°C, 30 cycles, additional elongation in the last cycle for 3 min at 72°C). To obtain sufficient PCR product (around 40 μg of DNA), the assembly reaction is performed in multiple (typically 5–10) 50 μL reactions. The PCR products are separated on a 1% agarose gel. The DNA band with the correct size is cut out with a razor blade and extracted. Typically, four columns of the kit are required to obtain 40 μg of DNA.
3. The genes are ligated into the phagemid vector pAK100 for the display of the AGT mutants on phage. The assembled DNA (40 μg) and the vector pAK100 (200 μg) are cleaved with *Sfi*I (2 units per 1 μg of DNA) by incubation in 0.1 and 1 mL reaction buffer for 2 h at 50°C. The digested PCR product is purified on four columns of an extraction kit. The correct DNA band is cut out of the gel with a razor blade and the DNA is extracted with a gel extraction kit using ten columns. The digested PCR product (20 μg) and vector (80 μg) are ligated overnight at 16°C in 1 mL of ligation buffer containing T4 ligase (1,000 units). The ligation product is purified on ten columns of a purification kit.
4. The ligated vector (80 μg) is electroporated into competent *E. coli* JM101 cells. For the preparation of electrocompetent cells, a 1 L culture of 2YT is inoculated with bacteria from an

overnight culture (1:100 dilution) and grown at 30°C until the culture has an optical density of 0.5–0.7 at 600 nm. The culture is chilled at 4°C for 2 h and the cells are washed three times by repeated centrifugation (15 min in a prechilled rotor at 4,000 rpm) and resuspended in buffer (1 L of ice cold H_2O with 1 mM HEPES, pH 7, 10% glycerol in the first round, 500 mL of ice cold H_2O with 1 mM HEPES, pH 7, 10% glycerol in the second round and 50 mL of H_2O with 1 mM HEPES, pH 7, 10% glycerol in the third round). The cells are pelleted by centrifugation (15 min, 4,000 rpm, 4°C) and resuspended in 2 mL of ice cold 10% glycerol. The DNA is mixed with the cells and transferred into 20 prechilled electroporation cuvettes (2 mm; 100 μL per cuvette). The cells are electroporated at 2.5 kV, 25 μF, and 200 Ω and mixed immediately with 2YT media (3 mL per electroporation; a total of 60 mL), incubated for 1 h (37°C, 200 rpm), plated on 15 large (20 cm diameter) 2YT agar plates containing 25 μg/mL chloramphenicol, and incubated overnight at 37°C. A small sample of the electroporated cells is plated at various dilutions (1:1,000, $1:10^4$, $1:10^5$ and $1:10^6$) on small agar plates (2YT with 25 μg/mL chloramphenicol) to estimate the number of transformants. Electroporation of 100 μg of ligated vector yields typically 10^8–10^9 colonies.

5. The bacterial colonies are resuspended with 2YT medium containing 15% glycerol (5–10 mL per plate), pipetted into 1.5 mL tubes, snap frozen in liquid nitrogen and stored at –80°C.

6. Phages are produced as follows (see also Notes 3 and 4): A 50 mL culture of 2YT containing 25 μg/mL chloramphenicol is inoculated with cells from the glycerol stock. Sufficient cells are added to the culture to obtain an optical density of 0.1 at 600 nm and hence to ensure the representation of the whole library diversity. The cultures are incubated in a shaking incubator (200 rpm) at 37°C until the OD at 600 nm reaches 0.4. The cells are infected by incubation (nonshaking) with helper phage M13KO7 (a tenfold excess of helper phage over bacterial cells) for 30 min at 37°C. The cells are then pelleted by centrifugation at 4,000 rpm for 10 min and resuspension in 2YT containing 25 μg/mL chloramphenicol and 70 μg/mL kanamycin. The culture is incubated at 30°C in an incubator shaker (200 rpm) for 16 h. The phages are separated from cells by centrifugation at 4,000 rpm for 20 min at 4°C. Phages can be stored at 4°C but it is recommended to prepare the phages freshly before they are used for the selections to minimize proteolysis of the displayed AGT.

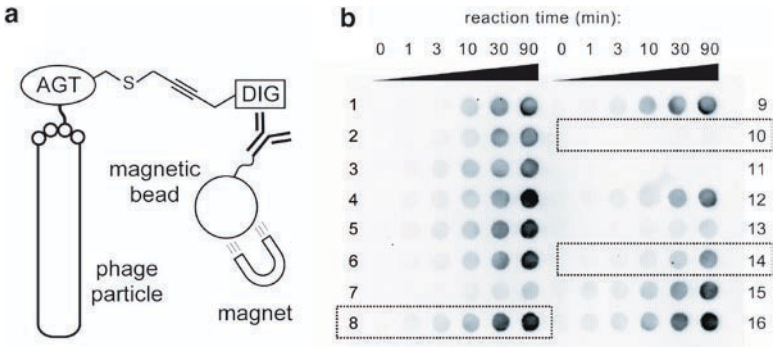


Fig. 3. (**a**) Schematic depiction of the strategy for the isolation of active AGT mutants by phage display. AGT mutants on phage that have reacted with PGDIG are captured with antidigoxigenin antibodies that are immobilized on magnetic beads. (**b**) Activity of AGT mutants determined with a dot-blot-based assay. Affinity purified AGT mutants are incubated with PGDIG for different time periods and blotted onto a hydrophobic membrane. The amount of reacted PGDIG is quantified with an antidigoxigenin-HRP conjugate and a chemiluminescence-based assay. Some mutants react faster (e.g., clone 8) compared to others (e.g., clone 14) and some are completely inactive (e.g., clone 10).

3.3. Phage Selection to Identify Loop Insertion Variants with Desired Activities

1. Phages that react with the PG (O^6-propargylguanine) substrate are isolated according to the principle outlined in Fig. 3a. The supernatant (0.5 mL) containing the phage particles (typically 10^{11} t.u./mL; t.u. are transforming units) are blocked by addition of 0.5 mL PBS/milk. The mixture is slowly inverted for 30 min at room temperature on a rotating wheel.
2. The substrate PGDIG (5 μL of 2 mM in DMSO) is added to the blocked phage to obtain a final concentration of 20 μM PGDIG. The reaction tube is inverted for 10 min on a rotating wheel. As a positive and negative control of the phage panning procedure, blocked phage particles of the same preparation are incubated in parallel either with BGDIG (20 μM final concentration) or without any substrate.
3. Nonreacted substrate is removed as follows: phages are mixed with 250 μL of PEG solution and incubated on ice for 10 min for precipitation in 1.5 mL tubes. The tubes are spun at 13,000 rpm for 10 min at 4°C and the supernatant is removed by aspiration without touching the phage (a white, hardly visible pellet). The phages are resuspended in PBS and precipitated again with PEG solution as before. After the second precipitation, the phages are resuspended in 0.5 mL PBS/milk.
4. The phages are added to preblocked antidigoxigenin magnetic beads and incubated on a rotating wheel at room temperature for 30 min. The magnetic beads are then washed once with PBS/milk, six times with PBS/Tween-20, and twice with PBS. The reaction tube is replaced at least once during the washing procedure. The phages are eluted by incubation in

100 μL acidic elution buffer for 5 min. The magnetic beads are drawn to the side of the tube with a magnet and the phages in the solution are transferred to a new tube. A quantity of 50 μL of neutralization buffer is added to the eluate to neutralize the pH.

5. Eluted phages (120 μL) are incubated with 10 mL of exponentially growing (OD at 600 nm = 0.4) JM101 cells for 30 min at 37°C (the cells must not be shaken during the infection), plated on large (20 cm diameter) agar plates containing chloramphenicol (25 μg/mL) and incubated at 37°C overnight. For the determination of the number of selected phages, 20 μL of the eluted phage (and various dilutions ranging from 10- to 10^7-fold) are incubated with 180 μL of exponentially growing JM101 cells for 30 min at 37°C. 20 μL of the infected cells are spotted on chloramphenicol-containing agar plates. Upon pipetting of the dilutions onto the plates, the media is soaked into the agar and the plates are incubated at 37°C overnight. The number of bacterial colonies on the plates multiplied with the dilution factor represents roughly the number of selected phage particles (assuming that all phage particles can infect one cell). It is expected that around 10^5–10^6 phages are pulled out through nonspecific binding to the beads in each selection (also in the negative control where no substrate is added to the phage).
6. The bacterial colonies are resuspended from the plates and new phage particles are produced. The phages are subjected to 2–3 further rounds of phage selection. A good indication for the enrichment of clones with improved catalytic activity is a phage titre that is elevated by a factor of 10–10,000 compared to the titre of the negative control (a selection where the PGDIG substrate is omitted). Enrichment is generally observed in the second or third round of phage selection (see Note 5).

3.4. Characterization of Isolated Loop Insertion Mutants

1. After the last selection round, bacterial cells are infected with the phage and plated on chloramphenicol-containing plates and incubated overnight at 37°C as described under Subheading 3.3. The bacterial cells of the colonies are resuspended in PBS, pooled, and the plasmid DNA is extracted from cells.
2. The genes of the AGT mutants are PCR-amplified with the primers p1 and p4 to add *BamH*I and *Eco*RI restriction sites and a C-terminal polyhistidine tag. Standard conditions are used for the 50 μL PCR reaction (0.5 μg plasmid of the extracted plasmid DNA, PCR buffer, 50 μM dNTP, 0.5 μM primers, and 2 units Taq polymerase; reaction time and temperature: 30 s 95°C, 30 s 55°C, 1 min 72°C, 30 cycles,

additional elongation in the last cycle for 3 min at 72°C). The PCR products are separated on a 1% agarose gel and the DNA band with the expected molecular size (about 580 base pairs) is cut out. The DNA is extracted.

3. For the expression of GST-AGT fusion proteins, the PCR product is ligated into the vector pGEX-2T. The purified PCR product (5 μg) and the vector pGEX-2T (20 μg) are cleaved with *BamH*I and *EcoR*I in 50 and 100 μL, respectively, reactions containing digestion buffer and enzyme (2 units of enzyme per 1 μg of DNA). The digested PCR product is purified. The products of the vector cleavage reaction are separated on a 0.8% agarose gel, the vector is cut out of the gel and extracted with a gel extraction kit (two columns are used). The digested PCR product (0.5 μg) and vector (2 μg) are ligated overnight at 16°C in a 50 μL reaction containing ligation buffer and T4 ligase (100 units). The ligation product is purified.
4. The ligation (0.5 μg of DNA) is electroporated into electrocompetent *E. coli* XL1-Blue cells (50 μL). The electroporated cells are mixed with 2YT medium (5 mL), plated in different dilutions (no dilution, 1:10, 1:100, 1:1,000) on agar plates containing 100 μg/mL ampicillin, and incubated at 37°C overnight.
5. Cells of single colonies are picked and grown overnight in cultures of 3 mL 2YT media containing ampicillin (100 μg/mL). Cells of these cultures are diluted in 100 mL of SB media containing ampicillin (100 μg/mL) to an OD of 0.1 at 600 nm and grown at 37°C. Protein expression is induced by addition of 1 mM IPTG when the OD at 600 nm reaches 0.6. After the expression at 24°C for 4 h, the cells are harvested by centrifugation at 5,000 rpm for 30 min at 4°C.
6. The cell pellet is resuspended in 10 mL IMAC lysis buffer and sonicated (50 W, 20 kHz) for 1 min on ice. The cell debris is pelleted by centrifugation for 30 min at 13,000 rpm at 4°C and the supernatant is transferred to a new tube.
7. 0.5 mL Ni-NTA slurry is spun for 1 min at 500 rpm and resuspended in 2 mL IMAC washing buffer. This process is repeated twice, the slurry is added to the cell lysate, and the mixture is slowly inverted on a rotating wheel for 30 min at 4°C and then pipetted into a column. The flow-through is discarded and the column is washed three times with 1 mL IMAC washing buffer. The fusion protein is eluted with 2× 0.5 mL IMAC elution buffer.
8. The eluted protein is dialyzed overnight at 4°C in reducing HEPES buffer using a membrane with a molecular-weight cut-off of 5,000 Da and the protein concentration is determined by measuring the optical density at 280 nm.

9. AGT mutants are diluted to 1 μM in reducing HEPES buffer. 1 μL of PGDIG (2 mM) is pipetted to 99 μL of AGT in six replicates per AGT mutant. The reactions are quenched by addition of 1 μL of BG (10 mM) after different time periods (0, 1, 3, 10, 30, or 90 min).
10. The reaction mixtures are spotted onto a hydrophobic membrane (PVDF) with a 96-well dot-blot system. The membrane is blocked in PBS/milk, rinsed with PBS, blocked in PBS/BSA, and incubated with preblocked antidigoxigenin-HRP antibody conjugate for 30 min at room temperature. The membrane is washed by sequentially incubating in three containers with PBS on a rocking platform. The immobilized HRP is detected by addition of ECL reagent (GE Healthcare Life Sciences) to the filter for 5 min and measuring the chemiluminescence with a detector (Kodak Imagestation 440; Fig. 3b).

4. Notes

1. The protocols can easily be adapted for the evolution of proteins with new binding specificities. It is recommended to insert peptide loops at multiple sites (two to three) to obtain protein binders with good affinities. Tagged antigen can be used instead of the PG substrate (e.g., biotinylated antigen in combination with streptavidin-coated magnetic beads) in the phage selection procedures.
2. The selection of AGT by phage display is particularly suited because the enzyme remains covalently linked to the tagged moiety of the substrate after the reaction and catalytically active mutants can be captured. For the evolution of other enzymes, another selection or screening strategy may be applied.
3. Phages are highly stable and can resist washing with detergents. It is recommended that nondisposable glass- and plastic-ware is soaked for 1 h in 2% (v/v) hypochlorite before it is washed using standard detergents.
4. To prevent contaminations with phage, precautions may be taken such as the use of disposable gloves and filter tips. Regularly clean incubator cabinets, incubator shakers, and water baths.
5. Sequencing of a few clones after each round of selection can give valuable information about the selection process. A large number of mutants with identical or similar loop sequences can be an indication for the enrichment of a superior phenotype.

References

1. Lu Y, Berry SM, Pfister TD (2001) Engineering novel metalloproteins: design of metal-binding sites into native protein scaffolds. Chem Rev 101(10):3047–3080
2. Nicaise M, Valerio-Lepiniec M, Minard P, Desmadril M (2004) Affinity transfer by CDR grafting on a nonimmunoglobulin scaffold. Protein Sci 13(7):1882–1891
3. Martin L, Stricher F, Misse D et al (2003) Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. Nat Biotechnol 21(1): 71–76
4. Batori V, Koide A, Koide S (2002) Exploring the potential of the monobody scaffold: effects of loop elongation on the stability of a fibronectin type III domain. Protein Eng 15(12):1015–1020
5. Koide A, Koide S (2007) Monobodies: antibody mimics based on the scaffold of the fibronectin type III domain. Methods Mol Biol 352:95–109
6. Koide A, Bailey CW, Huang X, Koide S (1998) The fibronectin type III domain as a scaffold for novel binding proteins. J Mol Biol 284(4):1141–1151
7. Schlehuber S, Skerra A (2005) Lipocalins in drug discovery: from natural ligand-binding proteins to "anticalins". Drug Discov Today 10(1):23–33
8. Christmann A, Walter K, Wentzel A, Kratzner R, Kolmar H (1999) The cystine knot of a squash-type protease inhibitor as a structural scaffold for Escherichia coli cell surface display of conformationally constrained peptides. Protein Eng 12(9):797–806
9. Nuttall SD, Rousch MJ, Irving RA, Hufton SE, Hoogenboom HR, Hudson PJ (1999) Design and expression of soluble CTLA-4 variable domain as a scaffold for the display of functional polypeptides. Proteins 36(2): 217–227
10. McConnell SJ, Hoess RH (1995) Tendamistat as a scaffold for conformationally constrained phage peptide libraries. J Mol Biol 250(4): 460–470
11. Mathonet P, Deherve J, Soumillion P, Fastrez J (2006) Active TEM-1 beta-lactamase mutants with random peptides inserted in three contiguous surface loops. Protein Sci 15(10):2323–2334
12. Legendre D, Soumillion P, Fastrez J (1999) Engineering a regulatable enzyme for homogeneous immunoassays. Nat Biotechnol 17(1):67–72
13. Legendre D, Vucic B, Hougardy V et al (2002) TEM-1 beta-lactamase as a scaffold for protein recognition and assay. Protein Sci 11(6):1506–1518
14. Hayes F, Hallet B (2000) Pentapeptide scanning mutagenesis: encouraging old proteins to execute unusual tricks. Trends Microbiol 8(12):571–577
15. Hayes F, Hallet B, Cao Y (1997) Insertion mutagenesis as a tool in the modification of protein function. Extended substrate specificity conferred by pentapeptide insertions in the omega-loop of TEM-1 beta-lactamase. J Biol Chem 272(46):28833–28836
16. Heinis C, Schmitt S, Kindermann M, Godin G, Johnsson K (2006) Evolving the substrate specificity of O6-alkylguanine-DNA alkyltransferase through loop insertion for applications in molecular imaging. ACS Chem Biol 1(9): 575–584
17. Hallet B, Sherratt DJ, Hayes F (1997) Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein. Nucleic Acids Res 25(9): 1866–1867
18. Handa P, Thanedar S, Varshney U (2002) Rapid and reliable site-directed mutagenesis using Kunkel's approach. Methods Mol Biol 182:1–6
19. Aiyar A, Xiang Y, Leis J (1996) Site-directed mutagenesis using overlap extension PCR. Methods Mol Biol 57:177–191
20. Juillerat A, Heinis C, Sielaff I et al (2005) Engineering substrate specificity of O6-alkylguanine-DNA alkyltransferase for specific protein labeling in living cells. Chembiochem 6(7):1263–1269
21. Keppler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H, Johnsson K (2003) A general method for the covalent labeling of fusion proteins with small molecules in vivo. Nat Biotechnol 21(1):86–89
22. Krebber A, Bornhauser S, Burmester J et al (1997) Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. J Immunol Methods 201(1):35–55
23. Gautier A, Juillerat A, Heinis C et al (2008) An engineered protein tag for multiprotein labeling in living cells. Chem Biol 15(2): 128–136

Chapter 17

Massive Mutagenesis®: High-Throughput Combinatorial Site-Directed Mutagenesis

Julien Sylvestre

Abstract

Massive Mutagenesis® is a proprietary library creation method that enables the fast generation of high-quality genetic libraries. Starting from a single gene on a plasmid and hundreds to thousands of oligonucleotides, a one-step single-strand circular amplification method creates random combinations of several site-directed substitutions, insertions, or deletions. Libraries of up to a billion such variants have been routinely generated. Sequencing those variants demonstrated lower biases than alternative approaches such as error-prone PCR. Screening and selecting them has yielded improved biocatalysts, therapeutic proteins, and antibodies.

Key words: Protein engineering, Directed evolution, Site-directed mutagenesis, Combinatorial mutagenesis, High-throughput

1. Introduction

Thanks to its simplicity, error-prone PCR is still widely used to create artificial genetic libraries. This technique, however, suffers from a number of biases. The most severe bias stems from the low probability of having 2 consecutive nucleotides mutated and implies that from a given starting codon, only six different amino acids are accessible (on average). Another bias, the "amplification bias," is caused by the (roughly) exponential accumulation of products during PCR, which favors the accumulation of mutations integrated during the first rounds. A third bias results from unbalanced probabilities of transitions (between purines or between pyrimidines; favored) and transversions (other base exchanges; unfavored) substitutions. In the face of these limitations of error-prone PCR, and aiming at improving

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_17,

the first "library creation" step of most directed evolution protein engineering experiments, Biométhodes has developed a proprietary high-throughput mutagenesis technique termed Massive Mutagenesis®. Massive Mutagenesis® has been used successfully in several industrial projects to improve various properties of therapeutic proteins, antibodies, and industrial enzymes. Compared to random approaches, Massive Mutagenesis® enhances the directed evolution process by opening access to all 19 codons (saturation), thereby increasing library quality. It is often a profitable strategy to design large custom libraries containing "smart" combinations of mutations based on the understanding of the chemistry of the underlying protein.

The Massive Mutagenesis® technology is fully disclosed in (1, 2). A variant that makes use of chip-eluted oligonucleotide mixtures is described in (3). Below is a description of the reference protocol as well as some useful variations. The main steps of a standard Massive Mutagenesis procedure and the molecules used are represented schematically in Fig. 1.

2. Materials

2.1. Plasmid

Dilute a plasmid with a sequence of interest that will be expressed and an antibiotic selection gene prepared from a standard (methylation positive) bacterial culture to a 200 ng/µL concentration.

2.2. Oligonucleotides Phosphorylation

In a volume of 20 µl, add a total of 1 nmol oligonucleotide mixture, 2 µl polynucleotide kinase buffer (New England Biolabs), 2 µl of 10 mM ATP (Sigma Aldrich), and 1 µl 10 U/µl T4 polynucleotide kinase (New England Biolabs). Incubate for 1 h at 37°C (see Note 1).

3. Amplification Mixture

Prepare the following mixture to obtain a final volume of 25 µL: 1 µL plasmid, 200 pmol phosphorylated oligonucleotide mixture, 1 µL 25 mM deoxyribonucleotide phosphate (dNTP, Sigma), 1 µl 100 mM $MgSO_4$ (Sigma), 0.2 µl 1 M dithiothreitol (DTT, Sigma), 2.5 µL 10× Pfu Pol buffer (Promega), 1 µL 10 mM adenosine triphosphate (ATP, Sigma), 0.2 µl 100 mM nicotinamide adenine dinucleotide (NAD, Sigma), 0.8 µL Pfu DNA polymerase (10 U/µL, Promega), and 0.8 µl Tth ligase (10 U/µl; ABgene; see Note 2).

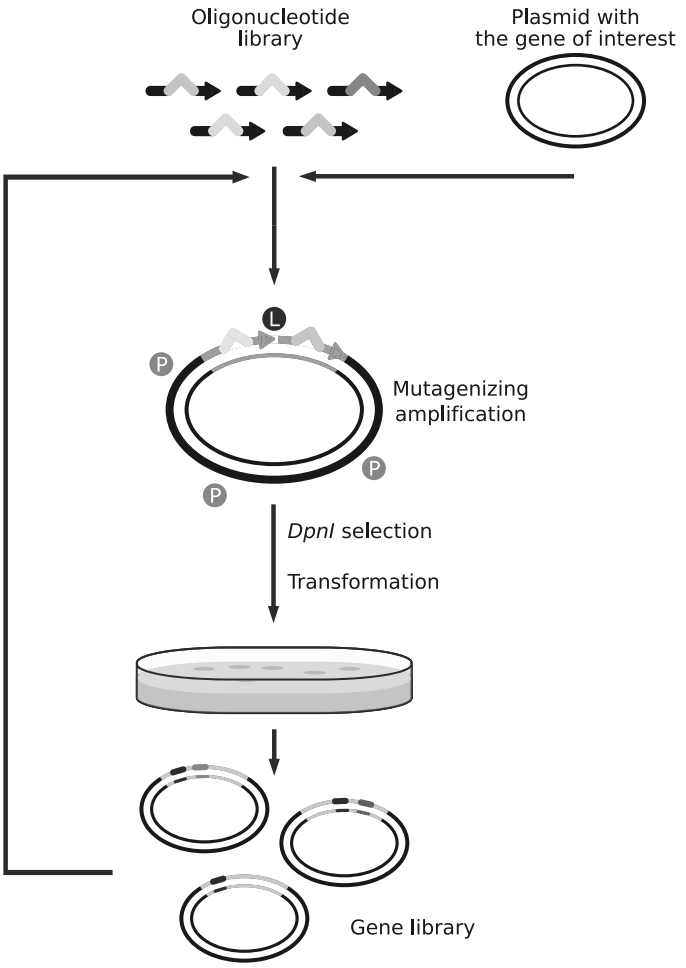


Fig. 1. Massive Mutagenesis® protocol. A single strand of a plasmid harbouring the gene of interest is amplified using a library of phosphorylated mutagenizing oligonucleotides, a thermostable polymerase (P), and a thermostable ligase (L). Oligonucleotide-encoded point mutations are incorporated in a combinatorial way in the amplification product. This in vitro synthesized product is then enriched by counter-selecting, in vivo, methylated parental molecules using DpnI. Transformation into cells creates a large, high-quality library of genes. The average number of mutations per molecule can be increased by scraping the clones containing the library, preparing their plasmid DNA, and submitting it to an additional round of Massive Mutagenesis®.

3.1. DpnI Digestion

To the 25 µL amplification product, add 4 µL 10× Buffer 4 (New England Biolabs), 0.5 µL *DpnI* (20 U/µL; New England Biolabs) and 15.5 µL deionized water and incubate for 30 min at 37°C (4).

3.2. Transformation

In order to desalt the digested product before electrotransformation, dialyze it on an MF hydrophilic membrane filter (0.025 µm pores; Millipore) against deionized water for 15 min.

Electroporate dialyzed DNA into 40 µL electrocompetent DH10B cells (Invitrogen), and then plate cells on LB Amp Petri dish (adapt antibiotic to the plasmid used; see Note 3).

4. Methods

4.1. Preparation of Plasmid and Oligonucleotides

Prepare a plasmid (up to 12 kb in length) and, based on the plasmid sequence, design 33-mer oligonucleotides that are all complementary to the same plasmid strand except in their center where they each incorporate one different desired point mutation (see Note 4). Said mutations can be a single-codon substitution, insertion, or deletion (see Note 5).

Prepare a library containing at least five oligonucleotides. Libraries with hundreds of different oligonucleotides have been used successfully. For saturation experiments, degenerate (NNN) or semi-degenerate (e.g., NNK) oligonucleotides can be used, though this impacts negatively on library quality due to genetic code degeneracy and the introduction of stop codons. Preferred alternatives include using equimolar mixtures of oligonucleotides synthesized from trinucleotide phosphoramidites (5) or eluted from cleavable DNA chips (6, 7) (see also Note 6).

Phosphorylate the oligonucleotides as a pool.

4.2. Amplification Reaction

Subject the amplification mixture containing the plasmid and oligonucleotides to the following thermal cycling: 12× (denaturation at 94°C for 1 min; annealing at 50°C for 2 min (adapt to the melting temperature of oligonucleotides); extension at 68°C for 20 min (use approximately 2 min per kb)).

4.3. Selection of Mutated Molecules

To counter-select methylated parent strands in favor of in vitro neosynthesized mutant molecules, the amplification product is digested by the methylation-dependent restriction enzyme *DpnI* (linearized molecules will not transform efficiently into cells at the next step).

4.4. Transformation and Screening or Selection

Desalt the library by dialysis and transform it into electrocompetent cells following the manufacturer's instructions (Invitrogen). Grow cells overnight. Individual mutant clones can then be sequenced or screened. Affinity-based selection methods include ribosome display or phage display techniques. Screening for catalytic activity typically involves fluorescent (or colored) substrates or products. Usually one round of Massive Mutagenesis® yields libraries with an average of 1–2 mutations per molecule; this should be checked by quality-control experiments using restriction analysis and sequencing. The mutation rate can be tuned depending on the gene of interest and experimental goals (see Note 7).

Directed evolution experiments based on Massive Mutagenesis® libraries are described in (1–3, 6, 8, 9). Enzymes with higher catalytic activity or thermostability, therapeutic proteins with an improved in vivo half-life as well as antibodies with an adjusted selectivity have been obtained starting from large Massive Mutagenesis® libraries and using appropriate screening or selection methods.

5. Notes

1. Alternatively, oligonucleotides can be ordered directly in a (chemically) phosphorylated form.
2. We observed that this mixture of enzymes works well but other proofreading thermostable DNA polymerases, or mixtures thereof and other thermostable DNA ligases have also been effective. A 10:1 Pfu:Taq mixture has been used successfully.
3. Heat-shock transformation can be used; however, it reduces effective library size (approximately 10^6–10^7 total clones).
4. The advantages of Massive Mutagenesis® can be leveraged by computer-aided library design.
5. Mutations of longer (2–3 codons) regions have been obtained by using appropriate oligonucleotides yet this might require empirical fine-tuning of oligonucleotides length and sequence to avoid secondary structure formation and adjust Km.
6. In order to find optimal mutation conditions it is sometimes useful to carry out an experiment with a range of oligonucleotides concentrations, e.g., 0.16, 0.64, 2.56, and 10 (plasmid: oligonucleotides).
7. It has been observed repetitively that the number of mutations per plasmid in a given library typically follows a Poisson distribution whose mean can be adjusted by iterating several rounds of Massive Mutagenesis® as explained in (8). A balance has to be found which depends on factors such as the target protein resistance to mutations and the sensitivity, price, and throughput of the screening assay.

Acknowledgments

Massive Mutagenesis® is a patented technology of Biométhodes, an industrial biotechnology company based near Paris, France. Please contact Biométhodes regarding licensing conditions and partnership opportunities (www.biomethodes.com). The author thanks Ida Swarczewskaja for conceiving and drawing the figure.

References

1. Delcourt M, Blesa S (2000) Method for massive directed mutagenesis Patents EP1311670B1 and US7202086B2
2. Sylvestre J, et al (In Press) Modern biopharmaceuticals In: Joerg Knaeblein (ed) Recent success stories. Wiley and Sons, Weinheim
3. Sylvestre J, Delcourt M (2004) Method of massive mutagenesis. Patent applications EP1 544296 and US2005153343
4. Li F et al (2002) Site-directed mutagenesis facilitated by DpnI selection on hemimethylated DNA. Methods Mol Biol 182:19–27
5. Randolph J et al (2008) Codon-based mutagenesis using trimer phosphoramidites. Nucleic Acids Symp Ser 52:479
6. Saboulard D et al (2005) High-throughput mutagenesis using oligonucleotides synthesized on DNA chips. Biotechniques 39(3): 363–68
7. Richmond KE et al (2004) Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. Nucleic Acids Research 32(17):5011–5018
8. Laffly E et al (2008) Improvement of an antibody neutralizing the anthrax toxin by simultaneous mutagenesis of Its six hypervariable loops. J Mol Biol 378:1094–1103
9. Chautard H et al (2007) An activity-independent selection system of thermostable protein variants. Nat Methods 4:919–921

Chapter 18

Directed In Vitro Evolution of Reporter Genes Based on Semi-Rational Design and High-Throughput Screening

Ai-Sheng Xiong, Quan-Hong Yao, Ri-He Peng, and Zong-Ming Cheng

Abstract

Marker genes, such as *gusA*, *lacZ*, and *gfp*, have been applied comprehensively in biological studies. Directed in vitro evolution provides a powerful tool for modifying genes and for studying gene structure, expression, and function. Here, we describe a strategy for directed in vitro evolution of reporter genes based on semi-rational design and high-throughput screening. The protocol involves two processes of DNA shuffling and screening. The first DNA shuffling and screening process involves eight steps: (1) amplifying the target gene by PCR, (2) cutting the product into random fragments with DNase I, (3) purification of 50–100 bp fragments, (4) reassembly of the fragments in a primerless PCR, (5) amplification of the reassembled product by primer PCR, (6) cloning into expression vector, (7) transformation of *E. coli* by electroporation, and (8) screening the target mutants using a nitrocellulose filter. The second DNA shuffling and screening process also involves the same eight steps, except that degenerate oligonucleotide primers are based on the sequence of the selected mutant.

Key words: Directed evolution, DNA shuffling, Degenerate oligonucleotide, High-throughput screening, Semi-rational design

1. Introduction

Directed in vitro evolution, especially combined with DNA shuffling, is a powerful molecular tool in biological studies of protein structure and function, and consequently in the creation of new proteins, and thus, new phenotypes. In 1994, Stemmer introduced the method of DNA shuffling for the in vitro formation of recombinant genes from a set of parental genes (1, 2). Since then, various in vitro DNA shuffling methods have been developed and applied to evolving enzymes for basic research or for improving proteins of industrial significance, such as improved kinetics, altered substrate, or product specificities. Many useful

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_18,

enzymes and peptides have been isolated following artificial evolution. These include enzymes with biocatalytic and industrial significances (3–8), enzymes for plant improvement (9–11), as well as vaccines for pharmaceutical applications (12–15).

The routine shuffling procedure was first established by Stemmer in 1994 (1, 2), which involves the following steps: amplifying the target gene by PCR, cutting the PCR product into random fragments, gel purification of the small fragments, reassembly of the fragments in a PCR-like reaction without primers, amplification of the reassembled product by standard PCR, cloning into the expression vector, and selection of the target mutants. Many modified methods of in vitro directed evolution have been developed, such as DNA family shuffling (16), genome shuffling (17), error-prone RCA (rolling circle amplification) (18), SSDNAs (single-stranded DNA shuffling) (19), DOGS (degenerate oligonucleotide gene shuffling) (20), RACHITT (random chimeragenesis on transient templates) (21), COBARDE (codon-based random deletion) (22, 23), CSR (compartmentalized self-replication) (24), RNDM (random drift mutagenesis) (25), frame shuffling (26), MURA (mutagenic and unidirectional reassembly) (27), and shuffling of critical segments of the gene (28).

To date, it is uncertain whether it is more efficient to mutate an enzyme aimlessly or to mutate just the active sites or key sites (29). The conventional view is that changes of amino acids near the active sites are more likely to obtain the target mutant more easily (30). However, many DNA shuffling experiments have demonstrated that specific and random amino acid changes distant from the active site can affect substrate specificity. Parikh and Matsumura found that site-saturation mutagenesis was more efficient than DNA shuffling for directed evolution of β-fucosidase from β-galactosidase (31).

DNA shuffling does not need information about the protein structure or the protein structure–function relationship, but the method has drawbacks, including the requirement for reliable, simple, and fast screening methods. The rational evolutionary design-based method utilizes structure and sequence alignment information to create new genes and proteins. Rational evolutionary design-based information of amino acid sequence and protein structure using site-directed mutagenesis offers another strategy for in vitro directed evolution (32). A suitable selection strategy is also required for in vitro directed evolution. A mutant library, consisting of less than 50,000 mutants, was screened successfully by 96-well or 384-well microplates (33). However, if a mutant library consists of over 100,000 colonies, it is tedious, laborious, and costly to screen by using microplates.

Approximately fifty reporter genes have been used for transgenic and transplastomic research, and they have been reviewed for efficiency, biosafety, scientific applications, and commercialization (34).

However, only a few reporter genes have been widely used because many of these genes have specific limitations or have not been sufficiently tested (34). Many significant studies have been carried out to improve reporter genes by directed evolution in vitro because of their easy detection. Stemmer's group has obtained *E. coli* β-galactosidase variants with enhanced β-fucosidase activity (35). GFP mutants with a stronger fluorescence were selected by directed evolution (36–38). A GFP-like protein with a shift in fluorescence to a longer wavelength region was also created by DNA shuffling (39). Some *E. coli* β-glucuronidase variants were selected following DNA shuffling with increased β-galactosidase activity (40), thermostability (41–43), increased resistance to both glutaraldehyde and formaldehyde (44), and higher activity for the substrate *p*-nitrophenyl-β-d-xylopyranoside (33). β-glucuronidase and β-galactosidase, two of the most widely used reporter genes in molecular biological studies, provide excellent model systems to study directed evolution in vitro (34, 42).

Here, we describe a strategy for directed in vitro evolution of reporter genes (*hlacz-sh*, GenBank accession no. EF090269, 1,533 bp) based on semi-rational design and high-throughput screening. The protocol involves two processes of DNA shuffling and screening by fixing the variant colonies onto a nitrocellulose filter before treatment. Ten thousand individual colonies can be screened and selected on a 9-cm filter. A library containing 300,000 individuals can be screened on only thirty filter papers. The colonies in the filter paper are treated, selected, and mutants recovered. By increasing the size of a mutant library, we were able to effectively obtain desirable mutant colonies more efficiently.

2. Materials

2.1. Reagents and Equipments

1. DNA polymerase: *Taq* and 10×*Taq* buffer (Takara).
2. DNA marker: DL2000 (Takara).
3. dNTP mixture: 2.5 mM dATP, 2.5 mM dGTP, 2.5 mM dCTP, and 2.5 mM dTTP (Takara).
4. X-Gal: 5-Bromo-4-Chloro-3-Indolyl-β-D-Galactoside (Takara).
5. DNase I (RNase Free): (Takara).
6. Restriction endonucleases: *Bam* HI and *Sac* I (Takara).
7. T4 DNA ligase: (Takara).
8. Agarose: (Takara).
9. Tryptone: (OXOID).
10. Yeast extract: (OXOID).

11. BigDye® Terminator v3.1 Cycle Sequencing Kit: (Applied BioSystems).
12. Oligonucleotides: 30 μM (see Note 1).
13. Acrylamide: (Sigma, St Louis, MO, USA) (see Note 2).
14. Bis-acrylamide: (Sigma, St Louis, MO, USA) (see Note 2).
15. TEMED: N,N,N′,N′-Tetramethyl-ethylenediamine (Bio-Rad, Hercules, CA, USA) (see Note 3).
16. Ethidium bromide: (Sigma, St Louis, MO, USA) (see Note 4).
17. QIAquick Gel Extraction Kit and QIAGEN Plasmid Mini Kit: (Qiagen, Stanford, CA, USA).
18. Nitrocellulose filters: (Amersham).
19. Other common chemicals and reagents were purchased from Sigma (St Louis, MO, USA) and Sinopharm Chemical Reagents (Shanghai, China).
20. Gel electrophoresis apparatus: digital Electrophoresis Power Supply (Shenergy Biocolor, Shanghai, China).
21. Perkin Elmer (PE) thermocycler 9600: (Perkin Elmer, Foster City, CA, USA).
22. Bio-Rad Genepulser: (Bio-Rad, Hercules, CA, USA).
23. Double distilled water: ddH_2O.

2.2. Plasmids and Strains

1. The prokaryotic expression vector pYPX251 containing the promoter of the *aacC1* gene and T_1T_2 transcription terminator was constructed in the laboratory of Plant Molecular Biology (Biotechnology Research institute, Shanghai Academy of Agricultural Sciences, Shanghai, China), and deposited in GenBank (GenBank number: AY178046).
2. *E. coli* strain DH5α (SupE44ΔlacU169 φ80 lacZΔM15 hsdR17 recA1 endA1 gyrA96 thi-1 relA1) used for amplification of plasmid DNA and screening was purchased from Stratagene (La Jolla, California, USA).

2.3. Cell Cultures and Solutions

1. LB medium (per liter): 10 g tryptone, 5 g yeast extract, 5 g NaCl, adjust pH to 7 with about 1 mL NaOH (1 M). Autoclave 15 min at 121°C.
2. SOC medium (per liter): 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 10 mL KCl (250 mM), 5 mL $MgCl_2$ (2 M), adjust pH to 7 with about 1 mL NaOH (1 M). Autoclave 15 min at 121°C.
3. X-Gal: dissolved at 20.0 mg/mL in dimethylformamide.
4. Ampicillin: dissolved at 100 μg/mL in sterile deionised H_2O.

2.4. Agarose Gel Electrophoresis

1. Separating buffer, Tris–acetate–EDTA (TAE, 50×): 2 M Tris acetate, 0.1 M Acetic acid (glacial), 0.1 M EDTA, adjust pH to 8.5 using NaOH.
2. Loading buffer (6×): 0.25% (w/v) bromophenol blue, 30 % (w/v) glycerol in water.
3. Ethidium bromide is dissolved at 0.01 mg/mL in sterile deionised H_2O (see Note 4).

2.5. Polyacrylamide Gel Electrophoresis (PAGE)

1. Separating buffer, Tris–Borate–EDTA (TBE, 5×): 0.445 M Tris acetate, 0.9 M Borate, 0.01 M EDTA, adjust pH to 8.3 using NaOH.
2. Loading buffer (6×): 0.25% (w/v) Bromophenol Blue, 30% (w/v) Glycerol in water.
3. Thirty percent Acrylamide/Bis-acrylamide solution (per 100 mL): 29 g Acrylamide, 1 g Bis-acrylamide in deionised H_2O (see Note 2).
4. Ammonium persulfate: prepare 10% solution in water and immediately freeze in single-use (200 µL) aliquots at –20°C.

3. Methods

3.1. Agarose Gel Electrophoresis

1. Measure 2.0 g agarose powder and add to a 500 mL flask. Add 200 mL 1× TAE buffer to the flask. Melt the agarose in a microwave until completely melted (see Note 5).
2. Add ethidium bromide to the gel (final concentration 0.5 µg/mL) after cooling the solution to about 55°C. Pour the solution into a casting tray containing a sample comb. Let gel solidify at room temperature (see Note 4).
3. Remove the comb carefully after the gel has solidified.
4. Place the gel in the electrophoresis chamber.
5. Add enough TAE buffer to cover the gel with a layer of about 2–3 mm depth.
6. Pipette into the sample wells the samples containing DNA mixed with 6× loading buffer. Record the order of each sample to be loaded on the gel. Pipette 5 µL of the DNA 2000 marker into one or two wells of each row on the gel.
7. Place the lid on the gel box and connect the electrode wires to the power supply. Run the electrophoresis at 150 V for 20 min.
8. When adequate migration has occurred, place the gel on an ultraviolet transilluminator to visualize the DNA (see Notes 4 and 7).

3.2. Polyacrylamide Gel Electrophoresis (PAGE)

Polyacrylamide gels are chemically cross-linked, formed by the polymerization of acrylamide with a cross-linking agent, usually N, N′-methylene bisacrylamide (Bis). Polymerization initiates by free radical formation usually carried out with ammonium persulfate as the initiator and N, N, N′, N′-tetramethylene diamine (TEMED) as a catalyst. Polyacrylamide gel electrophoresis (PAGE) is a very useful technique, which has been used for the separation and purification of DNA (see Notes 2 and 3).

1. Wash both of the glass plates thoroughly with deionised water and liquid detergent. Rinse the glass plates thoroughly with deionised water. Air-dry the glass plates by laying them on a Whatman filter paper. Do not allow the treated surfaces to come into contact with one another. Assemble the glass plates by placing the side spacers between the plates. Place the assembled plates in the loading section of the gel apparatus.
2. Prepare a 30 mL 10% gel by mixing 10 mL 30% acrylamide/bis solution, with 3 mL of 5× TBE separating buffer, 12.5 mL water, 210 μL ammonium persulfate solution, and 10 μL TEMED. Pour the gel and insert the square tooth comb into the gel so that the wells are submerged in the in gel solution, but leaving a bit of air on the top to prevent bubble formation and to allow for the addition of extra gel to make up for the leakage prior to polymerization. The gel should polymerize in about 30–60 min (see Notes 2 and 3).
3. Add 0.5× TBE buffer solution, and then gently remove the comb.
4. Fill both the top and bottom chambers of the apparatus with approximately 300 mL of 0.5× TBE buffer, ensuring that the platinum wires of the apparatus are submerged at both ends.
5. Pipette into the sample wells the samples containing DNA mixed with 6× loading buffer. Record the order of each sample to be loaded on the gel. Pipette 5 μL of the DNA 2000 marker into one or two wells of each row on the gel.
6. Place the lid on the gel box, and connect the electrode wires to the power supply.
7. Run at 80 V for 2 h.
8. When adequate migration has occurred, remove the plates carefully from the apparatus and place on blotting paper. Remove spacers and separate plates carefully so that gel remains on the smaller plate (see Note 2).
9. After staining in ethidium bromide (0.5 μg/mL), DNA fragments can be visualized on a standard UV transilluminator (see Notes 4 and 7).

3.3. The First Part of DNA Shuffling

1. *Amplify the target gene by PCR*. Set up a mixture of the two outer oligonucleotides (30 pmol for each primer) with the appropriate full-length gene as a template (about 100 ng), 4 μL 2.5 mM dNTP mix (2.5 mM each dNTP), 5 μL 10× *Taq* buffer (Mg^{2+} plus), 0.5 μL *Taq* polymerase (5 U/μL), and ddH_2O to 50 μL. PCR can be done on a Perkin Elmer (PE) thermocycler 9600 for 30 cycles. The PCR conditions are 30 s at 94°C, 30 s at 60°C, and 2 min at 72°C for each cycle. The final cycle is followed by an additional 10 min at 72°C to ensure complete extension for all PCRs. Run 4% (2 μL) of the PCR products on a 1.0% agarose gel and electrophorese at 150 V for 20 min. A single sharp band of the expected size (1,533 bp) should be observed (see Fig. 1a). Estimate the concentration of each fragment on the gel by comparison to quantitative DNA markers.
2. *Purify the target PCR product*. Good-quality DNA is required for subsequent cutting of DNA into fragments, so a DNA purification kit can be used (e.g., Qiagen Plasmid Mini Kit). Using ddH_2O to dissolve, the final purification PCR product should be perfect.
3. *Cut the PCR product into random fragments with DNase I*. About 5 μg of the target PCR product can be digested by 0.002 unit of DNase I and 10 μL 10× DNase I buffer and water to a final volume of 100 μL for 20 min at 25°C. Then, follow with an additional 10 min at 80°C to ensure complete deactivation of DNase I.
4. *Purify the small fragments by PAGE*. Run 2% (2 μL) of the yielded products on a 1.0% agarose gel and electrophorese at 150 V for 20 min. A smear of the expected size (<300 bp) should be observed. Fragments of 50–100 bp can be separated by 10% PAGE and then purified using a dialysis bag (see Fig. 1b) (see Note 8).

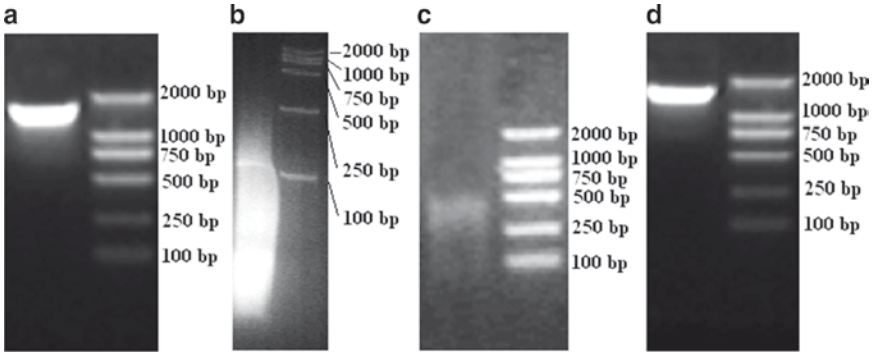


Fig. 1. The first part of DNA shuffling of *hlacz-sh* gene encoding a β-galactosidase of *P. woesei* (GenBank accession no. EF090269). (**a**) A 1,533 bp *hlacz-sh* DNA fragment amplified by PCR. (**b**) Random fragments with DNase I were separated by 10% PAGE. (**c**) Reassembly of the fragments in a PCR-like reaction without primers. (**d**) The final mutated products were amplified using the primerless PCR product as a template. Marker was DL 2000 marker.

5. *Primerless PCR and reassembly of the fragments in a PCR-like reaction without primers using the purified fragments as template*. Set up a mixture with 5 μL of purified small fragments from step 4 as template, 4 μL 2.5 mM dNTP mix (2.5 mM each dNTP), 5 μL 10× *Taq* buffer (Mg^{2+} free), 5 μL 25 mM $MgCl_2$, 0.5 μL *Taq* polymerase (5 U/μL), and ddH_2O to 50 μL. PCR can be done on a PE thermocycler 9600 for 45 cycles. The PCR conditions are 30 s at 90°C, 30 s at 42°C, and 1 min at 72°C for each cycle. The final cycle can be followed by an additional 10 min at 72°C to ensure complete extension for all PCRs. Run 4% (2 μL) of the PCR products on a 1.0% agarose gel and electrophorese at 150 V for 20 min. A smear should be observed (see Fig. 1c). Estimate the concentration of DNA in the gel by comparison to quantitative DNA markers.
6. *Amplify the shuffled product by PCR with primers*. Set up a mixture of the two outer oligonucleotides (30 pmol for each primer) with the primerless PCR product as a template (1 μL), 4 μL 2.5 mM dNTP mix (2.5 mM each dNTP), 5 μL 10× *Taq* buffer (Mg^{2+} plus), 0.5 μL *Taq* polymerase (5 U/μL), and ddH_2O to 50 μL. PCR can be done on a PE thermocycler 9600 for 20 cycles. The PCR conditions are 30 s at 94°C, 30 s at 60°C, and 2 min at 72°C for each cycle. The final cycle can be followed by an additional 10 min at 72°C to ensure complete extension for all PCRs. Run 4% (2 μL) of the PCR products on a 1.0% agarose gel and electrophorese at 150 V for 20 min. A single sharp band of the expected size (1,533 bp) should be observed (see Fig. 1d). Estimate the concentration of the fragment on the gel by comparison to quantitative DNA markers.

3.4. Construction of the Expression Vector

1. The gel slices containing the expected size band (1,533 bp) from primers PCR can be excised and extracted with QIAquick gel extraction kit.
2. The PCR products from the first part of DNA shuffling can be digested by restriction endonucleases *Bam* HI and *Sac* I. Set up a mixture with 1 μL (approximately 5 U) of restriction endonucleases *Bam* HI and *Sac* I, 10 μL (approximately 2 μg) of purified primer PCR product, 2.5 μL 10× K buffer, and ddH_2O to 50 μL. Then, incubate for 2 h at 37°C for digestion. High-quality DNA is required for subsequent ligation reaction, and a DNA purification kit can be used (e.g., Qiagen Plasmid Mini Kit). Using ddH_2O to dissolve the final purification products results in high-efficiency ligation reaction.
3. The digested fragments can be inserted into the prokaryotic expression vector pYPX251 at the *Bam* HI and *Sac* I sites. Set up a mixture with 1 μL (approximately 300 U) of T4 DNA

ligase, 10 μL of digestion products (approximately 1.5 μg), 1 μL of pYPX251 vector (approximately 0.5 μg), 5 μL T4 DNA ligase buffer, and ddH_2O to 50 μL. Then, incubate for 16 h at 16°C for ligation. High-quality DNA is required for subsequent transformation of *E. coli* by electroporation, and a DNA purification kit can be used (e.g., Qiagen Plasmid Mini Kit). Using ddH_2O to dissolve the final ligation products resulted in high-efficiency electro-transformation.

3.5. Transformation of E. coli by Electroporation

1. Prepare two new sterile Eppendorf tubes: one for mixing of the constructed library DNA and *E. coli* DH5α competent cells and one for incubation after electroporation. Place Eppendorf tubes, used for mixing of competent cells (50 μL) and DNA solution (1 μL, approximately 100 ng) by gently pipetting, on wet ice. Total six tubes.
2. Transfer the library DNA and DH5α competent cell suspension to a cold 0.1 cm electroporation cuvette (see Note 9).
3. Transform by electroporation using Bio-Rad Genepulser. Electroporate the cells at 1.7 kV/200 Ohm/25 μF with about 4.3–4.6 s between pulses.
4. Immediately after pulsing, 1 mL of SOC medium is added to the cuvette.
5. Transfer the cells out of the cuvette into the second 1.5 mL sterile Eppendorf tube.
6. Grow cells for 45 min to 1 h at 37°C and with rotation shaking at 150 rpm.
7. Plate cells on a 9-cm Petri dish containing LB solid medium and 100 μL/mL ampicillin.
8. Grow for 16 h at 37°C.

3.6. Screening and Selecting of Target Mutants

1. Transform the mutant library into the *E. coli* strain DH5α by electroporation, and plate the cells (approximately 200 μL) on thirty 9-cm Petri dishes containing LB medium and 100 μg/mL ampicillin.
2. Grow for 16 h at 37°C. About 300,000 variant colonies in the first part of DNA shuffling can be obtained using the pYPX251 vector.
3. The colonies formed on the culture medium are fixed onto a nitrocellulose filter and transferred colony-side-up to another Petri dish (see Fig. 2a).
4. The filter is incubated at 25°C in 0.03–0.04 mg/mL X-Gal.
5. One of the mutants with a higher activity, YH6754, was identified because it became blue more quickly and was of a deeper hue than the wild-type enzyme (see Fig. 2b) (see Note 10).

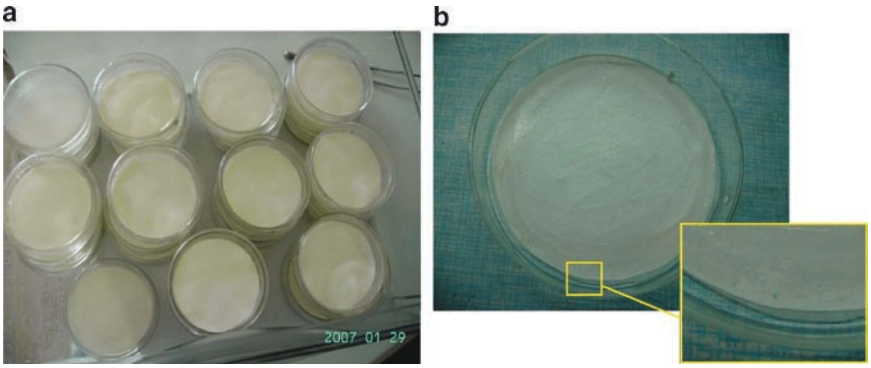


Fig. 2. High-throughput screening and selection of target mutants in directed evolution in vitro of a reporter gene. (**a**) The colonies formed on the culture medium were fixed onto a nitrocellulose filter and transferred colony-side-up to another Petri dish. (**b**) One of the higher activity mutants was identified because it became blue more quickly and was of a deeper hue than the wild-type enzyme.

3.7. DNA Sequencing, Analyzing, and Designing Degenerate Oligonucleotides

1. The evolved mutant gene was sequenced using the forward sequencing primer W73 (5′-GTA,GCG,TAT,GCG,CTC,AC G,CAA,CTG,G-3′) and the reverse sequencing primer W2467 (5′-GGC,AAA,TTC,TGT,TTT,ATC,AGA,CCG-3′) on the vector pYPX251.
2. The variant YH6754 revealed one AA deletion (P3) and seven AA substitutions (T29A, N106S, G177S, L217M, N277H, I387V, and T455A).
3. Eight mutated sites were identified as potential key amino acid residues. The degenerate domains were sites 2–11, 25–34, 104–113, 171–180, 210–219, 269–278, 382–391, and 450–459, which contained the mutated sites (see Fig. 3).
4. Designing degenerate oligonucleotides. The degenerate domain length was ten amino acid residues and contained the mutated site. Two oligonucleotides (one forward, one reverse) were designed according to each degenerate domain (see Fig. 3) (see Note 11).

3.8. The Second Part of DNA Shuffling Based on Degenerate Oligonucleotides

1. Amplify the control product (named product-ck) gene by PCR with the control *hlacz-sh* gene as the template. Amplify the gene (named product-0) by PCR with the plasmid DNA of YH6754 as the template. Set up a mixture of template (about 100 ng), the two outermost oligonucleotides (Hlacz-sh-S1 and Hlacz-sh-S2, 30 pmol for each primer), 4 μL 2.5 mM dNTP mix (2.5 mM each dNTP), 5 μL 10× *Taq* buffer (Mg^{2+} plus), 0.5 μL *Taq* polymerase (5 U/μL), and ddH_2O to 50 μL. PCR is carried out in a PE thermocycler 9600 for 30 cycles. The PCR conditions are 30 s at 94°C, 30 s at 60°C, and 2 min at 72°C for each cycle. The final cycle is followed by an additional 10 min at 72°C to ensure complete extension

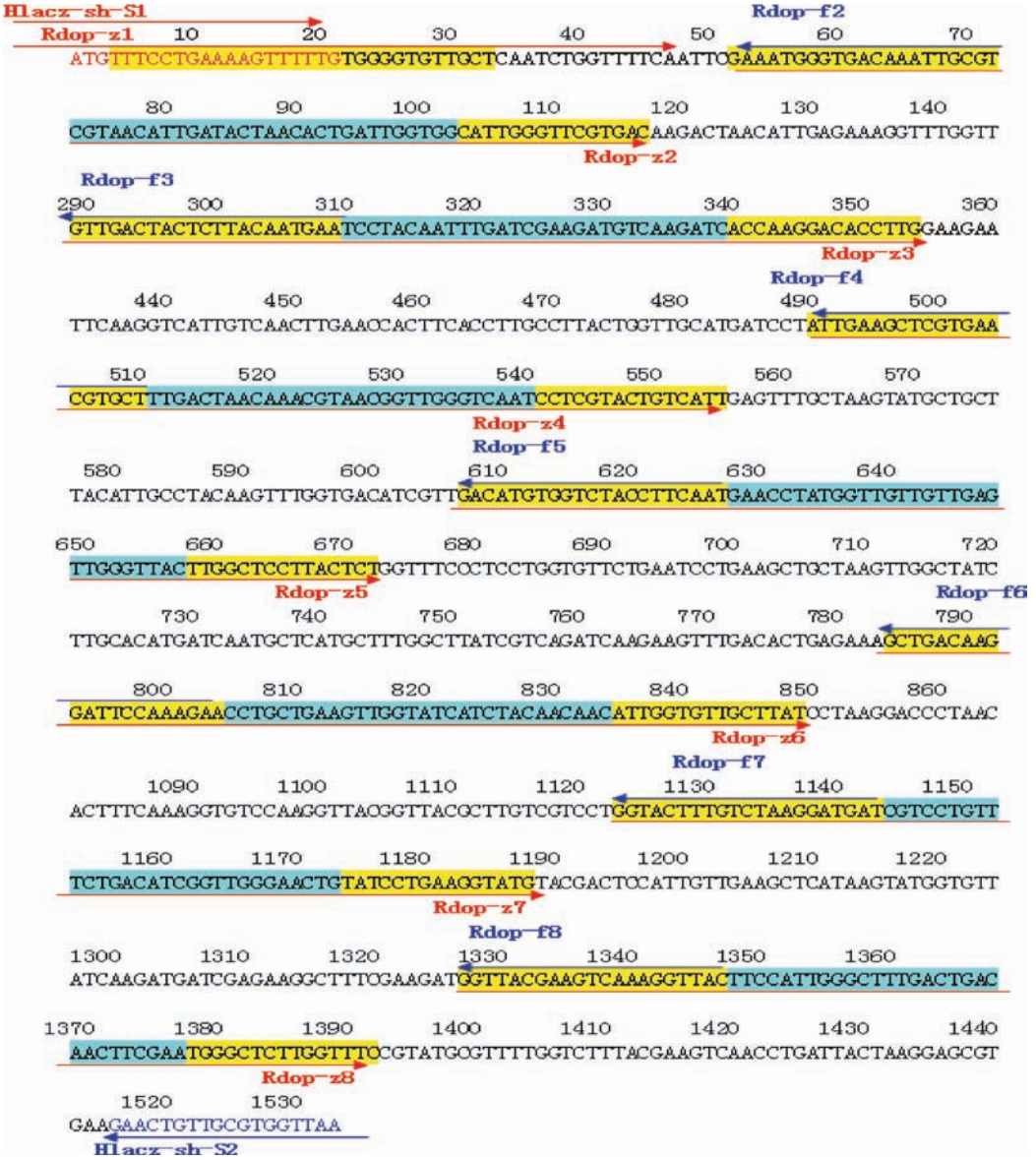


Fig. 3. The positions of designed degenerate oligonucleotides. Fifteen degenerate oligonucleotides, according to those eight AAs, and the oligonucleotide primers Hlacz-sh-S1 and Hlacz-sh-S2 were used to complete the second part of DNA shuffling. (Reproduced from ref. (29) with permission from Walter de Gruyter GmbH & Co. KG).

for all PCRs. Run 4% (2 μL) of the PCR products on a 1.0% agarose gel and electrophorese at 150 V for 20 min. A single sharp band of the expected size (1,533 bp) should be observed (see Fig. 4a).

2. The full-length product (named product-1) was amplified from the plasmid of YH6754 using the oligonucleotides Rdopz1 and Hlacz-sh-S2 as primers, which contain degenerate domain 1.

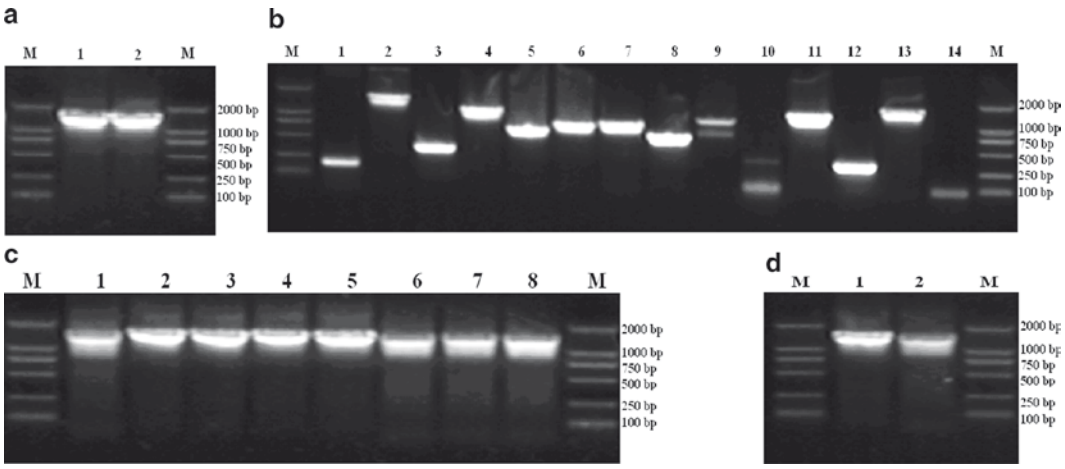


Fig. 4. The PCR-amplified products of the second part of DNA shuffling based on degenerate oligonucleotides. (**a**) The full-length product-ck (line 1) and full-length product-0 (line 2). (**b**) DNA segments I and II were amplified using eight groups of degenerate oligonucleotides separately: line 1, 2; 3, 4; 5, 6; 7, 8; 9, 10; 11, 12; 13, 14; and 15, 16 were segments I and II of product-2, product-3, product-4, product-5, product-6, product-7, and product-8. (**c**) The full-length products from the second step use overlap extension PCR with DNA segments I and II as the template and oligonucleotides Hlacz-sh-S1 and Hlacz-sh-S2 as primers. Line 1: product-1, Line 2: product-2, Line 3: product-3, Line 4: product-4, Line 5: product-5, Line 6: product-6, Line 7: product-7, and Line 8: product-8. (**d**) The final mutated gene. M: DL 2000 marker. (Reproduced from ref. (29) with permission from Walter de Gruyter GmbH & Co. KG).

The PCR conditions were the same as with the amplified product-ck gene. Run 4% (2 μL) of the PCR products on a 1.0% agarose gel and electrophorese at 150 V for 20 min. A single sharp band of the expected size (1,533 bp) should be observed (see Fig. 4c, Line 1).

3. The protocol for obtaining the full-length product (named product-2, which containing degenerate domain 2) involves two steps of PCR according to the following steps (Steps 4–7).
4. The first step is to use PCR to amplify the DNA segments I and II, which contain the degenerate domains 2, with the plasmid of YH6754 as a template using Hlacz-sh-S1/Rdopz2, Hlacz-sh-S2/Rdopf2 as primers, respectively.
5. Set up a mixture of template (about 100 ng), the two oligonucleotides (30 pmol for each primer), 4 μL 2.5 mM dNTP mix (2.5 mM each dNTP), 5 μL 10× *Taq* buffer (Mg^{2+} plus), 0.5 μL *Taq* polymerase (5 U/μL), and ddH_2O to 50 μL. PCR is carried out in a PE thermocycler 9600 for 30 cycles. The PCR conditions are 30 s at 94°C, 30 s at 60°C, and 1 min at 72°C for each cycle. The final cycle is followed by an additional 10 min at 72°C to ensure complete extension for all PCRs. Run 4% (2 μL) of the PCR products on a 1.0% agarose gel, and electrophorese at 150 V for 20 min (see Fig. 4b).

6. Mix the DNA segments I and II containing the degenerate domains 2.
7. The second step uses overlap extension PCR to recombine the entire mutant gene using DNA segments I and II as the template and oligonucleotides Hlacz-sh-S1 and Hlacz-sh-S2 as primers. Set up the mixture of DNA segment I and II (about 100 ng each), the two oligonucleotides (30 pmol for each primer), 4 μL 2.5 mM dNTP mix (2.5 mM each dNTP), 5 μL 10× *Taq* buffer (Mg^{2+} plus), 0.5 μL *Taq* polymerase (5 U/μL), and ddH_2O to 50 μL. PCR is carried out in a PE thermocycler 9600 for 30 cycles. The PCR conditions are 30 s at 94°C, 30 s at 60°C and 2 min at 72°C for each cycle. The final cycle is followed by an additional 10 min at 72°C to ensure complete extension for all PCRs. Run 4% (2 μL) of the PCR products on a 1.0% agarose gel and electrophorese at 150 V for 20 min. A single sharp band of the expected size (1,533 bp) should be observed (see Fig. 4c, Line 2).
8. The protocols for obtaining the other six full-length products (named product-3, product-4, product-5, product-6, product-7, and product-8, which contained degenerate domain 3, 4, 5, 6, 7, and 8, respectively) are similar with the protocol to obtain product-2. The first step is to use PCR to amplify the DNA segments I and II, which contain the degenerate domains 3, 4, 5, 6, 7, and 8, with the plasmid of YH6754 as a template using Hlacz-sh-S1/Rdopz3, Hlacz-sh-S2/Rdopf3; Hlacz-sh-S1/Rdopz4, Hlacz-sh-S2/Rdopf4; Hlacz-sh-S1/Rdopz5, Hlacz-sh-S2/Rdopf5; Hlacz-sh-S1/Rdopz6, Hlacz-sh-S2/Rdopf6; Hlacz-sh-S1/Rdopz7, Hlacz-sh-S2/Rdopf7; and Hlacz-sh-S1/Rdopz8, Hlacz-sh-S2/Rdopf8 as primers, respectively. Run 4% (2 μL) of those PCR products on a 1.0% agarose gel and electrophorese at 150 V for 20 min. A single sharp band of the expected size (1,533 bp) should be observed (see Fig. 4c, Line 3–8).
9. Purify the ten full-length PCR products (product-ck, product-0, product-1, product-2, product-3, product-4, product-5, product-6, product-7, and product-8). Then mix the purified PCR products. High-quality DNA is required for subsequent digesting DNA into fragments, and a DNA purification kit can be used (e.g., Qiagen Plasmid Mini Kit). Using ddH_2O to dissolve the final purification products resulted in perfect PCR products.
10. Cut the mixed purified PCR products into random fragments with DNase I. About 5 μg of the mixed PCR products were digested by 0.002 unit of DNase I with 10 μL 10× DNase I buffer and water in a final volume of 100 μL for 20 min at 25°C, followed by an additional 10 min at 80°C to ensure DNase I complete deactivation. Run 2% (2 μL) of the cutting

products on a 1.0% agarose gel and electrophorese at 150 V for 20 min.

11. Gel purification of small fragments by PAGE. Fragments of 50–100 bp separated quite well by 10% PAGE, and they were purified using a dialysis bag (see Note 8).
12. Primerless PCR. Reassembly of the fragments in a PCR-like reaction without primers using the purified fragments as a template. Set up a mixture with 4 μL 2.5 mM dNTP mix (2.5 mM each dNTP), 5 μL 10× *Taq* buffer (Mg^{2+} free), 5 μL 25 mM $MgCl_2$, 0.5 μL *Taq* polymerase (5 U/μL), and ddH_2O to 50 μL. PCR is performed in a PE thermocycler 9600 for 45 cycles. The PCR conditions are 30 s at 90°C, 30 s at 42°C, and 1 min at 72°C for each cycle. The final cycle is followed by an additional 10 min at 72°C to ensure complete extension for all PCRs.
13. Primer PCR for amplification of the full-length shuffling product with the primerless PCR product as a template and two outermost oligonucleotides (Hlacz-sh-S1 and Hlacz-sh-S2) as primers. Set up a mixture of the two outer oligonucleotides (30 pmol for each primer) with the primerless PCR product as a template (1 μL), 4 μL 2.5 mM dNTP mix (2.5 mM each dNTP), 5 μL 10× *Taq* buffer (Mg^{2+} plus), 0.5 μL *Taq* polymerase (5 U/μL), and ddH_2O to 50 μL. PCR is performed in a PE thermocycler 9600 for 20 cycles. The PCR conditions are 30 s at 94°C, 30 s at 60°C, and 2 min at 72°C for each cycle. The final cycle is followed by an additional 10 min at 72°C to ensure complete extension for all PCRs. Run 4% (2 μL) of the PCR products on a 1.0% agarose gel and electrophorese at 150 V for 20 min. A single sharp band of the expected size (1,533 bp) should be observed (see Fig. 4d). Estimate the concentration of the band on the gel by comparison to quantitative DNA markers.

3.9. The Second Part of Construction of the Expression Vector, Transformation of E. coli by Electroporation, Screening, and Selection of Target Mutants

1. The protocols of "Construction of the expression vector," "Transformation of *E.coli* by electroporation," "Screening", and "Selecting of target mutants" are similar to the first part of the directed evolution protocol, except that the filter is incubated at 25°C in 0.01–0.02 mg/mL X-Gal.
2. Two of the mutants with higher activity, YG6755 and YG8252, were identified because they became blue more quickly and were of deeper hue than YH6754 (see Note 10).
3. The specific β-galactosidase activities of those variants were increased by 68.8% (YG6755), and 72.7% (YG8252) compared to the wild type. The specific β-galactosidase activities of those variants by the second part of the strategy were increased by another 41.4% (YG6755) and 45.3% (YG8252) compared to the YH6754.

4. Notes

1. Poor-quality oligonucleotides is the major cause for errors in the end product. Because the use of PAGE-purified oligonucleotides can reduce error rates several fold, PAGE-mediated removal of oligonucleotides of incorrect length is essential for the synthesis of high-quality DNA. However, mutations caused by nucleotide substitutions cannot be identified by PAGE (45, 46).
2. CAUTION: Acrylamide and Bis-acrylamide are neurotoxins. Handle solids and solutions containing acrylamide and bisacrylamide with care. Wear a mask and gloves when weighing out solid acrylamide.
3. CAUTION: TEMED is hazardous. Always wear gloves when handling gels and any solutions containing TEMED.
4. CAUTION: Ethidium bromide (EtBr) is a mutagen. Always wear gloves when handling gels and any solutions containing ethidium bromide.
5. CAUTION: Harmful exposure to microwave energy.
6. The gel is placed on an ultraviolet transilluminator. Be aware that DNA will diffuse within the gel over time, and examination or photography should take place shortly after cessation of electrophoresis.
7. CAUTION: Always wear protective eyewear when observing DNA on a transilluminator to prevent damage to the eyes from UV light.
8. In DNA shuffling, after DNase I shearing, it was difficult to obtain the full-length sequence using 10–50 bp fragments when the full-length gene was over 1.5 kb. When the fragments are over 100 bp, the mutation rate was relatively low, especially when the full-length of DNA was less than 1.0 kb.
9. Throughout this procedure of transformation of *E. coli* by electroporation, all reaction solutions should be stored and handled on ice, unless otherwise stated.
10. The selected variant could not be isolated easily if the colony density was too high. The small area of the nitrocellulose filters containing the bluest colonies were cut out using scissors. The small nitrocellulose filter was placed into a 500 μL tube, and 40 μL LB medium was added. The eluted bacteria were plated on a 9-cm Petri dish containing LB medium and 100 μL/mL ampicillin, the mutant was selected again.
11. Forward oligonucleotides should be 66 bp long, with 21 bp (~56°C Tm) overlaps between neighboring reverse oligonucleotides. The degenerate domain should be positioned

in the middle of the forward oligonucleotides. The reverse oligonucleotide should be 21 bp long, with 21 bp (approximately 56°C Tm) overlaps between neighboring forward oligonucleotides. In order to avoid the stop codon, designed degenerate oligonucleotides were used: N(A/C/G/T), and some others, such as K(G/T), R(A/G), M(A,C), W(A/T), S(C/G), Y(C/T), V(A/C/G), H(A/C/T), D(A/G/T), and B(C/G/T). (Reproduced from ref. (29) with permission from Walter de Gruyter GmbH & Co. KG).

Acknowledgments

The author would like to thank Walter de Gruyter GmbH & Co. KG for permission to reproduce some figures. This research was supported by the Project for International Scientific and Technological Cooperation (Shanghai-Alberta, 08540706500); Shanghai Rising-Star Program (08QH14021); Hi-tech research and development program of China (2006AA10Z117-06Z358, 2008AA10Z401); Shanghai and Nation Natural Science Foundation (08ZR1417200, 30670179); Shanghai Key laboratory project (07dz22011). For this chapter, Professor Quan-Hong Yao is the corresponding author. We also thank Lori Osburn for critical reading of the manuscript.

References

1. Stemmer WP (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370:389–391
2. Stemmer WP (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc Natl Acad Sci USA 91:10747–10751
3. Arnold FH, Volkov AA (1999) Directed evolution of biocatalysts. Curr Opin Chem Biol 3:54–59
4. Chirumamilla RR, Muralidhar R, Marchant R, Nigam P (2001) Improving the quality of industrially important enzymes by directed evolution. Mol Cell Biochem 224:159–168
5. Zhao H, Chockalingam K, Chen Z (2002) Directed evolution of enzymes and pathways for industrial biocatalysis. Curr Opin Biotechnol 13:104–110
6. Alexeeva M, Carr R, Turner NJ (2003) Directed evolution of enzymes: new biocatalysts for asymmetric synthesis. Org Biomol Chem 1:4133–4137
7. Hibbert EG, Baganz F, Hailes HC, Ward JM, Lye GJ, Woodley JM, Dalby PA (2005) Directed evolution of biocatalytic processes. Biomol Eng 22:11–19
8. Otten LG, Quax WJ (2005) Directed evolution: selecting today's biocatalysts. Biomol Eng 22:1–9
9. Lassner M, Bedbrook J (2001) Directed molecular evolution in plant improvement. Curr Opin Plant Biol 4:152–156
10. Dixon DP, McEwen AG, Lapthorn AJ, Edwards R (2003) Forced evolution of a herbicide detoxifying glutathione transferase. J Biol Chem 278:23930–23935
11. Castle LA, Siehl DL, Gorton R, Patten PA, Chen YH, Bertain S, Cho HJ, Duck N, Wong J, Liu D, Lassner MW (2004) Discovery and directed evolution of a glyphosate tolerance gene. Science 304:1151–1154
12. Chartrain M, Salmon PM, Robinson DK, Buckland BC (2000) Metabolic engineering and directed evolution for the production of pharmaceuticals. Curr Opin Biotechnol 11:209–214
13. Whalen RG, Kaiwar R, Soong NW, Punnonen J (2001) DNA shuffling and vaccines. Curr Opin Mol Ther 3:31–36

14. Locher CP, Paidhungat M, Whalen RG, Punnonen J (2005) DNA shuffling and screening strategies for improving vaccine efficacy. DNA Cell Biol 24:256–263
15. Locher CP, Soong NW, Whalen RG, Punnonen J (2004) Development of novel vaccines using DNA shuffling and screening strategies. Curr Opin Mol Ther 6:34–39
16. Crameri A, Raillard SA, Bermudez E, Stemmer WP (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature 391:288–291
17. Zhang YX, Perry K, Vinci VA, Powell K, Stemmer WP, Cardayre SB (2002) Genome shuffling leads to rapid phenotypic improvement in bacteria. Nature 415:644–646
18. Fujii R, Kitaoka M, Hayashi K (2004) One-step random mutagenesis by error-prone rolling circle amplification. Nucleic Acids Res 32:e145
19. Kikuchi M, Ohnishi K, Harayama S (2000) An effective family shuffling method using single-stranded DNA. Gene 243:133–137
20. Gibbs MD, Nevalainen KM, Bergquist PL (2001) Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling. Gene 271:13–20
21. Coco WM (2003) RACHITT: Gene family shuffling by Random Chimeragenesis on Transient Templates. Methods Mol Biol 231:111–127
22. Murakami H, Hohsaka T, Sisido M (2002) Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. Nat Biotechnol 20: 76–81
23. Osuna J, Yáñez J, Soberón X, Gaytán P (2004) Protein evolution by codon-based random deletions. Nucleic Acids Res 32:e136
24. Ghadessy FJ, Ong JL, Holliger P (2001) Directed evolution of polymerase function by compartmentalized self-replication. Proc Natl Acad Sci USA 98:4552–4557
25. Bergquist PL, Reeves RA, Gibbs MD (2005) Degenerate oligonucleotide gene shuffling (DOGS) and random drift mutagenesis (RNDM): two complementary techniques for enzyme evolution. Biomol Eng 22:63–72
26. Kashiwagi K, Isogai Y, Nishiguchi K, Shiba K (2006) Frame shuffling: a novel method for in vitro protein evolution. Protein Eng Des Sel 19:135–140
27. Song JK, Chung B, Oh YH, Rhee JS (2002) Construction of DNA-shuffled and incrementally truncated libraries by a mutagenic and unidirectional reassembly method: changing from a substrate specificity of phospholipase to that of lipase. Appl Environ Microbiol 68:6146–6151
28. Barriault D, Plante MM, Sylvestre M (2002) Family shuffling of a targeted *bphA* region to engineer biphenyl dioxygenase. J Bacteriol 184:3794–3800
29. Xiong AS, Peng RH, Zhuang J, Liu JG, Gao F, Xu F, Cai B, Yao QH (2007) A semi-rational design strategy of directed evolution combined with chemical synthesis of DNA sequences. Biol Chem 388:1291–1300
30. Morley KL, Kazlauskas RJ (2005) Improving enzyme properties: when are closer mutations better? Trends Biotechnol 23:231–237
31. Parikh MR, Matsumura I (2005) Site-saturation mutagenesis is more efficient than DNA shuffling for the directed evolution of beta-fucosidase from beta-galactosidase. J Mol Biol 352:621–628
32. Xiong AS, Yao QH, Zhang Z, Peng RH, Zhuang J, Xu F, Liu JG, Zhu H (2006) Rational evolutionary design and modification of gene: a short-path of direct evolution. Yi Chuan 28:92–96
33. Geddie ML, Matsumura I (2004) Rapid evolution of beta-glucuronidase specificity by saturation mutagenesis of an active site loop. J Biol Chem 279:26462–26468
34. Miki B, McHugh S (2004) Selectable marker genes in transgenic plants: applications, alternatives and biosafety. J Biotechnol 107:193–232
35. Zhang JH, Dawes G, Stemmer WP (1997) Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. Proc Natl Acad Sci USA 94:4504–4509
36. Crameri A, Whitehorn EA, Tate E, Stemmer WP (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. Nat Biotechnol 14:315–319
37. Matsumura I, Ellington AD (1996) DNA shuffling brightens prospects for GFP. Nat Biotechnol 14:366
38. Nam SH, Oh KH, Kim GJ, Kim HS (2003) Functional tuning of a salvaged green fluorescent protein variant with a new sequence space by directed evolution. Protein Eng 16:1099–1105
39. Fradkov AF, Chen Y, Ding L, Barsova EV, Matz MV, Lukyanov SA (2000) Novel fluorescent protein from Discosoma coral and its mutants possesses a unique far-red fluorescence. FEBS Lett 479:127–130
40. Matsumura I, Ellington AD (2001) In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. J Mol Biol 305:331–339

41. Flores H, Ellington AD (2002) Increasing the thermal stability of an oligomeric protein, beta-glucuronidase. J Mol Biol 315:325–337
42. Xiong AS, Peng RH, Liu JG, Zhuang J, Qiao YS, Xu F, Cai B, Zhang Z, Chen JM, Yao QH (2007) High efficiency and throughput system in directed evolution in vitro of reporter gene. Appl Microbiol Biotechnol 74:160–168
43. Xiong AS, Peng RH, Cheng ZM, Li Y, Liu JG, Zhuang J, Gao F, Xu F, Qiao YS, Zhang Z, Chen JM, Yao QH (2007) Concurrent mutations in six amino acids in beta-glucuronidase improve its thermostability. Protein Eng Des Sel 20:319–325
44. Matsumura I, Wallingford JB, Surana NK, Vize PD, Ellington AD (1999) Directed evolution of the surface chemistry of the reporter enzyme beta-glucuronidase. Nat Biotechnol 17:696–701
45. Xiong AS, Yao QH, Peng RH, Li X, Fan HQ, Cheng ZM, Li Y (2004) A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences. Nucleic Acids Res 32:e98
46. Xiong AS, Yao QH, Peng RH, Duan H, Li X, Fan HQ, Cheng ZM, Li Y (2006) PCR-based accurate synthesis of long DNA sequences. Nat Protoc 1:791–797

Chapter 19

Ribosome Display for Rapid Protein Evolution by Consecutive Rounds of Mutation and Selection

Hayato Yanagida, Tomoaki Matsuura, and Tetsuya Yomo

Abstract

Directed evolution experiments are performed to improve the properties of proteins by creating a library of mutated genes of interest and selecting those genes that encode proteins exhibiting desired properties. Here, we present one of the methods to carry out an evolutionary experiment called ribosome display. Ribosome display allows this process to be carried out entirely in vitro, and it is therefore a rapid and robust method for protein evolution.

Key words: Random mutagenesis, Ribosome display, Directed evolution, PURE system, In vitro translation

1. Introduction

Evolutionary methods have been used to improve the biophysical properties of proteins or even create new proteins with novel functions. Directed evolution methods mainly consist of two steps: diversification and selection. A gene encoding the protein is mutated to generate a library of genes, and proteins exhibiting the properties of interest are selected. Genes encoding selected proteins are amplified, and subjected to further rounds of mutagenesis, selection, and amplification. The consecutive rounds of diversification and selection allow us to generate proteins with specific desired properties. To fulfill this purpose, it is necessary to have a linkage between the phenotype exhibited by the proteins and the gene that encodes the phenotype.

All living cells establish the linkage between genotype and phonotype by compartmentalizing genes according to cell membranes, whereas evolutionary methods, such as phage

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_19, © Springer Science+Business Media, LLC 2010

display (1), mRNA display (2), ribosome display (3), and in vitro compartmentalization (4), establish the linkage in different ways. Among these methods, in vitro selection technologies such as mRNA display and ribosome display are performed entirely in vitro, which provides mainly two advantages over in vivo selection technologies such as phage display. First, very large libraries of up to 10^{14} members can be handled because the size is not limited by cellular transformation efficiency. Second, random mutations can be introduced without the subsequent transformation steps. Here, we describe an efficient and rapid procedure to perform ribosome display (Fig. 1).

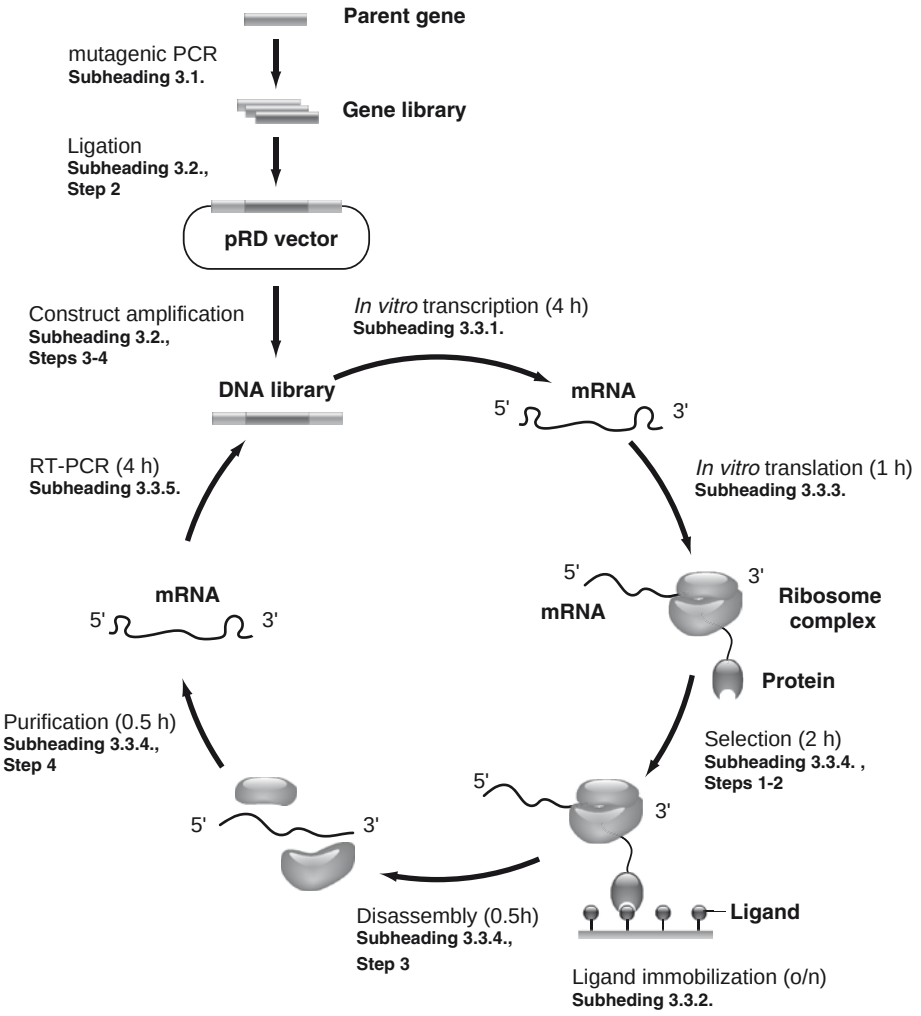


Fig. 1. Preparation of ribosome display construct and principle of ribosome display selection. A library of DNA molecules encoding the proteins of interest is transcribed and translated in vitro. The resulting mRNA lacks a stop codon, giving rise to ternary complexes of mRNA-ribosome-encoded protein. These are used directly for affinity selection. The mRNA from the bound complexes is eluted and purified. RT-PCR can introduce mutations and yields a DNA pool enriched for binders that can be used for the next iteration.

In ribosome display, an in vitro translation system is used to generate ribosome complexes consisting of a nascent chain, ribosome, and mRNA during translation, thereby establishing the physical linkage between genotype and phenotype (Fig. 1). Such complexes can be stabilized by the absence of a stop codon in the mRNA template. In addition, low temperatures and high magnesium ion concentrations are known to be key factors to stabilize this ternary complex. The stabilized complexes can then be used directly for affinity selections (panning experiments). During the experiments, mutagenesis procedures can be carried out as desired. If a low-fidelity DNA polymerase such as *Taq* DNA polymerase is used for the PCR amplification step, the amplification step after each round of selection can lead to diversification of the DNA pool. On the other hand, if screening and selection of an unaltered library are required for enrichment, a high-fidelity DNA polymerase can also be used for the PCR purification step. Diversification during PCR can be further enhanced by mutagenesis methods, such as oligonucleotide-directed mutagenesis (5), error-prone PCR in the presence of non-biological metal ions such as Mn^{2+} (6), or dNTP analogs (7). Other PCR methods for diversification involve homologous recombination in vitro, such as DNA shuffling (8) and the staggered extension process (StEP) (9). As both mutagenesis and selection in ribosome display take place entirely in vitro, the evolution cycles can be carried out within one day.

We describe a strategy for carrying out ribosome display experiments using the PURE system (10), which is an in vitro translation system in which all components necessary for translation are highly purified and reconstituted to carry out the protein synthesis reaction in a test tube. Mainly due to the very low RNase activity of this system compared to that of the *Escherichia coli* S30 extract, more than tenfold greater amounts of mRNA can be recovered after selection (11). In addition, the complexes generated in the PURE system were found to be highly stable over a wide range of temperatures in contrast to the previous assumption using *E. coli* S30 extract (12). Using this system, a process of compensatory evolution of a defective protein variant of the WW domain lacking a strictly conserved amino acid residue was demonstrated (13). The WW domain is one of the smallest protein modules (30–40 residues) consisting of a three-stranded antiparallel β-sheet structure, which mediates protein–protein interaction by recognizing proline-rich peptide sequences (14–16). The WW domain carries two highly conserved Trp residues separated by a stretch of 20–22 amino acids, for which it is named. A previous study on the human Yes-associated protein (hYAP) WW domain demonstrated the functional and structural importance of the two highly conserved Trp residues (16, 17) and indicated that the N-terminal Trp17 is involved in the hydrophobic

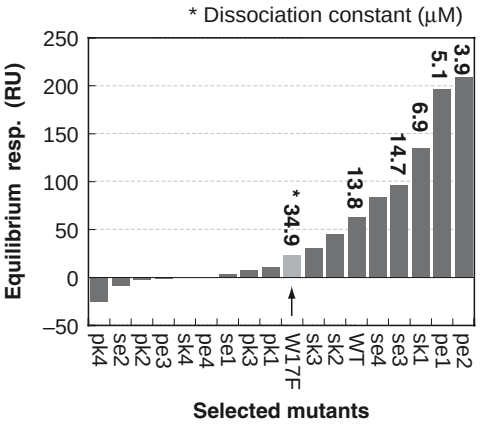


Fig. 2. Binding of the wild-type and W17F mutant of the WW domain, and 16 variants obtained after four rounds of ribosome display to the PY ligand evaluated by surface plasmon resonance (13). Each bar represents the binding signal (RU) at equilibrium with samples at 10 μM. Numbers above the bars show dissociation constants with the PY ligand. The parent W17F is indicated with an *arrow*.

patch and is essential for maintaining the overall β-sheet structure. Indeed, replacement of Trp17 with Phe resulted in significant loss of structural integrity and reduced binding affinity (K_d = 15.1 μM) to the PY ligand (EYPPYPPPPYPSG) compared to the wild type (K_d = 5.9 μM) (17). Thus, the W17F mutant of the hYAP WW domain lacks the strictly conserved residue and the overall structural integrity, and exhibits reduced binding function (17).

The W17F mutant of the WW domain was used as a model protein with a defective structure and function. First, a gene library of W17F mutants was prepared with random mutagenesis by error-prone PCR, and selection experiments were performed based on PY ligand binding function. After four rounds of ribosome display selection, second-site mutations could be selected that compensated for the defective function with up to tenfold higher affinity than the W17F mutant (Fig. 2). The results presented here together with those obtained by other groups (11, 18) demonstrated the efficacy of using the PURE system for rapid in vitro protein evolution by ribosome display.

2. Materials

2.1. Plasmid and Primers

1. pRD-n1n2_2 plasmid (13): a derivative of pQE30 (Qiagen).
2. SDA-pqe primer: (5′-AGACCACAACGGTTTCCCTCTAG AAATAATTTTGTTTA ACTTTAAGAAAGAGGAGAAA TTAA CTATGAGA-3′).

3. pqe1- primer: (5′-GATCTATCAACAGGAGTCCAAGCT CA-3′).
4. T7B primer: (5′-ATACGAAATTAATACGACTCACTATA GGGAGACCACAAC GG-3′).
5. T3te_pD- primer: (5′-CGGCCCACCCGTGAAGGTGAG CC-3′).

2.2. Mutagenic and Standard PCR

1. 10× mutagenic PCR buffer: 670 mM Tris–HCl (pH 8.8), 166 mM $(NH_4)_2SO_4$, 1.7 mg/mL BSA, 67 μM EDTA (pH 8.0), 61 mM $MgCl_2$.
2. 10× dNTP mix: 2 mM dATP, 10 mM dGTP, 10 mM dCTP, 10 mM dTTP.
3. 100 μM forward and reverse primers.
4. 5 mM $MnCl_2$.
5. *Δ*Tth DNA polymerase (Toyobo).
6. 10× *Ex Taq* buffer (Takara).
7. dNTPs (Takara).
8. *Ex Taq* DNA polymerase (Takara).

2.3. Gel Electrophoresis for DNA and RNA

1. TAE buffer: 40 mM Tris–acetate, 2 mM EDTA. Adjust pH to 8.0.
2. TBE buffer: 50 mM Tris–borate, 10 mM EDTA. Adjust pH to 8.0.
3. 1.5% Agarose in TAE or TBE buffer.
4. 1 M guanidinium isothiocyanate.
5. Formamide.
6. 37% formaldehyde.
7. 10× MOPS buffer: 0.2 M MOPS (pH 7.0), 50 mM sodium acetate, 10 mM EDTA.

2.4. In Vitro Transcription

1. 5× T7 RNA polymerase buffer: 1 M HEPES–KOH (pH 7.6), 150 mM magnesium acetate, 10 mM spermidine, 200 mM DTT.
2. NTP mix: 50 mM each NTP.
3. T7 RNA polymerase (Takara).
4. RNase inhibitor RNasin (Promega).
5. DNaseI (RNase free; Takara).
6. RNeasy kit (Qiagen).

2.5. In Vitro Translation

1. PURE system S-S (Post Genome Institute).
2. Wash buffer (WBKT): 50 mM Tris–acetate (pH 7.5), 150 mM NaCl, 50 mM magnesium acetate, 0.5 M KCl, 0.1% Tween 20.

3. 200 mg/mL heparin.
4. 5% BSA.

2.6. Selection Round

1. Microtiter plates (Maxisorp; Nunc).
2. 2 μg/mL NeutrAvidin (Pierce).
3. Phosphate-buffered saline (PBS, pH 7.4): 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4.
4. PBS with Tween (PBST): PBS supplemented with 0.1% Tween 20.
5. Biotinylated PY ligand (EYPPYPPPPYPSG) (19). Synthesize peptide with standard Fmoc protocols and purify by reverse-phase HPLC. Biotinylation of PY ligand is performed by labeling the N-terminal amine of the ligand with EZ-link NHS-SS-biotin (Pierce, Rockford, IL) in accordance with the manufacturer's instructions.
6. Elution buffer: 50 mM Tris–acetate (pH 7.5), 150 mM NaCl, 50 mM EDTA.
7. 2.5 mg/mL *Saccharomyces cerevisiae* RNA.
8. OneStep RT-PCR kit (Qiagen).
9. QIAquick gel extraction kit (Qiagen).

2.7. Analysis of Selection

1. SuperScript III Platinum One-Step qRT-PCR kit (Invitrogen).
2. (^{35}S)Methionine (PerkinElmer).

3. Methods

3.1. Mutagenic PCR

To prepare a randomly mutated gene library, design primers to introduce, for example, an *Nco*I site just before the initiation codon (methionine) and a *Bam*HI site just after the coding sequence to exclude the stop codon and amplify the DNA encoding the gene of interest by error-prone PCR using *Δ*Tth DNA polymerase (see Note 1). *Nco*I and *Bam*HI sites are used to clone the genes into the plasmid (pRD-n1n2_2) used for ribosome display.

1. Set up PCR mixtures in a volume of 89 μL containing 1× mutagenic PCR buffer, 1× dNTP mix, 1 μM primers, and 10–100 ng template. Then, add 10 μL of 5 mM $MnCl_2$ followed by 1 μL (0.5 U) of *Δ*Tth DNA polymerase.
2. Perform PCR for 30 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min/kb of desired product.
3. Check the yield and size of the PCR products by agarose gel electrophoresis.

4. Purify the PCR product using a QIAquick PCR purification kit in accordance with the manufacturer's instructions.

3.2. Preparation of Ribosome Display Constructs

1. Digest the PCR product with *Nco*I/*Bam*HI, and purify using a QIAquick PCR purification kit or gel purification kit.
2. Ligate the digested PCR product into the *Nco*I/*Bam*HI sites of the vector pRD-n1n2_2.
3. PCR-amplify the ligation products directly using *Ex Taq* DNA polymerase in 50-μL reactions containing 1× *Ex Taq* buffer, 4 μL of dNTP mix, 1 μM primers SDA-pqe and pqe1-, 1 μL of ligation mixture, and 1 μL of *Ex Taq* DNA polymerase. Perform PCR for 20–25 cycles of 98°C for 10 s, 55°C for 1 min, and 72°C for 1 min/kb of desired product. Check the yield and the size of the PCR product by agarose gel electrophoresis.
4. PCR-amplify the product of the previous step with the primers T7B and T3te_pD- to introduce the T7 promoter and the 5′ and 3′ stem loops to make the final DNA construct useful for in vitro transcription (Fig. 3). Use the same PCR conditions as in the previous step. Check the yield and size of the PCR product by agarose gel electrophoresis.

3.3. Ribosome Display Experiments

3.3.1. In Vitro Transcription

1. Set up reaction mixtures in a volume of 100 μL containing 1× T7 RNA polymerase buffer, 14 μL of NTP mix, 4 μL (200 U) of T7 RNA polymerase, 2 μL (80 U) of RNasin, and 0.5–1 μg of the PCR product (see step 4 of Subheading 3.2).
2. Incubate the solutions for 3 h at 37°C.
3. Add 2 μL (10 U) of DNaseI to the reaction mixtures, and incubate for 10 min at 37°C.
4. Purify the reaction product using a Qiagen RNeasy purification kit in accordance with the manufacturer's instructions.
5. Determine the concentration of nucleic acid from the absorbance at 260 nm.

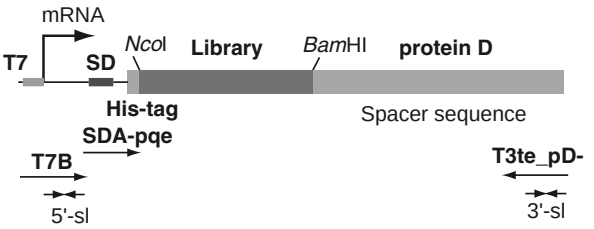


Fig. 3. The DNA construct used for ribosome display is described here. The T7promoter is followed by a Shine–Dalgarno sequence (SD), an N-terminal his-tag, and the DNA library of the protein of interest. The DNA library, which is cloned into the *Nco*I and *Bam*HI restriction sites of pRD-n1n2_2, is followed by protein D as a spacer sequence. Sequences encoding stem-loop structures are present at both ends (5′-sl and 3′-sl).

6. Check the quality of the mRNA by agarose gel electrophoresis. Cast agarose gels containing 2% (w/v) 1 M guanidinium isothiocyanate. Denature aliquots of 5 μg of mRNA by incubation for 10 min at 70°C in 15.5 μL of loading buffer (10 μL formamide, 3.5 μL 37% formaldehyde, 2 μL 5× MOPS) and analyze by electrophoresis.

3.3.2. Ligand Immobilization

1. Coat Maxisorp microtiter plates with 100 μL of 2 μg/mL NeutrAvidin dissolved in PBS at 4°C overnight.
2. Block the plates by incubation with 5% BSA for 2 h at room temperature.
3. Wash the coated plates three times with 300 μL of PBST.
4. Incubate the plates with 100 μL of 0.5 μM biotinylated ligand for 0.5 h at room temperature.
5. Wash the plates 5 times with 300 μL of ice-cold WBKT, and store the plates filled with 300 μL of WBKT at 4°C.

3.3.3. In Vitro Translation

1. Perform in vitro translation using the PURE system (see Notes 2 and 3), an *E. coli*-based reconstituted in vitro translation system. Briefly, set up 50-μL protein translation reactions containing 2 μg of mRNA and 1 μL of RNasin in accordance with the manufacturer's instructions.
2. Incubate the solutions for 0.5–1 h at 37°C.
3. Stop the reaction by adding 400 μL of ice-cold WBKT supplemented with 2.5 mg/mL heparin and 0.5% BSA.

3.3.4. Panning Procedure

1. Add an aliquot of 100 μL of the stopped reaction containing the ternary complexes into each well of ligand-immobilized plates prepared as described under Subheading 3.3.2 (see Note 4), and then shake gently at 4°C for 1 h.
2. Wash plates 5 times with 300 μL of WBKT, with 1 and 5 min of shaking between the first two and between the last three washes, respectively (see Note 5).
3. Elute the retained ternary complexes with 200 μL of ice-cold elution buffer containing 50 μg/mL *S. cerevisiae* RNA for 0.5 h at 4°C.
4. Purify the released mRNA using a Qiagen RNeasy kit.

3.3.5. Preparation of mRNA for the Next Round of Selection

Amplify the purified mRNA by RT-PCR using a OneStep RT-PCR kit (see Note 6) in accordance with the manufacturer's instructions using the same primers as described previously (see Subheading 3.2, step 4 and Note 7). To carry out an additional round of affinity selection on the isolated pools, the amplified products can be used directly or after gel purification, if necessary, for in vitro transcription (see Subheading 3.3.1).

1. Set up RT-PCR mixtures in a volume of 50 µL containing 1× Qiagen OneStep RT-PCR buffer, 2 µL of dNTP mix, 0.6 µM primers T7B and T3te_pD-, 2 µL of enzyme mix, and 5 µL of purified mRNA.
2. Carry out RT-PCR using the following thermocycling program: 50°C for 30 min, 95°C for 15 min, and 30–35 cycles of 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min.
3. Purify the products using a QIAquick PCR purification kit, or purify from the gel using a Qiagen gel purification kit.

3.4. Functional Analysis of the Selected Clones

Analysis of the selected pools can be done by analyzing the whole pool or single clones. To analyze single clones, digest and ligate the selected pools into a suitable expression vector and transform *E. coli* to obtain individual clones. The expressed proteins can be purified via the His-tag, which can then be used for binding analysis by, for example, surface plasmon resonance (SPR) (Fig. 2). Alternatively, the binding analysis can be done entirely in vitro. The binding can be evaluated by the amount of mRNA recovered after selection which can be quantified using a one-step quantitative RT-PCR kit, in accordance with the manufacturer's instructions. To test for binding of the proteins, in vitro translation can be carried out with (^{35}S)methionine (see Subheading 3.3.3). The radioactive translation mix can then be tested using similar assay setups as described in Subheading 3.3.4. The recovered radioactive proteins are analyzed by liquid scintillation counting or autoradiography.

With ribosome display experiments, not only the mutagenesis and selection process but also the analysis of the selected clones can be carried out entirely in vitro, and thus is an efficient and rapid procedure for protein evolution. With the procedures described above, we successfully enriched variants that showed improved binding in comparison with the parent W17F molecule (Fig. 2). The affinities of sk1, pe1, and pe2 were even higher than the wild type. Our results suggest that deleterious effects caused by mutations in highly conserved residues occurring through divergent evolution may not only be restored but also be improved even further by compensatory mutations.

4. Notes

1. Note that the mutation rate of mutagenic PCR by this method is 8.3×10^{-4}/nucleotide/cycle. Mutation rate may vary with conditions such as cycle number of PCR and amount of DNA template.

2. We compared the differences between two in vitro translation systems, i.e., the S30 extract-based system and the PURE system when used for ribosome display experiments and found that the enrichment factor judged by the band intensity after RT-PCR was markedly better with the PURE system.
3. We confirmed that the protein translation reaction can be done using DNA as a template, indicating that in vitro transcription and translation can be coupled in the PURE system, although the enrichment factors in selection experiments were higher when using mRNA as a template. Therefore, we suggest performing the protein translation reaction using mRNA as a template.
4. If the selection is not specific in later rounds, a "pre-panning step" can also be performed. Aliquots of 100 μL of the stopped translation reaction are first applied to the wells of microtiter plates, which are identical to the ligand-immobilized wells except lacking the ligand (e.g., containing NeutrAvidin and BSA). After incubation for 1 h at 4°C, the supernatant is then transferred to the ligand-immobilized wells. This step can be advantageous to reduce nonspecific binders adhering to the surface.
5. The stringency of selection can be controlled by changing the number of washing steps. For the initial round, it is essential not to lose the binding molecules, and tolerate lower selection stringency. In later rounds, the number of washing cycles and the incubation time should be increased.
6. We tested several types of reverse transcriptase and found that recovered mRNA was best amplified using the Qiagen OneStep RT-PCR kit, which allows both reverse transcription and PCR amplification to take place in a single tube.
7. When using the S30 extract-based in vitro translation system, it was necessary to amplify recovered mRNA with an internal primer (far inside the 5′ and 3′ ends of the mRNA), which was probably due to degradation of the terminal ends of mRNA, and then perform reamplification with nested primers. To obtain the full-length products from the PCR product, recloning into the pRD-n1n2_2 vector was required, including many processes as described under Subheading 3.2. On the other hand, when using the PURE system, one-step RT-PCR using primers located at the 5′ and 3′ ends could yield sufficient amounts of the full-length DNA, which was likely due to little degradation and high recovery of mRNA. Thus, the amplified DNA could be used directly for the subsequent in vitro transcription. In this way, the PURE system enabled us to shorten the time required for the experiments.

Acknowledgments

This research was supported in part by "Special Coordination Funds for Promoting Science and Technology: Yuragi Project" and "Global COE (Centers of Excellence) Program" of the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

References

1. Smith GP (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 228: 1315–1317
2. Roberts RW, Szostak JW (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc Natl Acad Sci U S A 94:12297–12302
3. Hanes J, Pluckthun A (1997) In vitro selection and evolution of functional proteins by using ribosome display. Proc Natl Acad Sci U S A 94:4937–4942
4. Tawfik DS, Griffiths AD (1998) Man-made cell-like compartments for molecular evolution. Nat Biotechnol 16:652–656
5. Hermes JD, Blacklow SC, Knowles JR (1990) Searching sequence space by definably random mutagenesis: improving the catalytic potency of an enzyme. Proc Natl Acad Sci U S A 87: 696–700
6. Lin-Goerke JL, Robbins DJ, Burczak JD (1997) PCR-based random mutagenesis using manganese and reduced dNTP concentration. Biotechniques 23:409–412
7. Zaccolo M, Williams DM, Brown DM, Gherardi E (1996) An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. J Mol Biol 255:589–603
8. Stemmer WP (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370:389–391
9. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. Nat Biotechnol 16:258–261
10. Shimizu Y, Inoue A, Tomari Y, Suzuki T, Yokogawa T, Nishikawa K, Ueda T (2001) Cell-free translation reconstituted with purified components. Nat Biotechnol 19:751–755
11. Villemagne D, Jackson R, Douthwaite JA (2006) Highly efficient ribosome display selection by use of purified components for in vitro translation. J Immunol Methods 313:140–148
12. Matsuura T, Yanagida H, Ushioda J, Urabe I, Yomo T (2007) Nascent chain, mRNA, and ribosome complexes generated by a pure translation system. Biochem Biophys Res Commun 352:372–377
13. Yanagida H, Matsuura T, Yomo T (2008) Compensatory evolution of a WW domain variant lacking the strictly conserved Trp residue. J Mol Evol 66:61–71
14. Bork P, Sudol M (1994) The WW domain: a signalling site in dystrophin? Trends Biochem Sci 19:531–533
15. Chen HI, Sudol M (1995) The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. Proc Natl Acad Sci U S A 92: 7819–7823
16. Macias MJ, Hyvonen M, Baraldi E, Schultz J, Sudol M, Saraste M, Oschkinat H (1996) Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. Nature 382:646–649
17. Koepf EK, Petrassi HM, Ratnaswamy G, Huff ME, Sudol M, Kelly JW (1999) Characterization of the structure and function of W→F WW domain variants: identification of a natively unfolded protein that folds upon ligand binding. Biochemistry 38: 14338–14351
18. Ohashi H, Shimizu Y, Ying BW, Ueda T (2007) Efficient protein selection based on ribosome display system with purified components. Biochem Biophys Res Commun 352:270–276
19. Sudol M (1998) From Src Homology domains to other signaling modules: proposal of the "protein recognition code". Oncogene 17:1469–1474

Part V

Protein Structure and Function Mutagenesis

Chapter 20

Fine-Tuning Enzyme Activity Through Saturation Mutagenesis

Holly H. Hogrefe

Abstract

Codon saturation is a powerful tool for analyzing protein structure–function relationships and fine-tuning enzyme activity. In this technique, one or more key amino acids are randomized by incorporating degenerate codon(s) into a gene of interest in a polymerase-mediated primer extension reaction. Traditional methods for codon saturation mutagenesis are labor-intensive and typically require multiple rounds of PCR and restriction/ligation-based cloning. In contrast, the QuikChange® Multi Site-Directed Mutagenesis kit provides an efficient one-day procedure for incorporating degenerate codons into any double-stranded plasmid DNA template. Originally developed for introducing point mutations at multiple sites, the Multi kit can be readily adapted for performing site-directed saturation mutagenesis of one, two, or three amino acids simultaneously. When coupled with a suitable screening assay, this method simplifies the process of surveying multiple side chain replacements at key locations.

Key words: Saturation mutagenesis, Degenerate primer, Degenerate codon, QuikChange Multi

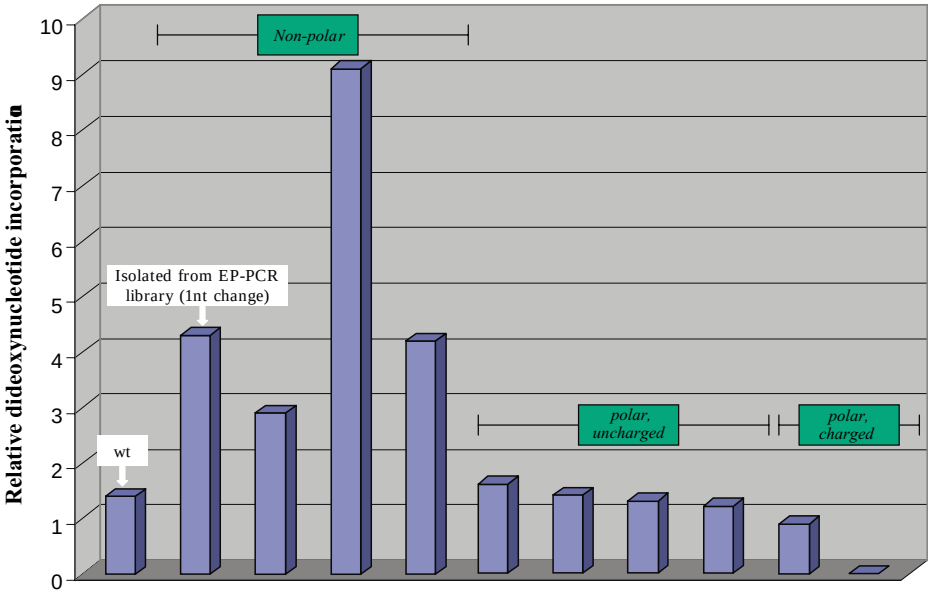
1. Introduction

Site-directed saturation mutagenesis refers to the process of replacing one or more amino acids with all 20 possible side chain replacements (1). By surveying the varying effects of different side chain substitutions one obtains a much more comprehensive analysis of structure–function relationships than can be achieved through single amino acid replacements. In addition to fine-tuning functional attributes, site-saturation provides key insights into molecular features (such as hydrophobicity, size, charge, and aromaticity) that contribute to enzyme activity, substrate/receptor binding, or protein stability. In contrast, beneficial mutations or critical attributes often go undetected when single amino acid

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_20,

replacement strategies, such as scanning alanine mutagenesis or random mutagenesis, are employed (1).

The power of saturation mutagenesis is illustrated in Fig. 1, where a variety of effects including complete loss of enzyme activity, no change in activity, and varying levels of improved utilization of a nucleotide analog were observed, depending on the amino acid replacement at a particular codon. In this example, saturation mutagenesis not only identified DNA polymerase mutants with increased activity (that were missed by random mutagenesis (2)),



Amino acid	P	L	M	I	G	Q	W	S	T	H	R
Codon	CC**G**	C**TG**	**ATG**	**A**TT	**GG**T	C**AG**	**TTG**	**T**CT	**A**CT	C**A**T	C**G**T
Frequency (non-CCT)	6%	6%	13%	6%	19%	13%	6%	6%	6%	31%	6%
polymerase activity	+++	+++	+++	+++	+++	++	++	+++	+++	++	-

Fig. 1. Saturation mutagenesis provides comprehensive analysis of structure–function relationships. Amino acid P410 (CCT codon) in the nucleotide-binding pocket of JDF-3 DNA polymerase was randomized using the QuikChange® Multi site-directed mutagenesis kit. Mutagenesis reactions consisted of 1× QuikChange Multi reaction buffer, 100 ng pBS-JDF-3 (5.2 kb), 100 ng primer 5′ TTTCGTAGTCTCTAC-NNG/T-TCAATCATAATCACC (TriLink Biotechnologies, Inc.; N denotes 25% each G,C,A, and T), dNTP mix, and 2.5 U QuikChange Multi enzyme blend. Reactions were cycled at 95°C for 1 min (1 cycle); 95°C for 1 min, 55°C for 1 min, 65°C for 10 min (30 cycles). After cycling, reaction products were digested with 10 U *Dpn* I for 1 h at 37°C and transformed into XL10-Gold cells as described in Subheading 3.3. The 25 μl Multi reaction produced a total of 1.2×10^4 mutants. Of the 30 clones randomly selected for further analysis, 16 (11 unique) mutants were recovered that encoded 10 different side chain replacements at position 410. Mutant protein samples were prepared, and assayed for DNA polymerase activity (+++, similar to wild-type activity; –, inactive) and dideoxynucleotide incorporation (see Note 6). One mutation (P410R) resulted in complete loss of enzyme activity, while other mutations had minimal or no effect on DNA polymerase activity with natural nucleotides. Four of the mutations (P410L; P410M; P410I; P410G) brought about significant improvements in dideoxynucleotide incorporation. The P410L (CCT → C**T**T) mutation was identified previously when a lambda phage library, created by amplifying the *jdf-3 pol* coding region under error-prone conditions, was screened in a plaque lift assay for increased (α-^{33}P)-ddNTP incorporation (2). The remaining mutants identified by saturation mutagenesis (JDF-3 P410M, P410I, and P410G) were not likely present in the error-prone PCR library at high frequency (1) because P → M, P → I, and P → G mutations require two to three simultaneous mutations in the 410 codon.

but also provided insights into the structural requirements for polymerase–nucleotide interaction.

Traditional methods for saturation mutagenesis are labor-intensive and require two rounds of PCR (to add degenerate-codon primers and assemble fragments by splice overlap extension), followed by restriction digestion, purification, and ligation of inserts (1). In contrast, the QuikChange® Multi Site-Directed Mutagenesis kits provide an efficient, one-day procedure for performing site-directed saturation mutagenesis of one, two, or three amino acids simultaneously (3). Originally developed for introducing point mutations at multiple sites, the Multi kits can be used to introduce degenerate codons into any gene of interest using a simple three-step procedure (Fig. 2). When coupled with a suitable screening assay, this method simplifies the process of evaluating multiple side chain replacements at key locations.

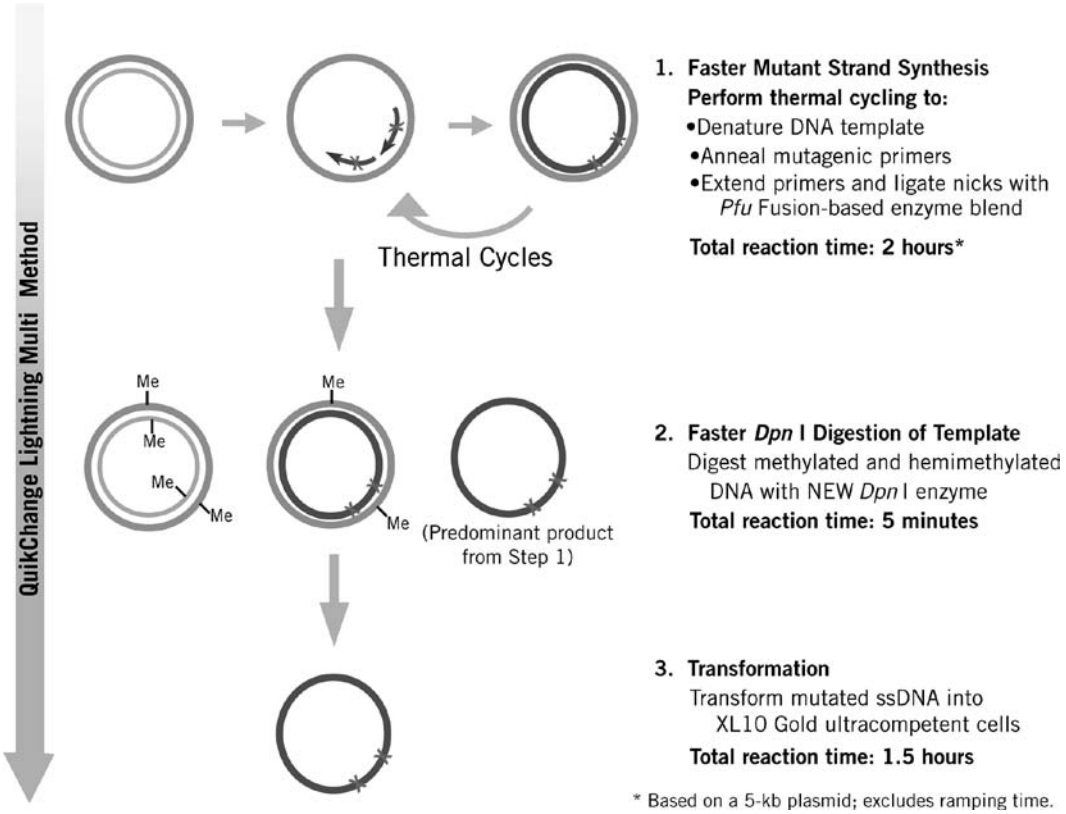


Fig. 2. Outline of the Three-Step QuikChange Lightning Multi Protocol for Saturation Mutagenesis. In the QuikChange Lightning Multi Site-Directed Mutagenesis Kit, degenerate codons (represented as ×) are introduced by annealing one or more mixed-base primers to either the sense or antisense strand of a plasmid DNA template. During thermal cycling, the mutagenic primers are extended and nicks are sealed with the Multi enzyme blend, containing *PfuUltra* II DNA polymerase and other enzymes. Reaction products are then digested with *Dpn* I (reformulated in the Lightning kit to allow shorter digestion time compared to the original kit) and transformed into high competency cells. Finally, clones are selected and screened by DNA sequencing to identify unique mutants. Alternatively, the mutant collection is screened to identify clones with new functional attributes. *Reprinted by permission of Agilent Technologies: Stratagene Products Division.*

A detailed protocol is provided below for designing and incorporating degenerate-codon primers into plasmid DNA templates using components of the QuikChange® Lightning Multi Site-Directed Mutagenesis kit. Compared to the original kit, the updated Lightning kit features faster enzymes and protocols for thermal cycling and mutant enrichment. Troubleshooting tips are provided for increasing mutation efficiency and library size, which typically averages 10^4 to 10^5 mutants per reaction (3). Finally, recommendations are included for randomizing multiple codons and for screening mutant collections for altered activity.

2. Materials

2.1. Mutagenesis Primers

1. *Degenerate codon(s)*. Degenerate codons are introduced with mixed-base oligonucleotides that are synthesized using a selected mixture of standard bases at three positions (e.g., NNN, where N refers to an equimolar mixture of G:C:T:A). When performing saturation mutagenesis, degenerate codon(s) should be positioned as close to the middle of the mutagenic primer as possible, with approximately 10–15 bases of template-complementary sequence on either side. To minimize redundancy, degenerate codons may be composed of NNG/T, where the wobble position is synthesized with an equimolar mixture of G and T. All 20 amino acids are encoded by NNG/T codons, and the distribution of mutation types in a library will vary with codon redundancy, from 3% for the least-represented amino acids (e.g., encoded by one NNG/T codon; N, D, C, E, Q, H, I, K, M, F, W, Y; frequency: $¼ \times ¼ \times ½ = 1/32$) to 9% for highly represented amino acids (encoded by 3 NNG/T codons; R, S, L). Similarly, stop codons (TAG) will be introduced at an expected frequency of 3%.
2. *General primer design*. With the Multi kits, one mutagenic primer is designed for each site to be mutagenized. Ideally, primers are 25–45 bases in length, with a melting temperature (T_m) of 75°C or higher. To estimate T_m of primers, the following formula can be used: $T_m = 81.5 + 0.41(\%GC) - 675/N - \%\ \text{mismatch}$, where N is primer length in bases and $\%GC$ and $\%\ mismatch$ values are whole numbers. When designing primers with degenerate codons, it is important to match the T_ms of the two primer sequences flanking the degenerate codon. If one side has a lower T_m due to lower GC content, primer length can be extended (by up to five nucleotides) on that side to increase the T_m. Whenever possible, primers should have a minimum GC content of 40% and one or more C or G nucleotides at the 3′-end.

3. *Primer quality and purity*. Mixed-base oligonucleotides should be ordered from a reputable vendor. PAGE purification is recommended to eliminate failure sequences (incomplete primer synthesis) and minimize unintended errors in the regions flanking the degenerate codon(s). The QuikChange Multi kits do not require 5′-phosphorylation of mutagenic primers.
4. *Multi primer design*. Although up to five primers can be incorporated with the Multi kit, the number of codons for simultaneous saturation should be limited to three to allow adequate representation of all mutation types (see discussion under Subheading 3.4). Amino acids located in close proximity (e.g., separated by <6 amino acids) can be randomized by incorporating two or three degenerate (NNG/T) codons in a single oligonucleotide. Amino acids located farther apart can be randomized simultaneously by introducing degenerate codons in separate primers. When using two or three primers, all of the primers should be designed to anneal to the same strand of the template plasmid (either to the sense or antisense strand). Primers should not overlap but can be designed to anneal immediately adjacent to each other. In general, primers located immediately adjacent to each other are incorporated as efficiently as primers that anneal farther apart.

When randomizing two or three codons with separate primers, one should consider the mutagenesis efficiency of multiple-primer reactions. The probability of incorporating a mutagenic primer at every site decreases with increasing primer number. For example, the average mutation efficiencies observed for the 4-kb kit control template (using multiple primers and primer combinations) varied from 91% (1–2 primers) to 58–60% (3–4 primers) to 30% (5 primers) (4). When randomizing two amino acids with separate primers, most but not all clones contain mutations at both codons. However, when three codons are saturated with separate primers, a substantial portion of the library consists of single and double mutants (3).

2.2. Plasmid DNA Template

1. Plasmid DNA templates must be isolated from a dam$^+$ *Escherichia coli* strain to allow *Dpn* I digestion (enrichment of mutant DNA). The majority of commonly used *E. coli* strains are dam$^+$.
2. Plasmid DNA template can be prepared by standard miniprep protocols or by cesium chloride gradient centrifugation.
3. Kit-provided XL10-Gold® Ultracompetent cells are resistant to tetracycline. If the plasmid DNA template contains only the tet^R resistance marker, the target gene sequence should be subcloned into an alternative vector backbone. Alternatively, a tetracycline-sensitive strain of competent cells must be used.

2.3. Site-Directed Mutagenesis Reagents

1. Degenerate codon primers are incorporated using the QuikChange Multi Site-Directed Mutagenesis kit or the newer QuikChange Lightning Multi Site-Directed Mutagenesis kit (Agilent Technologies: Stratagene Products Division). Both kits include a multienzyme blend, 10× reaction buffer, QuikSolution, dNTP mix, *Dpn* I, XL10-Gold cells, β-mercaptoethanol mix, and kit controls. The original kit employs *PfuTurbo*® DNA polymerase, while the Lightning kit uses a higher-processivity *Pfu* fusion enzyme and unique reaction buffer to allow shorter thermal cycling times. The multienzyme blend, reaction buffer, and *Dpn* I enzyme have been specifically optimized for use with the recommended cycling and digestion protocols for each kit. The remaining components (QuikSolution, dNTP mix, XL10-Gold cells, and β-mercaptoethanol mix) can be used interchangeably between the two versions of the multisite kit (see Note 7).
2. The XL10-Gold cells included with the kit are very sensitive to small variations in temperature (see Note 1b). To ensure highest efficiency, the cells should be stored at –80°C immediately upon receipt and kept on ice as recommended in the transformation protocol (see Subheading 3.3).
3. Although not provided, Falcon polypropylene 14 ml round-bottom tubes are highly recommended for the transformation step (BD Biosciences, location). Not only are they resistant to β-mercaptoethanol, but the duration and temperature of the heat-pulse step (30 s at 42°C) have also been optimized for the thickness and shape of these tubes.
4. Additional components not included in the Multi kits include NZY$^+$ culture media and LB agar plates containing the appropriate antibiotic. After transformation and overnight growth, colonies are picked for DNA template preparation, followed by DNA sequencing or activity screening.
5. Multi kits include control primers and template, which can be run alongside sample reactions. When employing controls, mutation efficiency can be monitored by blue-white color screening using LB-ampicillin agar plates that are prepared in advance and contain X-gal (5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside) and IPTG (isopropyl-1-thio-β-d-galactopyranoside) (see Note 1a and c).

3. Methods

The following protocol can be followed using components from the QuikChange Lightning Multi Site-Directed Mutagenesis kit. Please refer to thermal cycling and *Dpn* I digestion protocols

provided in the QuikChange Multi Site-Directed Mutagenesis kit when using the original kit components.

3.1. Mutant Strand Synthesis Reaction (Thermal Cycling)

1. Assemble mutagenesis reactions by adding components in the order listed below. Vortex gently.

Reaction component	Templates ≤ 5 kb	Templates > 5 kb
10× QuikChange Lightning Multi buffer	2.5 µl	2.5 µl
Double-distilled water	X µl to final volume of 25 µl	X µl to final volume of 25 µl
QuikSolution	–	0–0.75 µl (titrate for each template)
Plasmid DNA template	50 ng	100 ng
Degenerate-codon primer(s)	100 ng each primer[a]	100 ng each primer[a]
dNTP mix	1 µl	1 µl
QuikChange Lightning Multi enzyme blend	1 µl	1 µl

[a]When using two or three primers, primers should be added in approximately equimolar amounts

2. Cycle the mutagenesis reactions using the cycling parameters outlined in the table below.

Segment	Cycles	Temperature	Time
1	1	95°C	2 min
2	30	95°C	20 s
		55°C	30 s
		65°C	30 s/kb of plasmid length
3	1	65°C	5 min

3. Following temperature cycling, place QuikChange Multi-reactions on ice.

3.2. Dpn I Digestion of Amplification Products

1. Add 1 µl *Dpn* I restriction enzyme to each amplification reaction, and mix gently.
2. Briefly spin down the reaction mixtures in a microcentrifuge, and then incubate reactions at 37°C for 5 min to digest the parental (nonmutated) DNA template. If using *Dpn* I from the original QuikChange Multi kit, extend the incubation time to 60 min.

3.3. Transformation of QuikChange Multi reactions

1. Gently thaw the XL10-Gold cells on ice. For each mutagenesis reaction, transfer 45 µl of cells to a *prechilled* 14-ml BD Falcon polypropylene round-bottom tube.
2. Add 2 µl of kit-provided β-mercaptoethanol to each tube, and swirl gently to mix. Incubate the cells on ice for 10 min, swirling gently every 2 min.
3. Transfer 1.5 µl of *Dpn* I-digested DNA from each mutagenesis reaction to a separate aliquot of cells. Swirl the transformation reactions gently to mix, and incubate the reactions on ice for 30 min.
4. Store the remaining *Dpn* I-digested DNA at –20°C. It can be used to transform the entire library once titer and mutation efficiency have been determined.
5. Preheat NZY$^+$ broth in a 42°C water bath for use in step 8.
6. Heat-pulse the tubes for 30 s in a 42°C water bath (do not exceed 42°C). The duration of the heat pulse is critical for obtaining the highest efficiencies.
7. Incubate the tubes on ice for 2 min.
8. Add 0.5 ml of preheated 42°C NZY$^+$ broth to each tube, and incubate the tubes at 37°C for 1 h while shaking at 225–250 rpm.
9. Plate 1 µl, 10 µl, and 100 µl of each transformation reaction on agar plates containing the appropriate antibiotic for the plasmid DNA template. Incubate the transformation plates at 37°C for >16 h. Between 10 and 1,000 colonies are expected.
10. Library titer can be calculated using the following formula: $(\#\ colonies) \times (culture\ volume/volume\ plated) \times (Multi\ reaction\ volume/volume\ transformed)$. QuikChange Multi kits typically yield approximately 10^4 to 10^5 clones per 25 µl reaction using the recommended competent cells and transformation protocol (multiple transformations of 1.5 µl portions of *Dpn* I-digested DNA; see Note 2).

3.4. Screening

When analyzing single-codon saturation libraries, the minimum number of clones containing all possible mutations is dictated by the frequency of the least-represented mutants (e.g., encoded by only one codon), the efficiency of QuikChange Multi mutagenesis (typically >90% for one-primer mutagenesis), and the statistical distribution of mutation types. For NNG/T codons, the frequency of the least represented mutants can be calculated as: $(¼ \times ¼ \times ½) = 1/32$. Assuming 100% mutation efficiency, there is >95% likelihood of observing all possible mutants in a random sampling of approximately 100 clones ($0.95 = 1 - (1 - f)^n$, where f = frequency of the least represented mutants and n = # clones screened) (5). When incorporating two random NNG/T codons,

the frequency of the least-represented double mutants is 1/1024 ($(1/4 \times 1/4 \times 1/2)^2$), and to ensure representation of all possible double mutants, approximately 3,000 clones should be screened (>95% confidence). Similarly, when randomizing three amino acids, there is >95% likelihood of observing all possible triple mutants in a random sampling of approximately 10^5 clones ($f = 1/32{,}768$). To ensure fully representative libraries, no more than three codons should be randomized simultaneously due to the practical limits of bacterial transformation efficiency ($<10^6$ clones per 25 µl Multi reaction). Prior to screening, actual mutation efficiency should be determined by sequencing 10 or more randomly selected clones. Mutation efficiency will vary with primer sequence, number of primers, and the size and complexity of the plasmid DNA template (see Note 3). The minimal number of clones to screen (to ensure a statistically relevant sample size) should be adjusted according to actual mutation efficiency.

With single-codon saturation, random colonies can be picked for template preparation and DNA sequencing. As discussed above, all possible variants (19 different side chain replacements) should be recovered in a random sampling of approximately 100 colonies. In the absence of a suitable screening assay, unique mutants can be expressed, purified, and assayed to determine if the target amino acid plays a role in function (e.g., mutants show altered activity compared to the wild type) or structure (e.g., low expression and/or recovery of soluble mutant protein). When two or three codons are randomized, libraries should be assayed with an appropriate activity screen (see Note 5). As discussed above, all double or triple mutants should be represented in a random sampling of at least 3,000 or 100,000 clones, respectively (or higher, depending on the actual mutation efficiency achieved). Library titer can be scaled up as needed by performing replicate QuikChange Multi reactions and multiple transformations (see Note 4).

4. Notes

1. Kit controls
   (a) The mutagenesis kit control can be used to troubleshoot problems with low colony number or mutation efficiency. The control template and primers test the efficiency of simultaneous site-directed mutagenesis at three independent sites. The 4 kb control template is derived from pBluescript II SK$^-$ phagemid, which encodes the first 146 amino acids of β-galactosidase (*lacZ* gene product) and produces blue colonies when XL10-Gold transformants are plated on agar media containing X-gal (a chromogenic

substrate of β-galactosidase) and the inducer IPTG. The control template has been modified to introduce stop codons at three positions in the *lacZ* coding region, and XL10-Gold transformants appear white on X-gal/IPTG agar media. The control primer mix contains three primers, each of which reverts one stop codon to wild type. Restoration of active β-galactosidase requires all three reversion events to occur in the same molecule. To perform the kit control, a thermal cycling reaction is performed with the control template and primer mix. After *Dpn* I digestion and transformation, mutation efficiency is determined by blue-white color screening.

The mutagenesis control reaction is assembled by adding components in the order shown.

Reaction component	Volume added
10× QuikChange Lightning Multi buffer	2.5 µl
Double-distilled H_2O	18.5 µl
QuikSolution	–
QuikChange Multi control template	1 µl
QuikChange Multi control primer mix	1 µl
dNTP mix	1 µl
QuikChange Lightning Multi enzyme blend	1 µl

After mixing gently, the control reaction is cycled as described above (**see** step 2 under Subheading 3.1) using a 2.5-min extension time. *Dpn* I digestion and transformation are performed as described under Subheadings 3.2 and 3.3, except that transformations are plated on LB-ampicillin agar plates that have been prepared with 80 µg/ml X-gal and 20 mM IPTG. Alternatively, 100 µl of 10 mM IPTG (in sterile water) and 100 µl of 2% X-gal (in DMF) can be spread on LB agar plates 30 min prior to plating the control transformation reactions. When kit components and reaction protocols are working properly, anywhere from 50 to 800 colonies are expected from plating 10 µl of the control transformation. Greater than 50% of the clones should contain all three mutations and appear as blue colonies (β-gal$^+$) on X-gal/IPTG plates.

(b) When colony numbers are lower than expected for the mutagenesis control reaction (see Note 1a), transformation efficiency of the ultracompetent cells can be verified using the kit-provided pUC18 control plasmid. One microliter (1 µl) of pUC18 DNA (diluted to 0.01 ng/µl)

can be added to a separate 45 μl aliquot of XL10-Gold cells. Transformation is carried out as described under Subheading 3.3. When 5 μl of culture is plated on LB–ampicillin agar plates containing X-gal and IPTG, more than 100 colonies should be observed (>10^9 cfu/μg), and at least 98% should exhibit the blue phenotype. If colony number is lower than expected with pUC18, efficiency losses are most likely due to improper procedures for storing and handling XL10-Gold cells (see Subheadings 2.3, item 2**,** and 3.3).

(c) When the transformation efficiency is lower than expected for the mutagenesis control reaction, but within the expected range of >10^9 cfu/μg pUC18 (see Note 1a and b), cycling parameters may require further optimization. Alternatively, one or more kit components may have lost activity due to improper storage or repeated freezing and thawing (e.g., dNTP mix). When the mutagenesis control reaction is giving less than the expected mutagenesis efficiency (<50% blue colonies on X-gal/IPTG plates; see Note 1a), ensure that the agar plates were prepared correctly and incubated for at least 16 h at 37°C for better visualization of the blue (β-gal$^+$) phenotype.

2. When transformation and mutagenesis controls are giving the expected results (see Note 1), but sample colony number is lower than expected (<10^4 per 25 μl reaction), the following steps can be taken to increase the number of colonies recovered with experimental primer-template systems.

   (a) Ensure that an adequate amount of DNA template is used in mutagenesis reactions. Visualize the DNA template on a gel to verify the quantity and quality.

   (b) Ensure that sufficient primer is used in the mutagenesis reaction by measuring optical density at 260 nm. Visualize primers on a gel to verify quality.

   (c) Titrate the amount of QuikSolution in increments from 0 to 1.5 μl.

   (d) Transformation efficiency is expected to decrease as the plasmid size increases. Mutagenesis of longer templates typically requires optimization of template amount, which can be titrated in increments up to 300 ng.

   (e) To increase the number of colonies recovered per transformation reaction, the amount of *Dpn* I-treated DNA can be increased to 4 μl. Alternatively, *Dpn* I-digested DNA can be ethanol-precipitated, and after resuspension, the entire DNA sample can be transformed.

3. When the mutagenesis control is giving the expected efficiency (>50% blue colonies on X-gal/IPTG plates; see Note 1),

but sample reactions exhibit low mutation efficiency, the following steps can be taken.

(a) Repeat *Dpn* I digestion, especially if more than the recommended amount of DNA template was used. If mineral oil was used during thermal cycling, make sure that *Dpn* I enzyme is added below the mineral oil overlay and that the reaction is properly mixed. Verify that the template DNA was isolated from a *dam*$^+$ *E. coli* strain.

(b) Ensure that sufficient primer was used in the mutagenesis reaction by measuring optical density at 260 nm. Visualize the primer on a gel to verify quality. Titrate the amount of primer added to the thermal cycling reaction.

(c) If mutagenesis efficiency remains low (<50%) after addressing other possibilities, redesign the mutagenic primer(s) to bind to the opposite strand of the plasmid. In most cases, primers binding to either strand will be incorporated with equal efficiency. However, certain secondary structures or other features may influence the efficiency of the mutagenesis reaction.

(d) For multiprimer reactions, mutation efficiency at each site can be adjusted by varying primer concentrations. The appropriate amounts of each primer required to achieve the desired mutation efficiencies must be determined empirically for each primer-template system.

4. When the goal is to screen large numbers of mutant clones (e.g., 2- or 3-codon saturation libraries), libraries can be scaled up as needed by performing replicate QuikChange Multi reactions and multiple transformations. For each reaction, up to 4 µl of *Dpn* I-digested DNA can be transformed into separate 45 µl-aliquots of XL10-Gold cells. Although mutagenesis reactions can be scaled up to 50 µl (2×) or 100 µl (4×), thermal cycling conditions may require further optimization to ensure sufficient heat transfer in larger reaction volumes (e.g., increasing the duration of denaturation and extension steps). Higher colony numbers may also be achieved by using electroporation-competent cells (e.g., ElectroTen-Blue electroporation competent cells; Agilent). *Dpn* I-digested DNA should be purified to remove salts prior to electroporation.

5. The kit-provided XL10-Gold cells provide the highest transformation efficiencies possible for ligated DNA and large plasmids. Although ideal for constructing highly representative plasmid libraries, XL10-Gold cells may not provide sufficient protein expression for activity screening. The appropriate number of XL10-Gold colonies can be scraped from plates into media, and grown overnight.

Supercoiled plasmid DNA can then be prepared and used to transform a suitable high-expression bacterial strain (e.g., BL21-CodonPlus® or ArticExpress™ Competent cells). Alternatively, randomized gene sequences can be cloned into an appropriate mammalian expression vector (e.g., StrataClone™ Mammalian Expression system).

6. Protocols for preparing bacterial extracts and screening thermostable DNA polymerases have been published elsewhere (2, 6). In Fig. 1, DNA polymerase activity was measured by monitoring the incorporation of radiolabelled nucleotide (200 μM each dNTP, where 2.5% TTP is ^{3}H-TTP) into activated calf thymus DNA. Relative ddNTP incorporation was assayed by measuring ^{3}H-TTP incorporation in the absence and presence of 20% ddATP, ddCTP, and ddCTP. The tendency to incorporate ddNTPs is proportional to the decrease in counts incorporated in the presence of nucleotide analog inhibitor. Percent activity in the presence of ddNTPs was determined for wild-type and mutant JDF-3 DNA polymerases, and relative ddNTP incorporation was expressed as: (% activity exhibited by wild-type) ÷ (% activity exhibited by mutant).
7. In the absence of a Multi site-directed mutagenesis kit, the single-site QuikChange kit can also be used to incorporate degenerate codon primers. In one published report (7), two fully complementary oligonucleotides were designed with an NNN codon in the center of each primer. The primer pairs were incorporated using a QuikChange-like protocol. The authors reported mutation efficiencies over 95% and library sizes of 1–7.5×10^{3} clones per 25 μl reaction.

References

1. Miyazaki K, Arnold FH (1999) Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. J Mol Evol 49:716–720
2. Arezi B, Hansen CJ, Hogrefe HH (2002) Efficient and high fidelity incorporation of dye-terminators by a novel archaeal DNA polymerase mutant. J Mol Biol 322:719–729
3. Hogrefe HH, Cline J, Youngblood GL, Allen RM (2002) Creating randomized amino acid libraries with the QuikChange Multi site-directed mutagenesis kit. Biotechniques 33:1158–1165
4. Newsletter S (2001) Generate multiple mutations quickly and efficiently. Strategies 14(3):74–75
5. Jeltsch A, Lanio T, Jeltsch A, Lanio T (2002) Site-directed mutagenesis by polymerase chain reaction. In: Braman J (ed) In vitro mutagenesis protocols, 2nd edn, Methods in molecular biology. Humana, Totowa, NJ
6. Hogrefe HH, Cline J, Lovejoy AE, Nielson KB (2001) DNA polymerases from hyperthermophiles. Methods Enzymol 334:91–116
7. Steffens DL, Williams JGK (2007) Efficient site-directed saturation mutagenesis using degenerate oligonucleotides. J Biomol Tech 18:147–149

Chapter 21

Characterization of Structural Determinants of Type 1 Corticotropin Releasing Hormone (CRH) Receptor Signalling Properties

Danijela Markovic and Dimitris K. Grammatopoulos

Abstract

Mammalian adaptation to stressful stimuli requires activation of the type 1 corticotropin releasing hormone (CRH) receptor (CRH-R1), a 415 aminoacid protein that belongs to the large superfamily of 7 transmembrane domain receptors that relay signals across cells through activation of G-proteins. CRH-R1 expression and activity is regulated at the gene level by mRNA alternative splicing that results in a number of CRH-R1 variants. This process can generate putative CRH-R1 receptor variants with distinct structural and signaling properties; their study can provide important insights about the structural determinants of CRH-R1 functional characteristics. Using site-directed mutagenesis by overlap extension polymerase chain reaction (OE-PCR), we investigated the structure–function relationships of a CRH-R1 variant with an extended 1st intracellular loop (IC1) (CRH-R1β), a sequence modification that impairs signaling activity (such as cAMP production and MAPK activation). We identified a penta-aminoacid cassette within the 29-aminoacid insert of CRH-R1β rich in positive charged aminoacids (F^{170}-R^{174}), as an important structural determinant for the impaired cAMP response.

Key words: Site-directed mutagenesis, OE-PCR, CRH-R1, cAMP, Transient transfection

1. Introduction

The mechanisms controlling response to stressful stimuli are highly conserved across mammalian species, a reflection of their pivotal role in survival and evolution. These adaptive responses to stress are orchestrated by corticotropin releasing hormone (CRH), a 41-aminoacid hypothalamic peptide, which regulates the secretion of adrenocorticotropin from the anterior pituitary and exerts a wide spectrum of actions in the central nervous system (CNS) and the periphery. The complex process of stress

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_21,

adaptation is fine-tuned by several CRH-related peptides, the urocortins that exert complementary or sometimes contrasting actions (1).

This family of stress peptides exert their action through activation of two types of heptahelical G-protein coupled receptors (GPCR), each with distinct patterns of tissue distribution and pharmacology (2). Gene knockout animal models have demonstrated that the type 1 CRH-R receptor is principally responsible for mediating the CRH stress response. The mammalian CRH-R1 gene contains 13 exons that encode a 415 amino acid receptor protein (CRH-R1α). Interestingly, in humans, the CRH-R1 gene has undergone a modification during evolution by acquiring (or retaining) an additional exon. The complete human gene product is a 444-amino acid receptor variant, termed CRH-R1β, which is identical to the CRH-R1α, except for a 29 amino acid insert in the first intracellular loop (IC1) (3) (Fig. 1). This sequence modification has important functional consequences since this receptor variant exhibits impaired agonist binding and signaling properties (4). Although the physiological function of CRH-R1β is unknown, it can provide a useful model for studying the structural determinants of the CRH-R1 functional characteristics.

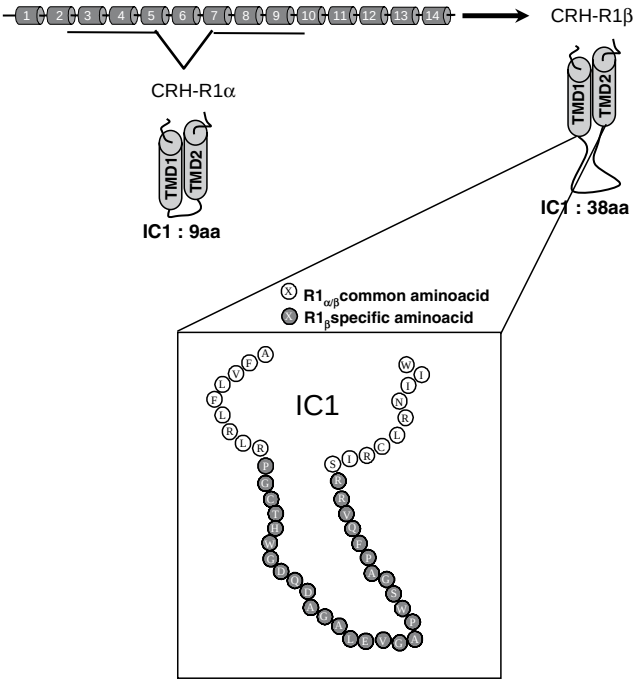


Fig. 1. Schematic representation of CRH-R1 gene organization and resulting proteins. The human CRH-R1 gene has 14 exons that encode a 444 amino acid, 7 transmembrane protein, termed CRH-R1β. Omission of exon 6 results in translation of a 415 amino acid protein, CRH-R1α. The difference between two receptor variants is within the first intracellular loop (IC1): CRH-R1α has only 9 amino acids in IC1, while CRH-R1β has additional 29 amino acids.

These studies require generation of mutant receptors with sequence modifications that result in either deletions or alanine substitutions of single or multiple aminoacids of interest. This can be achieved by site overlap extension polymerase chain reaction (OE-PCR) by usage of oligonucleotide primers that introduce a desirable mutation, resulting in amplification of DNA molecules with single or multiple nucleotide insertions, substitutions, or deletions. This technique is a powerful tool for the rapid generation of a large number of mutant receptors that can be further analyzed for their signaling properties. Using this technique and corticotropin releasing hormone receptor CRH-R1β DNA as a template, a series of mutant receptors were created containing sequential deletions (Δ) or alanine substitutions (A) of cassettes of five or six amino acids. Wild type and mutant receptors were transiently expressed in human embryonic kidney (HEK293) cells and tested for their abilities to increase intracellular cAMP levels. These experiments allowed the identification of a penta-amino acid cassette (F^{170}-R^{174}) within the 29-aminoacid insert of CRH-R1β, which contains multiple positive charged amino acids, and appears to be an important structural determinant of CRH-R1β-impaired cAMP response (5).

2. Materials

2.1. Molecular Biology Chemicals and Reagents (All Chemicals, If Not Stated Otherwise Were Purchased from Sigma). See Note 1

1. Primers (HPLC-purified): (Invitrogen) dissolved in molecular biology grade water to a working concentration of 10 μM, aliquoted, and stored at –20°C.
2. *Pfu* DNA polymerase and accompanying buffer: (Promega).
3. Deoxynucleotide triphosphate mix: 10 mM each dNTP (Bioline), aliquoted, and stored at –20°C.
4. DNA hyperladder I: (Bioline).
5. Agarose: (Sigma).
6. Ethidium bromide solution: 10 mg/ml (toxic and carcinogenic substance. Gloves and lab coat must be worn at all times. Care is required to avoid contact with skin and eyes. Disposal should be according to institutional protocols).
7. 5× DNA loading buffer: 1 mM Tris–HCl, pH 7.5, 40% glycerol (v/v), 0.1% bromphenol blue (w/v).
8. 10× TBE buffer stock solution: 0.9 M Tris-base, 0.9 M boric acid, 10 mM ethylenediamine tetraacetic acid (EDTA), pH 8.2. Dilution to 1× achieved with deionised water.

9. Recombinant CRH-R1α and -R1β: subcloned in pcDNA3.1(–) plasmid (Invitrogen) containing the inserted ampicillin resistance gene.
10. Restriction enzymes: *Hin*dIII, *Eco*RI, and accompanying buffers: (Fermentas).
11. LigaFast™ Rapid DNA Ligation System: (Promega).
12. QIAquick gel extraction kit: (Qiagen).
13. Mini and Midi plasmid preparation kit: (Qiagen).

2.2. Microbiology Chemical and Reagents (All Chemicals, If Not Stated Otherwise, Were Purchased from Sigma)

1. Ampicillin (Gibco) resuspended in sterile water to stock concentration of 100 mg/ml.
2. L-Broth: 10% tryptone (w/v), 5% yeast extract (w/v), 10% NaCl (w/v) in distilled water, autoclaved at 120°C for 15 min; ampicilline (Gibco) is added to a final concentration of 100 µg/ml, when temperature reaches approximately 40°C.
3. LB-agar: 1.5% agar (w/v) in L-Broth is autoclaved at 120°C for 15 min; when temperature drops to approximately 40°C, ampicilline (Gibco) is added to a final concentration of 100 µg/ml. The mixture is poured in 10 cm plastic Petri dishes under aseptic conditions.

2.3. Cell Culture and Transfection (All Chemicals If Not Stated Otherwise, Were Purchased from Gibco/Invitrogen)

1. HEK293-cells (human embryonic kidney cells), purchased from the European Collection of Animal Cell Culture (ECACC).
2. Normal growth medium: Dulbecco's Modified Eagle's Medium (DMEM) with Glutamax, supplemented with 10% fetal calf serum (FCS), 100 U/ml penicillin, and 100 µg/ml streptomycin.
3. Solution of trypsin (0.25%) and EDTA (1 mM) in Han's salt balanced solution.
4. Transfection medium: OptiMEM with Glutamax: (Gibco).
5. Lipofectamine2000, transfection reagent: (Gibco).
6. Poly-D-lysine (Sigma) is dissolved at 100 µg/ml in sterile distilled water and stored in single use aliquots at –20°C.
7. CRH (Bachem) is dissolved at 100 µM in 10% acetic acid and stored in single use aliquots at –80°C.

2.4. Equipment Required

1. Thermocycler PCR machine: (Eppendorf).
2. PCR tubes (see Note 2).
3. Gel electrophoresis apparatus with a power source, combs, and gel tray to pour agarose gels.
4. UV light source (UV light is harmful to skin and eyes. Need to wear protective goggles).

5. Spectrophotometer.
6. NanoDrop spectrophotometer (Thermo Scientific).
7. Bunsen burner for aseptic conditions when handling microbiology work.
8. Cell culture flow cabinet.
9. Incubator for bacterial growth.
10. Incubator supplied with CO_2 for cell culture growth.

3. Methods

3.1. General Considerations

Overlap extension PCR (OE-PCR) was employed to introduce desirable mutations (see Note 3). Primers that have mismatched nucleotide sequences to DNA template were used to introduce a mutation through PCR amplification as originally described in 1988 by Higuchi (6). OE-PCR utilizes four primers in two sequential PCR rounds to introduce an internal mutation into the target sequence (Fig. 2; (7)). In order to amplify separate, overlapping sequences of the template DNA, two independent PCR reactions were run simultaneously, using primers F (forward primer containing a sequence for a restriction site adjacent to the 5′-end of the template sequence) and RM (Reverse primer that contains the desired Mutation) in the reaction A, and primer R (reverse primer containing a sequence of the template 3′-end and a sequence for another restriction site) and FM (Forward primer that contains the desired Mutation) in the reaction B. Primers RM and FM have overlapping homologous regions containing the required mutation and are reverse complement sequences. PCR products from these reactions served as templates for the PCR reaction C, which utilizes primers F and R, the outer, conserved primers, resulting in generation of full length cDNA with the desired mutation (Fig. 2). The mutated cDNA was ligated in mammalian expression vector pcDNA3.1(−) and transformed into *E. coli*. Following maxi preparation of the plasmid, HEK293-cells were transiently transfected with it, and various biochemical and imaging techniques were employed to analyze the signaling properties and cellular localization of the mutated receptors.

An overview of CRH-R1β mutant receptors containing deletions (Δ) or alanine substitutions (A) of 5–6 amino acids is presented in Fig. 3.

3.2. Primer Design

See Note 4.

1. The outer primers (F and R in Fig. 2) were designed to amplify cDNA from the very 5′- and 3′-ends of sequence. Restriction sites were added to the primers in order to clone the insert into

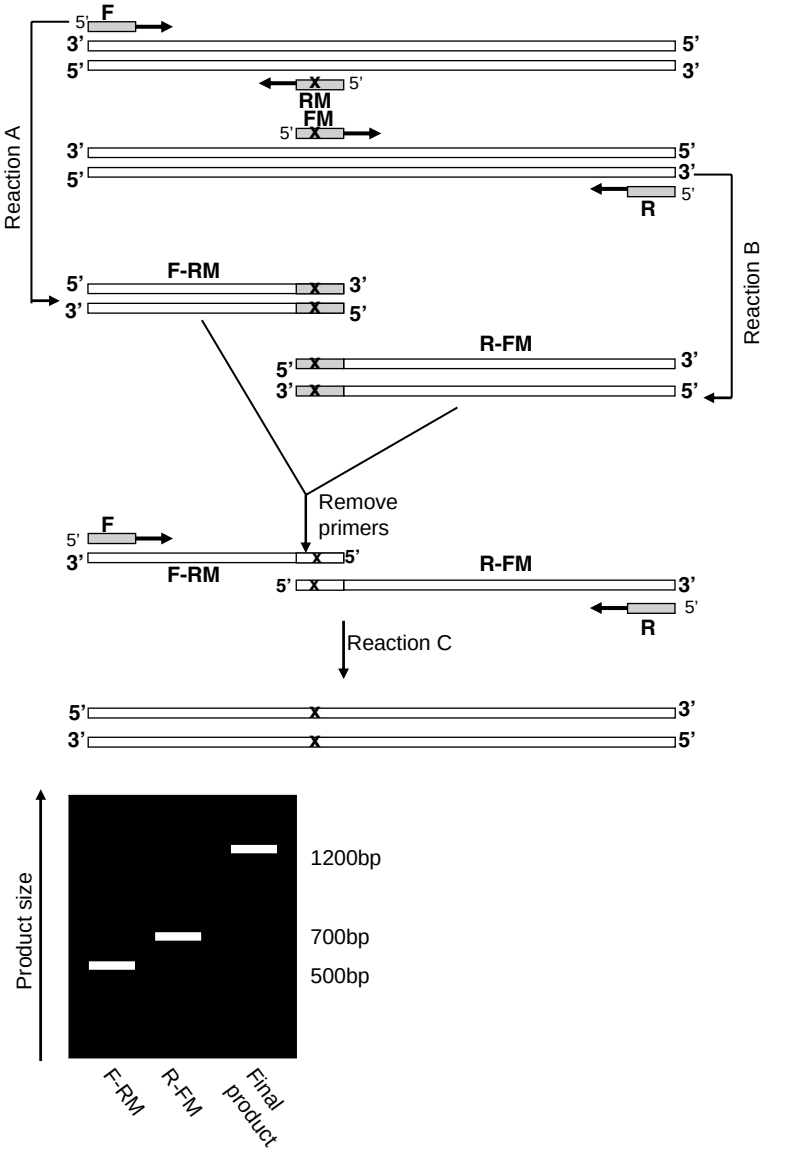
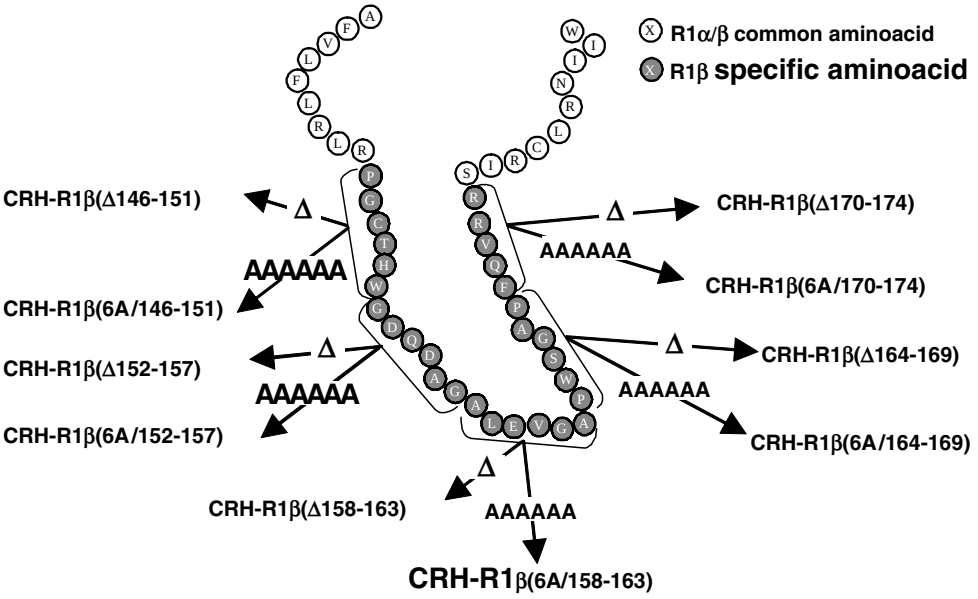


Fig. 2. Schematic representation of OE-PCR. Desired mutations (deletion, insertion, or substitution) were introduced with three separate PCR reactions. RM-reverse mutation containing primer and outer forward primer (F) were used in the reaction A, giving a hypothetical PCR product of 500 bp. In the reaction B, FM-forward mutation containing primer and outer reversed primer (R) were used to produce a product of 700 bp. The PCR products were separated on a 1.2% agarose gel and purified. The resulting products were used as templates for the PCR reaction C with F and R primers. The product of reaction C was a 1,200 bp nucleotide sequence that contains the desired mutations.

pcDNA(−) vector. Primer F was designed to introduce a novel *Eco*RI restriction site, and primer R was designed to introduce a novel *Hin*dIII restriction site (as shown in Fig. 4, *bold-italic letters represent nucleotides for the restriction sites*). For the design



CRH-R1α	^{145}R. . . . . SI^{147}
CRH-R1β	^{145}R PGCTHWGDQADGALEVGAPWSGAPFQVRR SI^{176}
CRH-R1β(6A/146-151)	^{145}R AAAAAAGDQADGALEVGAPWSGAPFQVRR SI^{176}
CRH-R1β(Δ146-151)	^{145}R . . . . . GDQADGALEVGAPWSGAPFQVRR SI^{171}
CRH-R1β(6A/152-157)	^{145}R PGCTHWAAAAAAALEVGAPWSGAPFQVRR SI^{176}
CRH-R1β(Δ152-157)	^{145}R PGCTHW . . . . . ALEVGAPWSGAPFQVRR SI^{171}
CRH-R1β(6A/158-163)	^{145}R PGCTHWGDQADGAAAAAAPWSGAPFQVRR SI^{176}
CRH-R1β(Δ158-163)	^{145}R PGCTHWGDQADG . . . . . PWSGAPFQVRR SI^{171}
CRH-R1β(6A/164-169)	^{145}R PGCTHWGDQADGALEVGAAAAAAAFQVRR SI^{176}
CRH-R1β(Δ164-169)	^{145}R PGCTHWGDQADGALEVGA . . . . . FQVRR SI^{171}
CRH-R1*β*(5A/170-174)	^{145}R PGCTHWGDQADGALEVGAPWSGAPAAAAA SI^{176}
CRH-R1*β*(*Δ*170-174)	^{145}R PGCTHWGDQADGALEVGAPWSGAP . . . . . SI^{171}

Fig. 3. Schematic overview of CRH-R1β mutant receptors. The snake-like amino acid sequence represents the IC1. Amino acids common for CRH-R1α and -R1β are presented with white circles, and amino acids specific for CRH-R1β are in *grey circles*. Cassettes of 5 or 6 amino acids were mutated either by deletion (Δ) or alanine substitution (A).

of primers containing novel restriction sites, it is important to add some extra nucleotides to the 5′-end of the sequence containing the restriction site. This will maximize enzymatic efficiency during digestion of the PCR product required for the insertion in an appropriate expression vector. Four to six base pairs is a suitable length to protect the restriction site, although this can vary; for some restriction enzymes, only 1 bp is enough, while for others, more than 20 bp are required (see Note 5).

2. The internal primers (FM and RM shown in Fig. 2) could range from 18 to 35 bp (8), but longer primers of 45 bp have

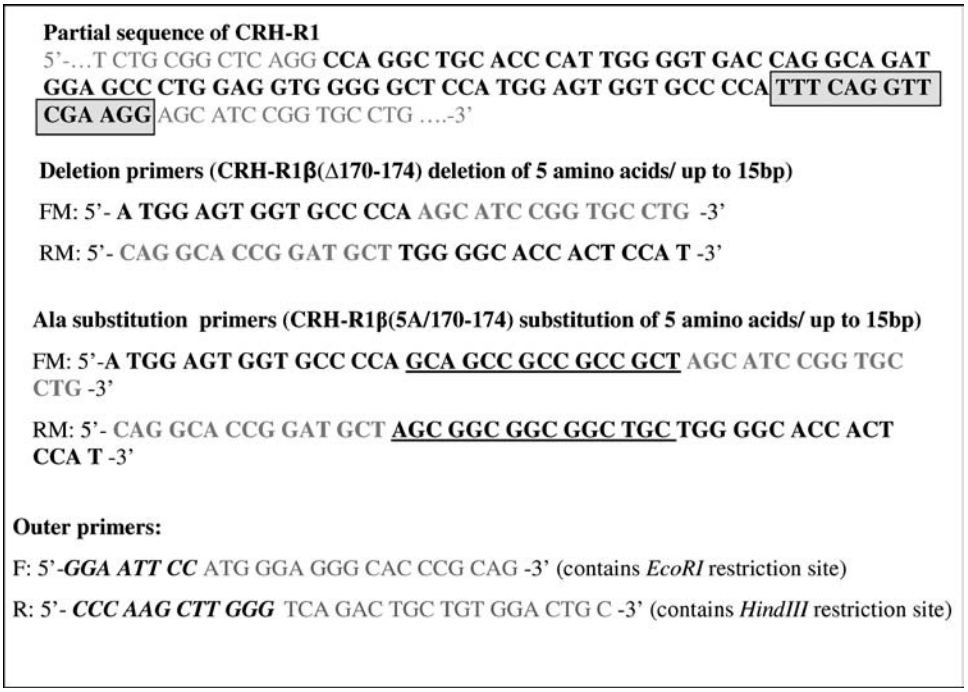


Fig. 4. Partial nucleotide sequence of CRH-R1β and primers used to create mutant receptors. In the partial nucleotide sequence (L23333-NCBI) of CRH-R1, the *bold letters* represent nucleotides specific for a CRH-R1β splice variant. The nucleotides highlighted in a *grey box* were mutated by using either deletion primers to create CRH-R1β(5Δ170-171) or Ala substitution primers to create CRH-R1β(5A/170-174) mutant receptor containing 5 alanines (*underlined nucleotides*). The outer primers contained 5′ restriction sites: F primer has an *Eco*RI restriction site adjacent to the 5′end of CRH-R1, and R primer has a sequence for *Hin*dIII restriction site after CRH-R1 stop codon.

also been employed (see Note 6). The internal primers contain nucleotides adjacent to sequences for site-directed mutagenesis. Depending on the type of mutation required, we used primers with 10–20 nucleotides on either side of the mutation (Fig. 4). Primers FM and RM are reverse complement sequences to each other.

3. Other standard guidelines of primer design (9) are also taken into consideration (see Note 7).
4. Once desired primers are designed, it is important to recheck the sequence for possible typographical errors.

3.3. Determination of DNA Quality and Concentration

As a starting DNA template, full length CRH-R1β cDNA was used, previously generated by reverse transcription and ligated into pcDNA3.1(–). The template was quantified by NanoDrop UV spectrophotometry as follows: (see Note 8)

1. Load 1.5 μl of DNA in the measurement chamber.
2. Record the DNA concentration (in ng/μl) and A_{260}/A_{280} ratio. A ratio of 1.7–2.0 indicates good quality DNA.

For information about different means of determining DNA concentration (see Note 9).

3.4. Setting Up PCRs to Generate Products R-FM and F-RM (Reactions A and B)

(see Note 10).

1. Primers, dNTP mix, *Pfu* DNA polymerase accompanying buffer, DNA template, and water were placed on ice to defrost. *Pfu* DNA polymerase was kept at −20°C until needed.
2. PCR tubes were labeled for two negative controls-NC (water instead of DNA) and products R-FM and F-RM, and placed on ice.
3. Reagents were combined in individual PCR tubes, as described below (Table 1) (see Note 11).
4. The tubes were briefly spun in a cold centrifuge at 6,000 × *g* for 30 s, placed in the PCR machine, and amplification performed using the program described in Table 2.
5. At the end of the reaction, PCR products were stored at 4°C for a day or at −20°C for up to two weeks, or run on a 1.2% agarose gel immediately.

3.5. Setting Up Agarose Gel Electrophoresis and Separation of PCR Products

(see Note 12)

1. The desired volume of 1× TBE is placed in an Erlenmayer flask, and agarose is added to a final concentration of 1.2% (w/v). The flask volume should be at least 3 times larger than

Table 1
PCR mixture for reactions A and B

Reagents (stock conc)	R-FM	R-FM NC	F-RM	F-RM NC	Final concentration
	Amount per reaction (µl)				
Pfu PCR buffer (10×)	5	5	5	5	1×
dNTP mix (40 mM)	1	1	1	1	0.8 mM
Primer F (10 µM)	–	–	2.5	2.5	0.5 µM
Primer RM (10 µM)	–	–	2.5	2.5	0.5 µM
Primer R (10 µM)	2.5	2.5	–	–	0.5 µM
Primer FM (10 µM)	2.5	2.5	–	–	0.5 µM
DNA (vary)	Variable	–	Variable	–	50–100 ng
Pfu DNA polymerase	Variable	Variable	Variable	Variable	3 U
Water	Up to 50 µl	Up to 50 µl	Up to 50 µl	Up to 50 µl	

Table 2
PCR program for reactions A and B

Number of cycles	1	2–35	36	37
Denature	95°C, for 5 min	95°C, for 45 s	N/A	N/A
Anneal	N/A	48°C, for 45 s	N/A	N/A
Elongate	N/A	72°C, for 90 s	72°C, for 7 min	N/A
Hold	N/A	N/A	N/A	4°C, hold

the volume of TBE. To set up a small gel (approximately 10×10 cm), 50 ml of 1× TBE and 0.6 g agarose are added in a 250 ml Erlenmyer flask.

2. The flask is placed in a microwave oven (other sources of heat can be used) and heated until agarose is completely melted.
3. The agarose solution is allowed to cool down to 45°C (flask should be warm, but not hot to touch), and then ethidium bromide is added to a final concentration of 1 μg/ml (5 μl of 10 mg/ml ethidium bromide solution in 50 ml melted agarose). The top of the flask is covered with aluminum foil and swirled gently to evenly distribute ethidium bromide (see Note 13).
4. The agarose mix is poured into the gel tray containing a comb to produce wells with a volume of least 60 μl. It should take approximately 15 min. for the gel to solidify at room temperature.
5. The solidified gel is placed into the gel electrophoresis tank, and the combs are removed.
6. 1× TBE buffer is poured in the electrophoresis apparatus until the gel is entirely covered with the buffer.
7. 10 μl of DNA ladder is loaded, usually into the first well.
8. 10 μl of 5× DNA loading buffer is added to each tube containing 50 μl of PCR reactions, and the entire amount is loaded onto the gel (see Note 14).
9. The gel is run at 80–120 V until the tracking dye front reaches halfway to the bottom of the gel. The progress of separation is checked with a UV light source (see Note 15). The gel should run until the DNA ladder is resolved and the bands from PCR reactions are well separated to allow size determination with confidence (see Notes 16 and 17).

10. Using a clean razor blade, the band is cut out of gel, and placed in a clearly labeled microfuge tube.
11. The weight of a gel fragment is determined; if it is more than 0.4 g, then the fragment should be cut into two pieces and placed into two tubes.

3.6. Purification and Quantification of PCR Products (R-FM and F-RM)

1. The PCR products were purified from the agarose slices using the QIAquick Gel Extraction kit, according to the manufacturer's guidelines (see Note 18).
2. Quantification of DNA was done as described above (Subheading 3.3).
3. DNA can be stored at −20°C indefinitely.

3.7. Setting Up PCRs to Generate Products FR (Reaction C)

1. The volumes of the PCR products R-FM and F-RM were calculated to contain at least 50 ng of DNA. Equal amounts of R-FM and F-RM products were used as templates for Reaction C.
2. Primers, dNTP mix, *Pfu* DNA polymerase accompanying buffer, DNA template, and water were placed on ice to defrost, while *Pfu* DNA polymerase was kept at −20°C until use.
3. PCR tubes were labeled including negative controls-NC (water instead of DNA) and products FR, and placed on ice.
4. The reagents were combined in individual PCR tubes as described below (see Table 3; see Note 19).

Table 3
PCR mixture for reaction C

Reagents (stock concentration)	FR	FR-NC	Final concentration
	Amount per reaction (µl)		
Pfu PCR buffer (10×)	5	5	1×
dNTP mix (40 mM)	1	1	0.8 mM
Primer F (10 µM)	2.5	2.5	0.5 µM
Primer R(10 µM)	2.5	2.5	0.5 µM
Product F-RM (variable)	Variable	–	50–100 ng
Product R-FM (variable)	Variable	–	50–100 ng
Pfu DNA polymerase (vary)	Variable	Variable	3 U
Water	Up to 50 µl	Up to 50 µl	

5. Tubes were centrifuged briefly, placed in the PCR machine, and run using the program described in Table 4 (see Note 20).

3.8. Isolation and Purification of Product FR

1. Agarose gel separation of PCR product was carried out, as described in Subheading 3.5.
2. The bands were purified, and the concentration of DNA was determined, as described in Subheading 3.3.

3.9. Digestion and Ligation of the Product FR into pcDNA3.1(–)

1. Two digestion reactions were prepared in 0.5 ml microfuge tubes, which were appropriately labeled. One digestion reaction was set to digest and linearize the vector pcDNA3.1(–), whereas the other reaction was used to cut the PCR product –FR- at the 5′ and 3′ ends.
2. The reagents were combined as described below (Table 5) and kept on ice, while the enzymes were added.
3. The tubes were centrifuged and placed at 37°C for 2 h (see Note 21).

Table 4
PCR program for reaction C

Number of cycles	1	2–35	36	37
Denature	95°C, for 5 min	95°C, for 45 s	N/A	N/A
Anneal	N/A	52°C, for 45 s	N/A	N/A
Elongate	N/A	72°C, for 90 s	72°C, for 7 min	N/A
Hold	N/A	N/A	N/A	4°C, hold

Table 5
Restriction enzyme digestion mixture

	Cut product FR	Cut vector	
Reagent	Amount per reaction (µl)		Final concentration
10× Tango buffer	3	3	1×
Product FR	Variable	–	Approx. three-fourth amount of product available
pcDNA3.1(–)	–	Variable	1–5 µg
*Eco*RI	Variable	Variable	2 U/µg of DNA
*Hin*dIII	Variable	Variable	2 U/µg of DNA
Water	Up to 30 µl	Up to 30 µl	

Table 6
Ligation mixture

Reagents	Amount per reaction (μl)	Final concentration
2× rapid ligation buffer	5	1×
Insert	Variable	72.2 ng
Vector	Variable	100 ng
T4 ligase	Variable	3 U
Water	Up to 10 μl	

4. During the digestion period, agarose gel was prepared, and after completion of the digestion reaction, the products were separated, as described above (see Subheading 3.5).
5. The bands were purified, and the DNA concentration was determined, as described above (see Subheading 3.3).
6. The amount of digested product FR (insert) and vector used for the ligation reaction was calculated. We usually use 100 ng of vector and 3:1 molar ratio of insert to vector, although it is a good practice to set up several ligation reactions with varying amounts of insert to vector (1:1, 3:1, 6:1) (see Note 22).
7. The ligation reaction was performed using LigaFast™ Rapid DNA Ligation System. The ligation mixture was prepared as described in Table 6.
8. The reaction was incubated at room temperature for 5 min. for cohesive-end ligation, and 15 min. for blunt-end ligation.
9. The products of the ligation reaction can be stored for several months at −20°C. However, we always proceed immediately to the transformation of *E. coli* DH5α.

3.10. Transformation of Plasmid into E. coli DH5α

(see Note 23).

1. Competent cells (*E. coli* DH5-α) were removed from −80°C freezer and thawed on wet ice.
2. Required number of 15 ml Falcon tubes were placed on ice.
3. *E. coli* DH5-α were gently mixed, and aliquots of 100 μl of competent cells were placed into Falcon tubes.
4. 10 μl of ligation reaction was added to the cells, moving the pipette through the cells while dispensing. Also, a negative control was set up in which no DNA was present.
5. Gently tap tubes to mix competent cells and DNA (see Note 24).

6. The cells were incubated with DNA on ice for 30 min. Then, cells were heat-shocked in a water bath for 45 s at 42°C and placed again on ice for 2 min. 0.9 ml of room temperature LB-broth (without antibiotics) was added to the cells.
7. The cells were incubated with shaking at 225 rpm at 37°C for 1 h.
8. 100 µl and 200 µl of transformed cells were spread on LB (amp) agar plates (100 µg/ml ampicillin) and incubated at 37°C overnight (16–18 h).
9. The following morning, plates were removed from the incubator, wrapped with parafilm, and stored at 4°C (see Note 25).
10. Late in the afternoon, 5 ml of LB-Broth with 100 µg/ml ampicillin was added to a 15 ml Falcon tube.
11. Using a sterile pipette tip, one colony from the plate was taken and placed in the 15 ml Falcon tube. The same was repeated for another 5–7 colonies. The tubes were incubated overnight at 37°C with shaking (225 rpm).

3.11. Mini-Prep of Plasmid DNA

1. 4 ml of overnight grown bacteria from each Falcon tube described in Subheading 3.10, step 11 were used to isolate plasmid DNA. The rest of the bacteria were stored at 4°C. QIAprep Spin Mini prep kit was used, according to the manufacturer's instruction.
2. DNA was eluted with molecular biology grade water.
3. The concentration of plasmid DNA was determined as described in Subheading 3.3.
4. A diagnostic restriction enzymes digest was set up to identify whether the insert was present in the isolated plasmids (see Note 26). The resulting products were run on 1.2% agarose gel (as described in Subheading 3.5)
5. A sequencing reaction was prepared only with the insert-positive plasmids. We combined 500 ng of plasmid DNA with 5.5 pmol of sequencing primers (these primers should anneal to at least 40 nucleotides upstream of mutation site) in total volume of 10 µl (see Note 27).
6. Insert-negative plasmid DNAs were disposed, whereas insert-positive plasmid DNA were stored at –20°C.

3.12. Analysis of Sequencing Results

1. Once clones with desired mutations were identified, the full length insert was sequenced to confirm that no spontaneous mutations have been introduced.
2. Various alignment softwares are available. We used Biology WorkBench web site. This site has been developed at the University of San Diego (http://workbench.sdsc.edu/) and

provides an excellent resource for different analysis of proteins and nucleic acids. Registration is required to access these services; however, they are free.

3. Once registered, the "Nucleic Acid" tool box is selected.
4. On scroll down menu, "Add new nucleic sequence" is selected. Then, the sequence of the gene of interest in FASTA format was uploaded (in our case, it was CRH-R1β NCBI nucleotide database entries L23333). The same was repeated for the sequence of the mutant gene received from the sequencing services.
5. Both entries were selected and aligned by selecting "CLASTALW-multiple sequence alignment" from the scroll down menu.
6. Alignment of the two sequences confirmed that the desired mutation was created and that there were no spontaneous mutations. Once this criterion was satisfied, it was safe to proceed to the next step.

3.13. Midi-Prep (or Maxi-Prep) of Plasmid DNA

1. 100–125 ml LB-Broth containing ampicillin (100 μg/ml) was added to a 300 ml sterile Erlenmeyer flask.
2. 100 μl of bacteria (see Subheading 3.11, step 1) was added to the flask and incubated overnight (16–18 h) at 37°C with shaking (250 rpm).
3. Before spinning the overnight culture for preparation of plasmid DNA, a glycerol stock of bacteria was made: a cryovial was labeled with name, date, plasmid, gene name, and bacterial info. 400 μl of sterile glycerol stock (50%) and 600 μl of bacteria from the overnight culture were added into the cryovial, mixed well, and kept it at −80°C.
4. To prepare a plasmid DNA midi prep, the Qiagen HiSpeed Plasmid midiprep kit was used, according to the manufacturer's instruction (see Note 28).
5. Concentration and quality of plasmid DNA was determined, as described before (see Subheading 3.3), and then stored at −20°C.

3.14. Transfection of HEK293 Cells and Validation of Mutant Receptors

1. Transient transfection of wild type and mutant receptors cDNA was carried out in 50–70% confluent cells. 5 μg or 10 μg DNA was transfected in cells seeded in 25 cm^2 or 75 cm^2 vented flasks, with 5 μl or 10 μl Lipofectamin2000 reagent, respectively, in 5 ml or 10 ml of OptiMEM + GlutaMax, for overnight (16–18 h) incubation, according to manufacturer's instructions.
2. The following morning, the transfection mixture was replaced with normal growth media (DMEM with Glutamax containing 10% FCS, 100 U/ml penicillin, and 100 μg/ml streptomycin.).

3. After 12 h, the transfected cells were transferred onto poly-D-lysine (Sigma) coated plates (6, 12 or 24-well plates) or on coated glass cover slips (for confocal microscopy studies).
4. All experiments were carried out 48–72 h after transfection.
5. Receptor expression was verified by western blot analysis (see Fig. 5 *top*) using CRH-R specific antibodies and *Enhanced Chemiluminescence* (ECL) detection (5). Membrane localization of expressed receptors was verified by indirect confocal microscopy analysis (see Fig. 5 *bottom*) (4). These experiments

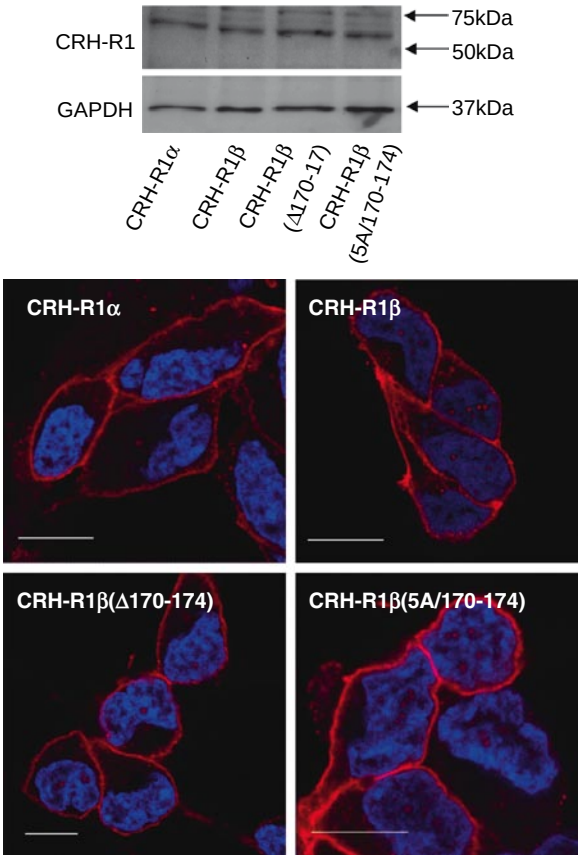


Fig. 5. Detection of mutant receptors expression by western blotting and confocal microscopy. *Top*: Comparison by Western blot analysis of CRH-R1α, -R1β, and mutant CRH-R1β receptor expression in HEK 293 cells. Cell lysates from HEK 293 cells transiently expressing CRH-R1α, -R1β, or mutant -R1β receptors were fractionated by SDS-PAGE, and subjected to immunoblotting with a specific CRH-R1/2 antibody (5). *Bottom*: Comparison of CRH-R1α, -R1β, and mutant CRH-R1β receptor plasma membrane expression in HEK 293 cells by indirect confocal microscopy (4). HEK 293 cells transiently expressing CRH-R1α, -R1β, or mutant -R1β receptors were fixed in 4% parafolmaldehyde. Following blocking of nonspecific binding and permeabilization of cells, CRH-R1 was detected using CRH-R1-specific antiserum and Alexa-Fluor®594 secondary antibody (*red*) (4). Cell nuclei were stained using DAPI. Scale bar is 10 μm.

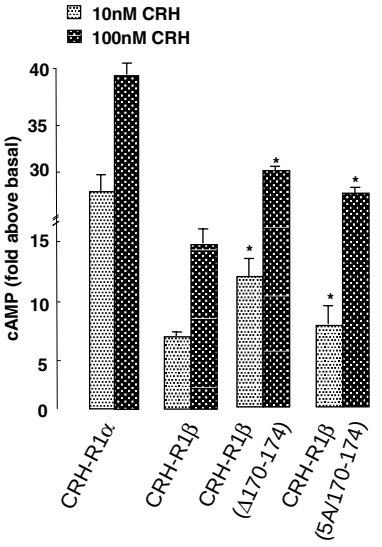


Fig. 6. Functional assessment of CRH-R1β mutant receptors. Transiently transfected HEK293 were grown in 12-well plates. Stimulation buffer (DMEM containing 10 mM $MgCl_2$ and 0.1 mg/ml 3-isobutyl-1-methylxanthine) was added on cells for 20 min before the stimulation. Cells were stimulated with 10 and 100 nM CRH for 15 min. Intracellular cAMP production was measured using ELISA kit (11).

demonstrated that deletion or Ala substitution of 5 amino acids in CRH-R1β (Δ170-174 or 5A/170-174) did not alter the levels of receptor expression or subcellular localization, since the mutant receptors were exclusively expressed in the plasma membrane.

6. Functionality of the mutant receptors was assessed by determining CRH induced cAMP production (see Fig. 6). Intracellular cAMP was measured with RIA or commercially available ELISA kits (4, 11). The results demonstrated that the deletion or substitution of 5 amino acids (F^{170}-R^{174}) within the 29-aminoacid insert of CRH-R1β led to a significantly improved signaling response of these mutant receptors compared to wild type CRH-R1β, suggesting that the penta-amino acid cassette might be an important structural determinant of CRH-R1β-impaired cAMP response (5).

4. Notes

1. Water used for PCR reactions, digestion, ligation, and any other molecular biology reactions was molecular biology grade water (DNase and RNase free). The water used to make buffers and other solutions was double distilled water (resistance of 18.2 MΩ/cm).

2. If instead of PCR thin well tubes, another type of tubes is used, always use the same kind of tubes for one type of PCR reactions. Different manufacturers have tubes with different wall thickness, and this can significantly affect PCR efficiency. Also, if other tubes are used instead of PCR tubes, they have to be heat sterilized, prior to use.
3. The development of commercially available mutagenesis kits (such as from Stratagene) can significantly increase the speed of DNA mutagenesis.
4. The primer design step is of pivotal importance for the success of mutagenesis. It is essential that the template DNA sequence is checked for any unwanted spontaneous mutations. The recombinant CRH-R1β cDNA was sequenced (using the core sequencing facility at the University of Warwick, UK) to confirm the fidelity of the nucleotide sequence, and compared to NCBI nucleotide database entry L23333. The partial nucleotide sequence of CRH-R1β is presented in Fig. 4. *Black bold letters* represent nucleotides that are unique for this splice variant and are not present in the CRH-R1α nucleotide sequence. These nucleotides were manipulated in order to create mutant receptors presented in Fig. 3.
5. For more info, see http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/cleavage_olignucleotides.asp.
6. The long primers are usually used when up to 15 adjacent nucleotides are mismatched to the template.
7. There are various companies that provide free primer design service. Stratagene has an excellent web-based resource for primer design (http://www.stratagene.com).
8. The quality of starting material is essential for the entire protocol. We have routinely determined concentration and quality of DNA using NanoDrop spectrophotometer. This system allows use of very small volumes of DNA (which is very important if large amounts of DNA are not available).
9. In the absence of a NanoDrop instrument, a spectrophotometer with a UV light source and quartz cuvettes might also be used. It is advisable to consult each manufacturer's instruction manual on how to use individual instruments.
   (a) A spectrophotometer with a UV lamp is turned on for at least 5 min prior to use.
   (b) 990 µl distilled water and 10 µl of plasmid DNA are added in an 1.5 ml eppendorf microfuge tube and mixed well
   (c) 1 ml of distilled water is placed in a quartz cuvette to measure absorbance of distilled water (blank) at 260 nm and 280 nm, and those values were set as 0.0 (blank).

(d) Water is removed from the cuvette, and the cuvette is dried using fine paper tissue. 1 ml of dilute DNA (from step 2) is added to the cuvette, and the absorbance is measured at 260 nm and 280 nm.

(e) DNA concentration is calculated using the formula

$$A_{260} \times 50 \times \text{dilution factor} = \mu\text{gDNA / ml in undiluted sample}$$

In our experimental example, the dilution factor is 100.

(f) The purity of DNA is determined by the ratio A_{260}/A_{280}. A ratio of A_{260}/A_{280} should be greater than 1.8, and it suggests little or no protein contamination in a DNA sample.

10. It should take approximately 20–30 min to set up PCR reactions and 2–3 h for PCR amplification. All chemicals should be kept on ice, and *Pfu* DNA polymerase at –20°C until needed. It is crucial to use a high-fidelity DNA polymerase to minimize random mutations and maximize the efficiency of PCR. We used *Pfu* DNA polymerase (Promega) in most experiments, although occasionally alternative enzymes from Invitrogen or Fermentas were also used with equally good performance.
11. *Pfu* DNA polymerase is always added last. NC should not contain template DNA, and therefore no PCR product should be amplified in those tubes, unless other reagents are contaminated with DNA. Most often, water and/or primers are contaminated with DNA. These are the cheapest to replace.
12. Lab coat and gloves must be worn when handling agarose gels. The gel contains ethidium bromide, which is highly carcinogenic.
13. While waiting for agarose solution to cool down, a gel tray is prepared. Depending on the electrophoresis system, sometimes, the sides of gel tray need to be sealed with a tape in order to prevent leakage of agarose solution. Since the agarose solution contains ethidium bromide (highly toxic and carcinogenic substance), it is important to secure the gel tray in a case of accidental leakage. We always place the gel tray in another larger plastic tray, which is embedded with paper tissue.
14. It is a good practice to skip a well between samples to avoid spill over from one well to the next and also to facilitate extraction of specific bands from the gel.
15. If the UV light source is visible, protective eyewear is required.
16. In our mutagenesis experiments, we usually stop the run once the tracking dye front has reached 1 cm from the bottom of the gel (the bands of interest were approximately 500 bp and

900 bp). If the size of bands is smaller (less that 100 bp), consider running the gel only halfway or preparing a gel with higher agarose concentration (for example 1.5%).

17. If possible, a photo of the gel should be obtained before cutting the bands from the gel, (an example of a gel obtained during the CRH-R1β mutagenesis experiments is shown in Fig. 7a). The expected size of the PCR products is checked by comparing them with the DNA ladder. Also, there should be no bands in the negative control PCR reactions. However, if random bands are present, this suggests contamination of PCR reagents, and questions accuracy and specificity of any bands observed in the other PCR reactions that contained DNA template.
18. The elution of DNA (the last step in the extraction protocol) was carried out with 30 µl molecular biology grade water.

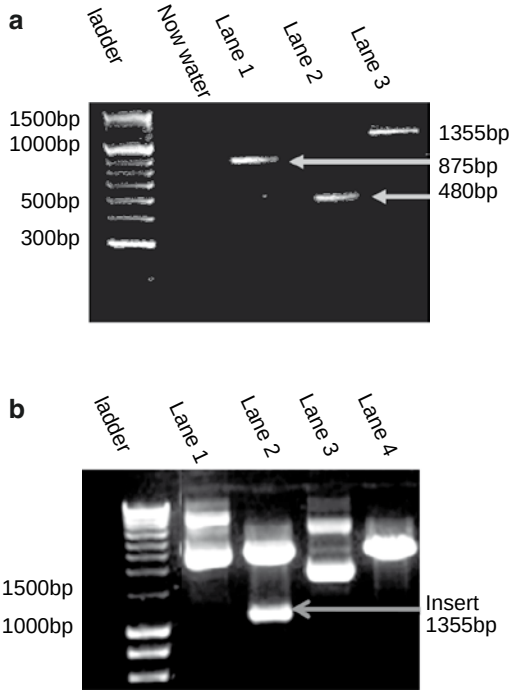


Fig. 7. (**a**) Example of an agarose gel DNA analysis used to separate PCR products from reactions A, B, and C. Following a PCR cycle products from reactions, A – lane 1 (480 bp), B – lane2 (875 bp), and C – lane 3 (1,355 bp) were separated on a 1.2% agarose gel (containing EtBr) and visualized under a UV light source. (**b**) Example of an agarose gel DNA analysis used to determine a presence of the insert in the plasmid. Following a diagnostic digestion reaction, the resulting products were separated on a 1.2% agarose gel (containing EtBr) and visualized under a UV light source. The samples were loaded in a following order: lane 1 – undigested plasmid A, lane 2 – digested plasmid A, lane 3 – undigested plasmid B, and lane 4 – digested plasmid B. In the example shown here, only the plasmid A contains the insert.

19. *Pfu* DNA polymerase is always added last. NC should not have template DNA; therefore, no PCR product should be amplified in those tubes, unless other reagents are contaminated with DNA.
20. Annealing temperature can be adjusted as needed. Annealing temperature used in the first round of PCR reactions was 48°C. This is considered to be low annealing temperature. Quite often, it is advised that annealing temperature should be between 50 and 60°C; generally at lower temperatures, primers can anneal to a nonspecific sequences leading to amplification of additional nonspecific products; also, higher temperatures increase the accuracy of the PCR. However, if large mutations need to be introduced at once (removal or insertion of several nucleotides in a row), using low temperatures (down to 48°C) can result in success. The length of the elongation step depends on the length of the desired PCR product. There is a general rule that a 1 min period is sufficient for the synthesis of products up to 1,000 bp. It is always a good practice to allow more time for complete elongation than to end up with incomplete synthesis of the full length products.
21. If different restriction enzymes are to be used, the manufacturer instructions should be checked for optimal digestion conditions. Some enzymes work optimally at different temperatures, also not all enzymes are suitable for double digestion reactions. Recently, various companies launched fast acting restriction enzymes; these enzymes are equally efficient as traditional with considerably faster digestion times (5 min).
22. The amount of insert is determined using the following formula

$$\text{Insert (ng)} = A \times (\text{vector (ng)} \times \text{insert size(kb)}) / \text{vector size(kb)}$$

where A is a number of insert moles to one mole of vector (for 3:1 molar ratio, $A = 3$). In our mutagenesis experiments, the calculation was

$$\text{Insert(ng)} = 3 \times (100\ \text{ng} \times 1.3\ \text{kb})) / 5.4\ \text{kb}$$

$$\text{Insert(ng)} = 72.2\ \text{ng}$$

23. Before starting plasmid transformation, it is advisable to ensure that LB-Broth, ampicillin, and LB(Amp) agar plates are available. All the steps described in this section are performed using aseptic technique.
24. Pipetting bacteria up and down should be avoided.

25. Sometimes, following ligation and transformation, there are no distinct bacterial colonies on LB-agar plate, but just thin layer of bacteria. In that case, the bacteria need to be streak out onto another plate and incubate for further 18 h at 37°C. There should be no colonies on negative control plate.
26. The reaction mixture used in our experiments was prepared as described below, and the incubation was performed at 37°C for 2 h.

    1°µl DNA

    1°µl *Eco*RI

    1°µl *Hin*dIII

    2°µl Tango buffer (Accompanying buffer)

    13°µl water

    20°µl total

    In our experiments, approximately 70% of screened plasmids had the insert ligated in. An example of gel containing products of a diagnostic digestion reaction of plasmids with and without the insert is shown in Fig. 7b.
27. All our sequencing reactions were performed by technical staff of the molecular biology services at the department of Biological Sciences, University of Warwick (http://www2.warwick.ac.uk/fac/sci/bio/services/molbiol/update2007_sequsersinstructionsv2.pdf).
28. Some cell lines do not transfect very well with plasmids made using commercially available kits. For example, for transfection of differentiated 3T3L1 adipocytes or MIN6 cells, plasmid needs to be prepared using the cesium chloride gradient method (10).

References

1. Smith SM, Vale WW (2006) The role of the hypothalamic–pituitary–adrenal axis in neuroendocrine responses to stress. Dialogues Clin Neurosci 8(4):383–395
2. Hillhouse EW, Grammatopoulos DK (2006) The molecular mechanisms underlying the regulation of the biological activity of corticotropin-releasing hormone receptors: implications for physiology and pathophysiology. Endocr Rev 27:260–286
3. Chen R, Lewis KA, Perrin MH, Vale WW (1993) Expression cloning of a human corticotropin-releasing-factor receptor. PNAS 90(19):8967–8971
4. Markovic D, Papadopoulou N, Teli T, Randeva H, Levine MA, Hillhouse EW, Grammatopoulos DK (2006) Differential responses of corticotropin-releasing hormone receptor type 1 variants to protein kinase C phosphorylation. J Pharmacol Exp Ther 319:1032–1042
5. Teli T, Markovic D, Hewitt ME, Levine MA, Hillhouse EW, Grammatopoulos DK (2008) Structural domains determining signalling

characteristics of the CRH-receptor type 1 variant R1beta and response to PKC phosphorylation. Cell Signal 20(1):40–49

6. Higuchi R, Krummel B, Saiki RK (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res 16(15):7351–7367
7. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51–59
8. Heckman KL, Pease LR (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. Nat Protoc 2(4):924–932
9. Kramer MF, Coen DM (2001) Enzymatic amplification of DNA by PCR: standard procedures and optimization. Curr Protoc Mol Biol:15.1.1–15.1.14
10. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Press, pp 1.42–1.46
11. Markovic D, Punn A, Lehnert H, Grammatopoulos DK (2008) Intracellular mechanisms regulating corticotropin-releasing hormone receptor-2beta endocytosis and interaction with extracellularly regulated kinase 1/2 and p38 mitogen-activated protein kinase signaling cascades. Mol Endocrinol 22(3):689–706

Chapter 22

Site-Directed Mutagenesis for Improving Biophysical Properties of V_H Domains*

Mehdi Arbabi-Ghahroudi, Roger MacKenzie, and Jamshid Tanha

Abstract

Recombinant antibody fragments are significant therapeutic and diagnostic reagents. As such, their efficacy depends heavily on their affinities and biophysical properties. Thus, mutagenesis approaches have been extensively applied to recombinant antibodies to improve their affinity, stability, and solubility. Among the existing recombinant antibody variants, human V_H domains stand out as the ones with the general need of solubility engineering at some point during their development; this solubility engineering step transforms V_Hs into nonaggregating, functional entities, rendering them useful as therapeutic and diagnostic reagents. Here, we present one of several approaches that have been employed to develop nonaggregating human V_H domains. We apply an in vitro site-directed mutagenesis approach to an aggregating human V_H domain by means of a splice overlap extension technique. The resultant mutant V_Hs are nonaggregating in contrast to the parent wild type V_H and less prone to aggregation following thermal unfolding.

Key words: In vitro mutagenesis, Splice overlap extension, V_H, Biophysical properties, Reversible thermal unfolding, Nonaggregating, Surface plasmon resonance, Circular dichroism

1. Introduction

Mutagenesis approaches have contributed enormously to our understanding of protein structure–function relationship and to improving biophysico-chemical properties of proteins (1–9). In the field of recombinant antibody engineering, in vitro and in vivo mutagenesis approaches have been employed to improve antibody affinity and stability (3, 6, 7, 10). One class of recombinant antibodies is the single-domain antibody (sdAb) encompassing a

* This is National Research Council of Canada Publication 50001.

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_22, © Springer Science+Business Media, LLC 2010

number of varieties (11–14). sdAbs offer several advantages over other recombinant antibody fragments, which make them highly desirable for biotechnological applications (11, 12, 14, 15). Of particular interest are human V_H sdAbs, which have the added advantage of being less immunogenic in human therapy compared to other sdAbs, e.g., camelid V_HHs. Thus, it is worthwhile to spend time and focus resources on devising ways of obtaining nonaggregating human V_H domains, as V_Hs are notoriously prone to aggregation. In fact, the number of laboratories working on developing human V_H domains has been steadily on the rise, and the focus of one particular company, Domantis/GSK, is exclusively on developing human V_H (and V_L) therapeutics.

A popular source of human V_H binders is V_H display libraries made by complementarity-determining region (CDR) randomization on a single V_H scaffold (16–19). Various approaches for obtaining nonaggregating V_H domains from V_H libraries are depicted in Fig. 1 (10, 16, 18–23). The approach pertaining to the present chapter involves the prestep of solubilizing a V_H by site-directed mutagenesis before its use as a library scaffold.

PCR-based site-directed mutagenesis, in particular its significantly improved version involving splice overlap extension (SOE), is a rapid, consistent technique with virtually 100% efficiency (24–28). SOE has been commonly employed to introduce mutations at any desired DNA site (29–38). In the first step of the SOE approach, two overlapping DNA sub-fragments with the desired mutations are generated by polymerase chain reaction (PCR) (5, 25, 26, 39). In the next step, the two subfragments are mixed, heat-denatured, and then hybridized (spliced) in their terminal overlapping regions, forming duplex DNA with long overhangs at either end. The resultant staggered duplex is then extended in the overhang regions by a thermostable DNA polymerase, e.g., *Taq* DNA polymerase, forming a complete duplex. Several rounds of DNA denaturation, splicing, and extension provide enough mutagenized template, which is then amplified by PCR, performed following the inclusion of appropriate primers.

In this chapter, we describe protocols for improving biophysical properties of an aggregating human V_H by site-directed mutagenesis. SOE was employed to introduce site-specific mutations at several amino acid positions, and the resultant V_H mutants outperformed the wild type V_H in terms of nonaggregation and reversible thermal unfolding properties. As a scaffold, one of the mutant V_Hs led to a stable and functional library in contrast to the wild type V_H scaffold, which gave a dysfunctional library.

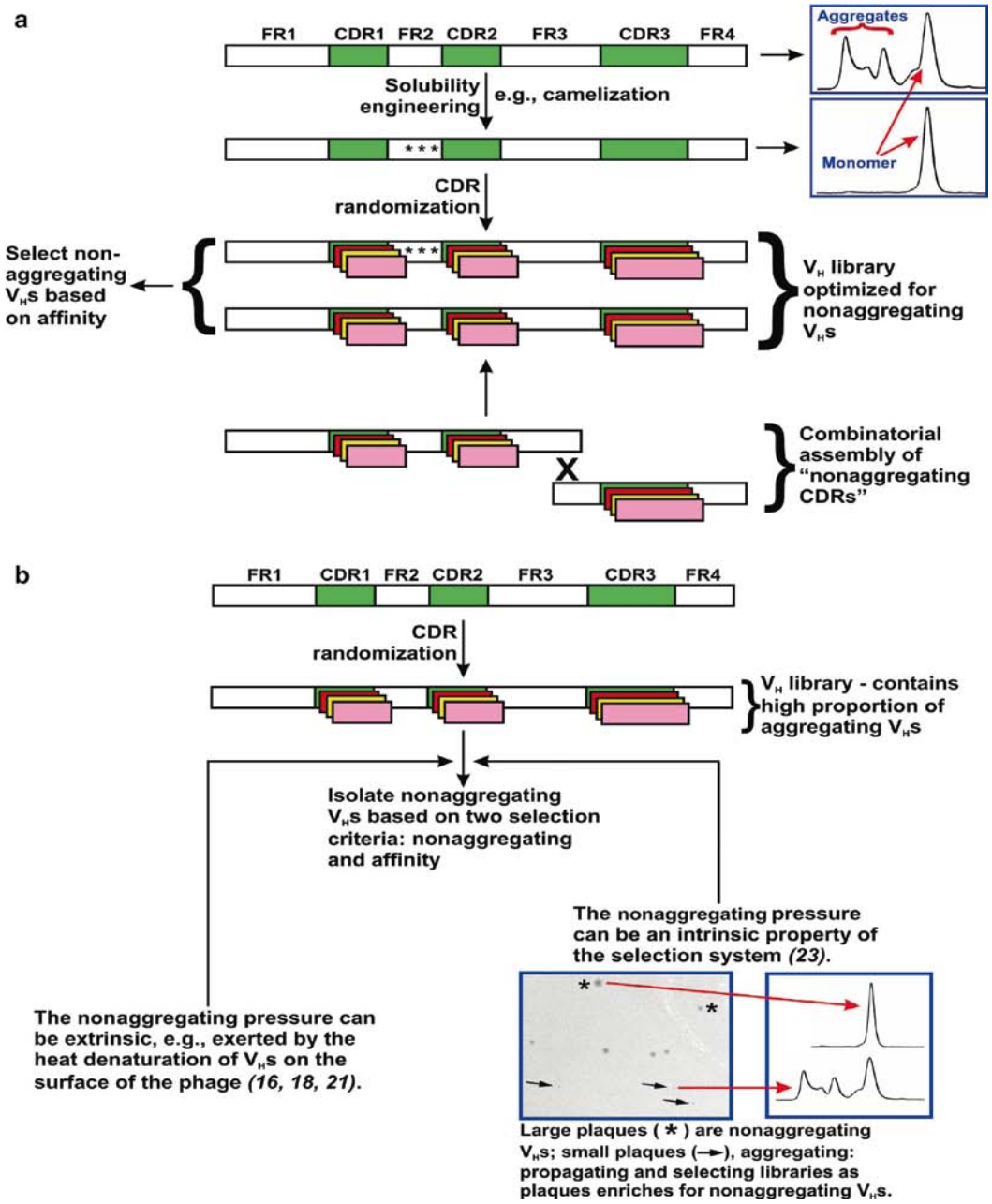


Fig. 1. Approaches for selecting nonaggregating human V_H binders from V_H phage display libraries. V_H libraries are constructed by randomizing the CDRs of a single V_H scaffold. In one approach (**a**), the proportion of nonaggregating V_Hs in the library is optimized, and the selection is performed based on affinity. Library optimization can be achieved by including a solubility engineering step to generate a nonaggregating library scaffold (10, 19, 22) or by the combinatorial assembly of CDRs from repertoires with a high frequency of aggregation-resistant V_Hs (20). In a second approach (**b**), library optimization is bypassed, but selection is performed based on affinity and nonaggregation (16, 18, 21, 23). Asterisks (**a**) represent amino acid substitutions. The right hand panels show the size exclusion chromatogram profiles for aggregating and nonaggregating V_Hs.

2. Materials

2.1. Construction of Mutant V_Hs by the SOE Approach

1. VHBACKBbsI:5′-TATGAAGACACCAGGCCGAGGTCCAGCTGCAGGCGTCT-3′.
2. VHBACKMut1:5′-GAGGTCCAGCTGCAG**GCG**TCTGGGGGAGGCTTAGTCCAGCCTGGGGGGTCCCTGAGACTCTCCTGT**GCG**GCTAGCGGATTCACC-3′ (E6A/S23A).
3. VHFORFR3Mut1:5′-GTAATACACAGCCGTGTCCTCAGCTCTCAGACT**ATT**CATTT GAAGATA-3′ (S82aN).
4. VHFORFR3Mut1a: 5′-GTAATACACAGCCGTGTCCTC**CGGTTT**CAGACT**ATT**CATTTGAAGATACAGAGTGTTCTT**CGC**ATTGTC-3′ (S74A/S82aN/R83K/A84P).
5. VHBACKFR3Mut2:5′-TGAGAGCTGAGGACACGGCTGTGTATTACTGT**GCG**AA AGACAGG-3′ (V93A).
6. VHBACKFR3Mut2a: 5′-AGTCTG**AAACCG**GAGGACACGGCTGTGTATTACTGT**GCG**AAAGACAGG-3′ (V93A/R83K/A84P).
7. VHFORBamHIMut3: 5′-CGATGTCTACGGACGCGAAGGATCCTGATGAGACGGTGAC**CTG**TGTCCCTTG-3′ (T108Q).
8. VHASSUF: 5′-GTCAGACGATGTCTACGGACGCGAA-3′.

   (Primers were purchased from *Sigma Genosys, The Woodlands, TX*. The underlined sequences denote the overlapping regions for SOE purposes (see Fig. 2). Mutated codons are shown in bold, and the corresponding amino acid substitutions are in brackets.)
9. dNTPs: *NEB, Pickering, ON, Canada*.
10. Expand High Fidelity Buffer (10×, with 15 mM $MgCl_2$): *Hoffmann-La Roche Ltd., Mississauga, ON, Canada*.
11. Expand High Fidelity Enzyme mix (Taq DNA polymerase): *Hoffmann-La Roche Ltd*.
12. pSJF2H-BT32/A6 plasmid[1] (contains the gene for BT32/A6 V_H) (19, 40).
13. Autoclave-sterilized double-distilled water (sddH_2O) (41). Sterilize by autoclaving for 20 min at 15 lb/sq. in. on liquid cycle. Store at room temperature.
14. QIAquick PCR purification™ kit: *QIAGEN, Mississauga, ON, Canada*.
15. Agarose gel electrophoresis equipment.
16. ND-1000 spectrophotometer: *NanoDrop Technologies, Wilmington, DE*, or a similar instrument.

[1] All plasmid constructs are available to researchers free of charge.

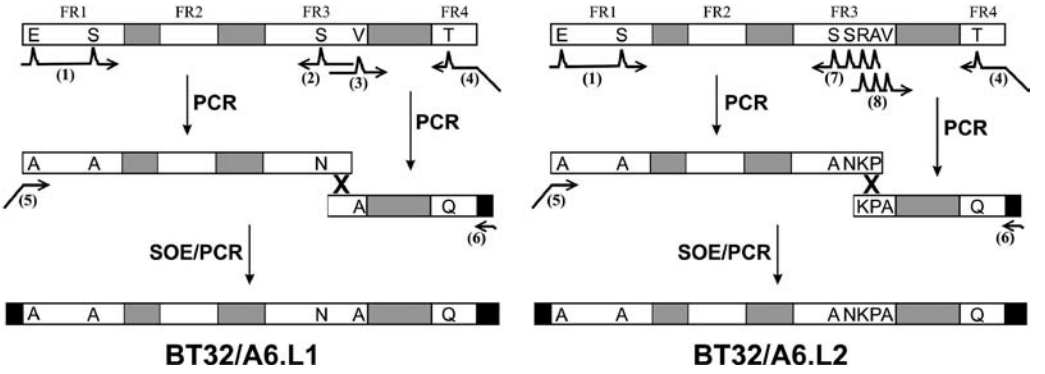


Fig. 2. Schematic representation of steps taken to construct BT32/A6.L1 and BT32/A6.L2. Using pSJF2H-BT32/A6 vector as the template (for simplicity only the DNA insert coding for BT32/A6 V_H is shown) and mutagenic primers, two sets of overlapping fragments containing a total of five (BT32/A6.L1) and eight (BT32/A6.L2) mutations were constructed by PCR. Primer 4 also added a flanking *Bam*HI restriction endonuclease site for subsequent cloning purposes. The two sub-fragments were purified and subsequently spliced, through their complementary overlapping region (shown by "X"), and amplified by SOE/PCR using primers 5 and 6. Primer 5 also added a *Bbs*I restriction endonuclease site. The SOE/PCR products (BT32/A6.L1 and BT32/A6.L2) were purified, digested with *Bam*HI and *Bbs*I restriction endonucleases, and cloned into pSJF2H vector, as described in Subheading 3.2.1. The additional sequences, shown as black blocks at the 3′ end of the small sub-fragments, were incorporated in order to prevent the amplification of the wild type V_H from the pSJF2H-BT32/A6 carry over (see Note 6). (1) VHBACKMut1; (2) VHFORFR3Mut1; (3) VHBACKFR3Mut2; (4) VHFORBamHIMut3; (5) VHBACKBbsI; (6) VHASSUF; (7) VHFORFR3Mut1a; (8) VHBACKFR3Mut2a.

2.2. Cloning, Expression, Extraction, and Purification of V_Hs

2.2.1. Cloning

1. Restriction endonucleases *Bbs* I and *Bam*H I: *NEB.*
2. 10× *Bbs*I and *Bam*H I buffers: *NEB.*
3. Expression vector pSJF2H (15).
4. LigaFast™ Rapid DNA Ligation System: *Promega, Madison, WI.*
5. Electroporation-comptent TG1 cells: *Stratagene, La Jolla, CA.*
6. SOC (per liter: bacto-trypton, 20 g; bacto-yeast extract, 5 g; NaCl, 0.5 g; KCl, 2.5 mM; $MgCl_2$, 10 mM; glucose, 20 mM; deionized H_2O), LB (per liter: bacto-trypton, 10 g; bacto-yeast extract, 5 g; NaCl, 10 g; deionized H_2O), and LB/agar (per liter: bacto-trypton, 10 g; bacto-yeast extract, 5 g; NaCl, 10 g; bacto-agar, 15 g; deionized H_2O) media: (41).
7. Filter-sterilized ampicillin.
8. Taq DNA polymerase: *Hoffmann-La Roche Ltd.* (see Note 1).
9. M13RP: 5′-GCGGATAACAATTTCACACAGGAA-3′: *Sigma Genosys.*
10. M13FP: 5′-CGCCAGGGTTTTCCCAGTCACGAC-3′: *Sigma Genosys.*
11. Autoclaved-sterilized glycerol: *BDH Inc., Toronto, ON, Canada* (41). Sterilize by autoclaving for 20 min at 15 lb/sq. in. on liquid cycle. Store at room temperature.

12. 0.2 µm MILLEx®-GV filter unit: *Millipore, Cambridge, ON, Canada* (see Note 2).
13. Electroporation cuvettes: *Bio-Rad Laboratories, Mississauga, ON, Canada*.
14. MicroPulser™ electroporator: *Bio-Rad Laboratories* or a similar instrument.
15. DNA sequencing equipment (3100 Genetic Analyzer): *Applied Biosystems, Foster City, CA*.
16. Autoclave-sterilized toothpicks or P10 pipette tips. Autoclave on a dry cycle.

2.2.2. Protein Expression and Extraction

1. M9/Amp: M9 medium (per liter: $Na_2HPO_4.7H_2O$, 12.8 g; KH_2PO_4, 3 g; NaCl, 0.5 g; NH_4Cl, 1 g; $MgSO_4$, 2 mM; glucose, 0.4%; $CaCl_2$, 0.1 mM) (41) supplemented with 5 µg/mL vitamin B1, 0.4% casamino acids, and 100 µg/mL ampicillin.
2. 10× induction medium: per liter: bacto-trypton, 12 g; bacto-yeast extract, 24 g; glycerol, 4 mL; deionized H_2O. Sterilize by autoclaving for 20 min at 15 lb/sq. in. on liquid cycle.
3. Filter-sterilized IPTG (isopropyl-β-D-thio-galactopyranoside).
4. Wash solution: 10 mM Tris–HCl pH 8.0, 154 mM NaCl.
5. Sucrose solution: 10 mM Tris–HCl pH 8.0, 1 mM EDTA, 25% filter-sterilized sucrose.
6. Shock solution: 10 mM Tris–HCl pH 8.0, 0.5 mM $MgCl_2$. Always keep refrigerated.
7. Anti-c-Myc tag monoclonal antibody: *Jackson Immuno-Research Inc., Cambridgeshire, UK*.
8. Starting buffer: 10 mM HEPES (*N*-[2-hydroxyethyl]piperazine-*N*′-[2- ethanesulfonic acid]), 10 mM imidazole, 500 mM NaCl, pH 7.0.
9. Dialysis membrane (10 kDa MWCO).
10. SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) and Western blotting equipment.

2.2.3. Purification

1. $NiCl_2.6H_2O$ (5 mg/mL).
2. Elution buffer: 10 mM HEPES, 500 mM imidazole, 500 mM NaCl, pH 7.0.
3. Sodium phosphate buffer A: 6.7 mM Na_2HPO_4, 3.3 mM NaH_2PO_4, 150 mM NaCl, 0.5 mM EDTA, pH 7.0.
4. Sodium azide.
5. 5-mL HiTrap™ Chelating HP column: *GE Healthcare, Baie d'Urfé, QC, Canada*.

6. 0.2 µm GP Express™ Plus Membrane filtration system: *Millipore* (see Note 3).
7. ÄKTA FPLC purification system: *GE Healthcare*.

2.3. Assessment of Aggregation State of V_Hs by Size Exclusion Chromatography

1. Superdex™ 75 10/300 GL gel filtration column (bed volume: 24 mL; bed dimensions: 10 mm × 300 mm): *GE Healthcare*.
2. Filtered and degassed ddH_2O. Degas the filtered water with a conventional water aspirator.
3. PBS buffer (per liter: NaCl, 8 g; KCl, 0.2 g; Na_2HPO_4, 1.44 g; KH_2PO_4, 0.24 g; deionized H_2O): (41), filtered and degassed.

2.4. Circular Dichroism (CD) Spectroscopy Determination of Thermal Refolding Efficiency (TRE) of V_Hs

1. Sodium phosphate buffer B: 6.7 mM Na_2HPO_4, 3.3 mM NaH_2PO_4, pH 7.0.
2. Circular CD cuvettes: *Hellma Canada Limited, Concord, ON, Canada*.
3. Jasco J-600 spectropolarimeter: *JASCO Corporation, Tokyo, Japan*.
4. Neslab RTE-110 water bath: *NESLAB Instruments Inc., Portsmouth, NH*.
5. 6 N HCl: *EMD Chemicals, Inc., Gibbstown, NJ*. Prepared by diluting in ddH_2O a stock of 12.1 N HCl.
6. Methanol (min 99.8%): *EMD Chemicals, Inc*.

2.5. TRE Determination by Surface Plasmon Resonance (SPR)

1. HBS-EP buffer: 10 mM HEPES buffer pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% surfactant P20 (see Note 4).
2. Surfactant P20: *GE Healthcare*.
3. Amine coupling kit containing *N*-hydroxysuccinimide (NHS), *N*-ethyl-*N*′-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and 1 M ethanolamine: *GE Healthcare*.
4. 10 mM acetate buffer, pH 4.5: *GE Healthcare*.
5. Sensorchip CM5: *GE Healthcare*.
6. BIACORE 3000: *GE Healthcare*, or other surface plasmon resonance instrument with similar capabilities.
7. BIAevaluation software 4.1: *GE Healthcare*.

3. Methods

3.1. Construction of V_H Mutants by the SOE Approach

In this step, site directed mutagenesis involving SOE is employed to construct two V_H mutants, namely BT32/A6.L1 and BT32/A6.L2, from the human V_H domain BT32/A6 template (19, 40), which has been shown to have significant aggregation tendency, a

common characteristic of V_Hs (23). BT32/A6.L1 is mutated at five positions with respect to the wild type V_H (E6A, S23A, S82aN, V93A, and T108Q). BT32/A6.L2 is mutated at three additional positions with respect to BT32/A6.L1 (S74A, R83K and A84P). The mutations were hypothesized to eliminate the aggregation tendency of V_Hs (19, 42). Figure 2 shows a schematic representation of steps taken to construct the BT32/A6.L1 and BT32/A6.L2 mutants.

1. Generate and amplify by PCR two overlapping sub-fragments containing the above-mentioned BT32/A6.L1 mutations according to the following protocol (see Note 5):

dNTPs (2.5 mM each)	4 µL
Expand High Fidelity Buffer	5 µL
VHBACKMut1 (10 pmol/µL)	0.5 µL
VHFORFR3Mut1 (10 pmol/µL)	0.5 µL
pSJF2H-BT32/A6 template (1 ng/µL)	1 µL
Expand High Fidelity Enzyme mix (3.5 units/µL)	0.5 µL
sddH_2O	38.5 µL

   For the second sub-fragment, use VHBACKFR3Mut2/VHFORBamHIMut3 primer pair in the reaction mixture above.

   Place the reaction tubes in a thermocycler, and synthesize the two DNA sub-fragments with a program consisting of a preheating step at 94°C for 5 min, 30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s.

2. Apply 5 µL of the PCR products into a 1% agarose gel. The expected sizes for the two sub-fragments are 273 bp (B) and 153 bp (S). Purify the remaining PCR products (45 µL) with a QIAquick PCR purification™ kit (see Note 6), eluting the purified products in a volume of 30 µL. Determine the concentration of DNA based on OD_{260} measurements (41).

3. Splice the purified sub-fragments, and subsequently amplify the spliced products by SOE-PCR. To splice, prepare the following reaction mixture (see Note 7):

dNTPs (2.5 mM each)	4 µL
Expand High Fidelity Buffer	5 µL
Sub-fragment B	89 ng
Sub-fragment S	50 ng
Expand High Fidelity Enzyme mix (3.5 units/µL)	0.5 µL
Increase the volume to 50 µL with sddH_2O.	

Place the reaction tube in a thermocycler, and subject it to a program consisting of a preheating step at 94°C for 5 min, 10 cycles of 94°C for 30 s, 57°C for 30 s, and 72°C for 1 min.

Subsequently, to amplify the spliced products, heat the reaction mixture in the thermocycler at 94°C for 3 min, add 0.5 µL (5 pmol) of each of the two primers VHBACKBbsI and VHASSUF, and perform 30 PCR cycles as described in step 1.

4. Apply 5 µL of the SOE/PCR reaction mixture into a 1% agarose gel, and check for the presence of a single fragment of appropriate size (approximately 370 bp) (see Note 8). Purify the remaining PCR product with a QIAquick PCR purification™ kit.
5. Use VHBACKMut1/VHFORFR3Mut1a and VHBACKFR3-Mut2a/VHFORBamHIMut3 primer pairs in step 1, and follow steps 1–4 to construct BT32/A6.L2.
6. Proceed with the cloning of BT32/A6.L1 and BT32/A6.L2, as described in Subheading 3.2.

3.2. Cloning, Expression, Extraction, and Purification of V_Hs

In this step, BT32/A6.L1 and BT32/A6.L2 are cloned in an expression vector and transformed into an *E. coli* host employing standard cloning techniques described below. Positive clones are identified by colony-PCR and DNA sequencing. V_Hs are subsequently expressed in a periplasmic mode in fusion with a C-terminal His_6 tag. The presence of the His_6 tag allows for (1) detection of V_H expression by Western blotting and (2) one-step V_H purification by immobilized metal affinity chromatography (IMAC).

3.2.1. Cloning

1. Digest the purified V_H fragments and pSJF2H vector in separate reactions with restriction endonuclease *Bbs*I:

V_H fragment or pSJF2H vector	3 µg or 10 µg, respectively
10× *Bbs*I buffer	5 µL
*Bbs*I (10 units/µL)	3 µL
Increase the volume to 50 µL with sddH_2O.	

Incubate the reaction mixture at 37°C for 5–6 h.

2. Apply 5 µL of the reaction mixture into a 1% agarose gel to check for the success of the digestion. Purify the remaining DNA with a QIAquick PCR purification™ kit, and elute in a volume of 44 µL (the actual eluted volume is 42 µL). Digest with *Bam*HI:

Purified V_H fragment or pSJF2H vector	42 µL
10× *Bam*HI buffer	5 µL
*Bam*HI (10 units/µL)	3 µL

Incubate the reaction mixture at 37°C for 4 h.

3. Perform ligation in a total volume of 10 µL with 300 ng total DNA (1:3 molar ratio of vector to insert) using the LigaFast™ Rapid DNA Ligation System.
4. Transform 50 µL of electroporation-competent (TG1 cells) with 3 µL of the ligated material as described (43), using a MicroPulser™ electroporator. Transfer the electroporated cells to a tube containing 1 mL SOC medium, and incubate for 1 h at 37°C with shaking at 180 rpm.
5. Following transformation, spread 100 µL of cells on LB/agar plates containing 100 µg/mL ampicillin (LB/Amp). Incubate the plates inverted overnight at 32°C (see Note 9).
6. In the morning, perform colony-PCR in a total volume of 15 µL; prepare a master mix for 10 PCR reactions:

10× PCR buffer	16 µL
dNTPs (2.5 mM each)	13 µL
M13FP (10 pmol/µL)	3 µL
M13RP (10 pmol/µL)	3 µL
Taq DNA polymerase (5 units/µL)	1.6 µL
sddH_2O	123.4 µL

Aliquot 15 µL volumes from the master mix in 10 PCR tubes. Gently touch single colonies from the titer plates (step 5) with sterile toothpicks (or P10 pipette tips), and swirl in PCR tubes. Place the reaction tubes in a thermocycler, and perform PCR with a program consisting of a preheating step at 94°C for 5 min, 30 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min, and a final step of 72° for 7 min.

7. Apply 5 µL of the PCR mix into a 1% agarose gel to identify the clones with insert (approximately 600 bp). Confirm the positive clones by further sequencing their V_H genes using 0.5 µL of the PCR mixture and M13RP or M13FP as primers (see Note 10).
8. To have a permanent stock of positive clones, grow a single colony (step 5) in 2 mL LB/100 µg/mL ampicillin overnight at 200 rpm and 37°C. In the morning, centrifuge the cell

culture, discard the supernatant, resuspend the cells in LB/15% glycerol, and store at –80°C.

9. Perform protein expression, extraction, and purification as described (15).

3.2.2. Protein Expression and Extraction

V_H genes are cloned in fusion with the OmpA leader sequence, expressed and exported to the periplasm. The following extraction protocol based on an osmotic shock method (44) is designed to increase the permeability of the outer membrane and release V_Hs from the periplasm, without lysing the cells. Since the endogenous protein content of the periplasm is far less than that of the cytoplasm, the periplasmic extraction step results in partial V_H purification. It is recommended to keep the fractions from various steps of the extraction at 4°C until it is verified by Western blotting which fractions contain the V_H. The presence of c-Myc and His_6 tags at the C-terminus of V_Hs allows for V_H expression verification by Western blotting using commercially available anti-c-Myc or antiHis tag antibodies.

1. Use a single positive clone to inoculate 100 mL of M9/Amp. Incubate in a rotary shaker at 200 rpm for 24 h at 25°C.
2. Transfer 30 mL of the above preculture to 1 L of M9/Amp in a 4-L baffled flask. Incubate the culture at 200 rpm for 24 h at 25°C, subsequently supplement with 100 mL of 10× induction medium and 100 µL of 1 M IPTG, and incubate for another 48 h at 25°C.
3. Retain a small aliquot for Western blotting (see step 7), and centrifuge the remaining culture at 5,000 × *g* for 20 min at 4°C in a high-speed centrifuge. Keep the supernatant fraction at 4°C.
4. Resuspend the pellet in 150 mL wash solution. Centrifuge at 14,000 × *g* for 10 min at 4°C. Keep the supernatant fraction at 4°C.
5. Resuspend the pellet in 50 mL sucrose solution and incubate at room temperature for 10 min. Centrifuge at 14,000 × *g* for 45 min at 4°C. Keep the supernatant fraction at 4°C.
6. Resuspend the pellet in 50 mL ice-cold shock solution and incubate in an ice bath for 10 min. Centrifuge at 14,000 × *g* for 25 min at 4°C. Keep the supernatant fraction at 4°C.
7. Verify expression by detecting the presence of V_Hs in fractions from steps 3–6 by Western blotting against the c-Myc tag (45) using anti-c-Myc antibody (see Note 11). Pool the fractions that contain V_H and dialyze against 6 L of starting buffer overnight at 4°C using a dialysis membrane of 10 kDa MWCO.
8. Proceed with protein purification.

3.2.3. Purification

The presence of the C-terminal His_6 tag in V_Hs allows for one-step protein purification by IMAC using a 5-mL HiTrap™ Chelating HP column.

1. Charge the column with Ni^{2+} by applying 30 mL of a 5 mg/mL $NiCl_2.6H_2O$ solution, and subsequently wash the column with 15 mL deionized water.
2. Perform purification on an ÄKTA FPLC purification instrument according to the instructions provided by the manufacturer using the starting buffer, and elute bound protein using the elution buffer with a 10–500 mM imidazole gradient.
3. Examine the fractions corresponding to the "eluted" peaks on the chromatogram for the presence and purity of the V_Hs by SDS-PAGE (46). Pool the "V_H fractions" and dialyze extensively against sodium phosphate buffer A. Measure OD_{280} for determination of protein concentration from molar extinction coefficients (47) (see Note 12), add sodium azide at a final concentration of 0.02%, and store the V_Hs at 4°C.

3.3. Assessment of Aggregation State of V_Hs by Size Exclusion Chromatography

In this step, size exclusion chromatography employing Superdex™ 75 is used to assess the aggregation state of V_H domains. Nonaggregating V_Hs should give chromatograms with a single, symmetrical peak with elution volumes expected for a monomeric V_H (see Note 13). In contrast, the chromatogram profiles of aggregating V_Hs, in addition to the monomeric peaks, consist of additional peaks, which elute earlier (see Note 14). Percent monomer can be calculated by area integration of the peaks and used as a quantitative measure of V_H aggregation tendency (the higher the % monomer of a V_H, the lower its aggregation tendency).

1. Wash a Superdex™ 75 size exclusion column with 50 mL of filtered and degassed ddH_2O and subsequently equilibrate with 50 mL of PBS buffer at a pump speed of 0.5 mL/min (see Note 15).
2. Inject and elute 200 µL of purified V_H, and obtain the chromatogram (see Notes 16 and 17).
3. Integrate monomeric and aggregate peaks to obtain % monomer.
4. Figure 3 confirms that the incremental mutations significantly improved the solubility of BT32/A6. In contrast to BT32/A6, which forms multimeric and aggregated species, BT32/A6.L1 and BT32/A6.L2 are essentially devoid of such forms and exist primarily as monomers. The % monomer, which is 51% for the wild type V_H, increases significantly to 87% and 90%, for BT32/A6.L1 and BT32/A6.L2, respectively.

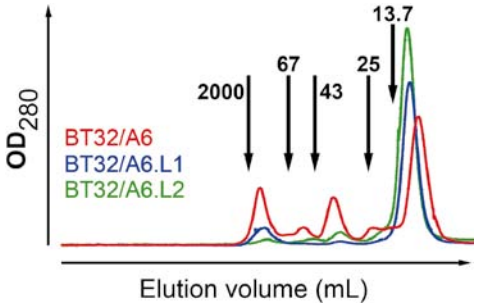


Fig. 3. Size exclusion chromatograms of BT32/A6, BT32/A6.L1, and BT32/A6.L2 V_Hs. The arrows denote the peak position of molecular weight markers BSA (M_r 67 kDa), ovalbumin (M_r 43 kDa), chymotrypsinogen A (M_r 25 kDa) and ribonuclease A (M_r 13.7 kDa) and blue dextran 2000 (M_r 2,000 kDa), which elutes at the void volume.

3.4. Determination of TRE of V_Hs

In this step, TRE, a quantitative measure of the extent to which a V_H resumes its native fold following heat denaturation, is determined by CD spectroscopy and SPR. Nonaggregating V_Hs are characterized by having a high TRE (16, 18, 23, 48). The high thermal refolding property of nonaggregating V_Hs not only extends their potential applications to needs that cannot be met by other antibody fragments (12, 20, 48) but also correlates with other useful properties such as high expression (18).

3.4.1. TRE Determination by CD Spectroscopy

The CD spectrum of a V_H is measured at its native state (30°C) and subsequently at its fully denatured state (85°C). The V_H is then allowed to cool down slowly to its "native temperature" (30°C), and its spectrum is measured again. TRE is calculated by the extent to which the CD spectrum of the heated-and-cooled V_H approaches that of the native one.

1. Dialyze V_Hs in sodium phosphate buffer B (see Note 18).
2. Refer to the manufacturer's instruction manual on how to operate the spectropolarimeter. Set the parameters for subsequent recording of CD spectra: wavelength, 215 nm–260 nm; wavelength interval, 0.2 nm; scan speed, 20 nm/min; bandwidth, 2 nm; integration time, 1 s. Choose a cuvette with appropriate pathlength (see Note 19).
3. Record the CD spectrum of the blank (empty cuvette) at 30°C with the spectropolarimeter connected to a waterbath.
4. Add the V_H (10 µg/mL) to the cuvette (see Note 20), and wait for 30 min for the temperature of the sample to equilibrate to 30°C; record the CD spectrum. This represents the CD spectrum of the native V_H.
5. Raise the temperature of the sample to 85°C to fully unfold the protein, leave at 85°C for 30 min for temperature equilibration, and record the CD spectrum (see Note 21).

6. Let the sample cool down to 30°C (approximately 3–4 h) and equilibrate as before; then, record the CD spectrum at 30°C.
7. Empty the cuvette, wash it 3× with 6 N HCl, and then rinse it 3× with ddH_2O and 3× with methanol. Dry by blowing nitrogen gas into the cuvette (see Note 22).
8. Subtract the "V_H spectra" from the blank spectrum, and subsequently smooth the spectra by Jasco software.
9. Calculate the TRE from the following formula

$$TRE = -\frac{R - U}{U - N} \times 100$$

where N and 1 represent the CD signals of the native and fully unfolded proteins at 30°C and 85°C, respectively, and R is the CD signal of the protein obtained at 30°C following its unfolding at 85°C (see Note 23). Calculate TRE as the average of at least two measurements.

10. Figure 4 shows the CD spectra of BT32/A6 and BT32/A6.L1. It is clear that BT32/A6.L1 has a higher TRE as it is heated-and-cooled since its CD spectrum is much closer to that for the native protein (compare elipticity values around 235 nm). In contrast, for BT32/A6, the heated and cooled CD spectrum almost superimposes on that for the unfolded V_H, indicating essentially irreversible thermal unfolding. The calculated TREs were 13% for the wild type, 44% for BT32/A6.L1, and 74% for BT32/A6.L2, a clear indication that the addition of the BT32/A6.L1 and BT32/A6.L2 mutation sets increases V_H refolding efficiency (see Note 24). Additionally, while a library based on BT32/A6 V_H scaffold was dysfunctional, the one based on BT32/A6.L1 was very stable and yielded V_Hs with good antigen binding activity and solubility (19).

3.4.2. TRE Determination by SPR

Alternatively, if the binding activity of a V_H is known, TRE can be determined by binding assays, e.g., enzyme-linked immunosorbent assay (49), SPR (16, 23). In the SPR approach, concentration analyses are performed to determine the amounts of active V_H following refolding. For the present V_Hs, however, the binding activity is not known, and thus, for the sake of providing a protocol for TRE determination by SPR, we resort to one with known binding activity, i.e., HVHAm304 V_H (16). The following protocol describes TRE measurements for HVHAm304 V_H, which has binding activity toward protein A.

1. Isolate monomeric V_Hs prior to SPR analysis by Superdex™ 75 size exclusion column chromatography. Equilibrate the column with 50 mL of HBS-EP buffer at a pump speed of 0.5 mL/min (see Note 25), inject 200 µL of IMAC-purified

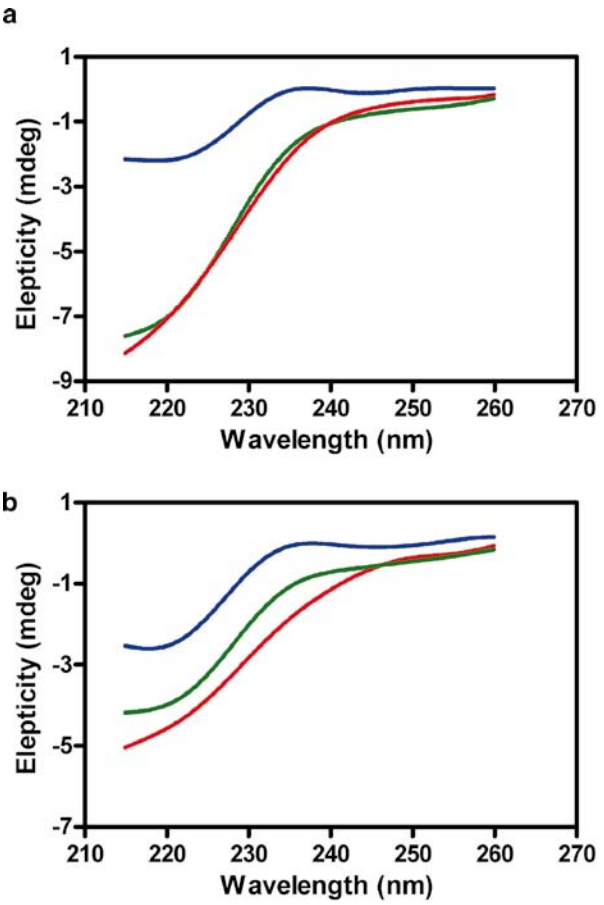


Fig. 4. CD spectra of native (*blue*), denatured (*red*) and heated-and-cooled (*green*) BT32/A6 (**a**) and BT32/A6.L1 (**b**). Elipticity values at 235 nm were used to calculate TREs, as described in Subheading 3.4.1.

V_H, and collect the monomer peak fraction. Measure protein concentration as described in Subheading 3.2.3, step 3.

2. Carry out SPR experiments at 25°C using a BIACORE 3000 instrument with HBS-EP as the running buffer.
3. Immobilize ligand (protein A in the present case) on a sensor-chip CM5 at a surface density that will give approximately 200 resonance units (RUs) of the theoretical surface capacity (see Note 26). Activate CM-dextran surface with a 7 min injection of a mixture of 50 mM NHS and 200 mM EDC at a flow rate of 5 µL/min. Inject 50 µg/mL ligand typically diluted in 10 mM acetate buffer, pH 4.5, for 3 min, and block the surface with a 7 min injection of 1 M ethanolamine, pH 8.5.
4. Prepare a standard curve for the active V_H concentration assay. Inject 10 or 20 µL of more than six different concentrations

of monomer V_H approximately in the range of 1/20 to 1/3 K_D (K_D = equilibrium dissociation constant) over both the ligand and reference surfaces at a flow rate of 40 µL/min (see Fig. 5a).

5. Analyze the data using BIAevaluation software 4.1. Plot steady state binding against the native V_H concentrations (see Fig. 5b). The linear range is used as a standard curve.

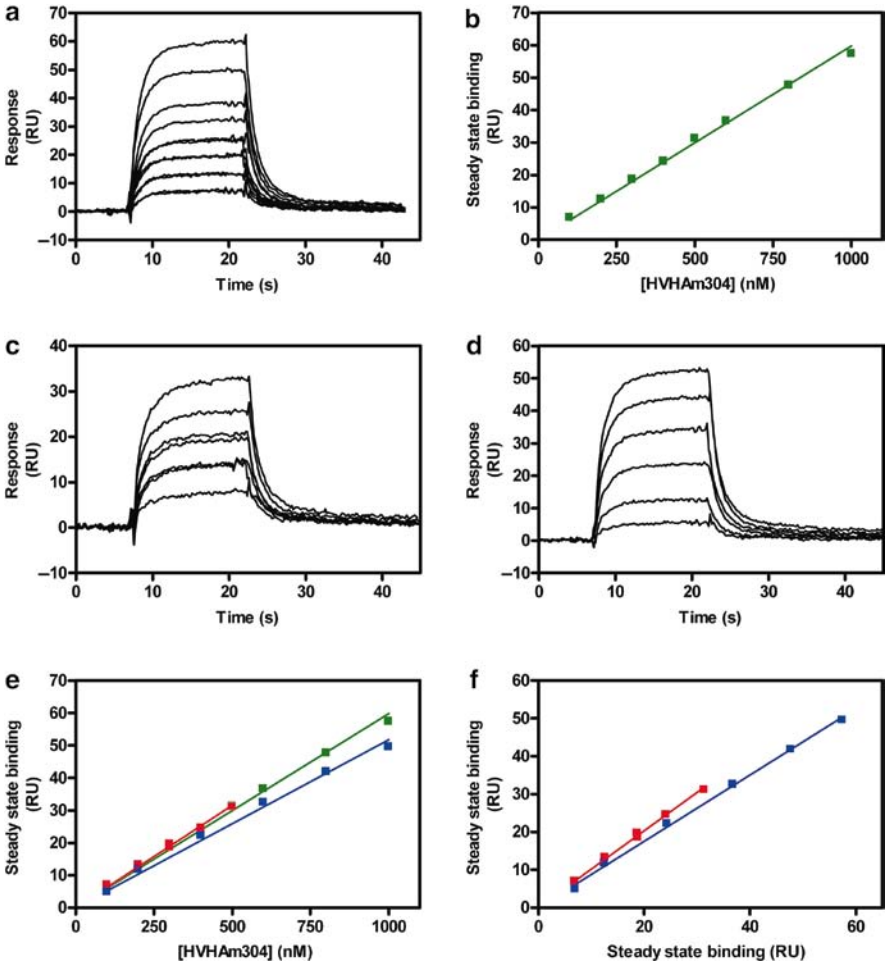


Fig. 5. TRE determination by SPR. (**a**) Sensorgram overlays of native HVHAm304 binding to immobilized protein A at concentrations of 100, 200, 300, 400, 500, 600, 800, and 1,000 nM. (**b**) Standard curve for native HVHAm304 steady state binding to protein A. (**c**, **d**) Sensorgram overlays for the binding of refolded HVHAm304 to protein A. (**c**) The protein was heat-denatured and cooled at 0.5 µM and passed over the sensorchip surface at concentrations of 100, 200, 300, 400, and 500 nM, based on the concentration of the unheated sample, with duplicate injections at 200 and 300 nM. (**d**) The protein was heat-denatured and cooled at 5 µM and passed over the sensorchip surface at concentrations of 100, 200, 400, 600, 800, and 1,000 nM, based on the concentration of the unheated sample. (**e**) Steady state binding of native (*green*) and refolded (*red and blue*) HVHAm304 at different concentrations, determined as described in **c** and **d**. Refolding was done at 0.5 (*red*) and 5 µM (*blue*) V_H concentrations. (**f**) Plots of the steady state binding to protein A of refolded (*y*-axis) versus native (*x*-axis) HVHAm304. Refolding was done at 0.5 (*red*) and 5 µM (*blue*) V_H concentrations. The slope values, m, are 1.0 (*red*) and 0.87 (*blue*).

6. Denature, at a concentration of 0.5 µM or higher, 50–200 µL V_Hs by incubation at 85°C for 20 min in a 250–600 µL plastic vial (see Note 27). Cool the V_Hs to room temperature for 30 min to refold, and subsequently centrifuge in a microfuge at 16,000 × *g* for 5 min at room temperature to remove any protein precipitate. Recover the supernatants, and determine the active V_H concentrations by measuring steady state binding.
7. Collect steady state binding data for the refolded V_H (see Fig. 5c, d). Plot steady state binding against native V_H concentrations (see Fig. 5e), and determine TRE:

$$\text{TRE} = \frac{m_{\text{refolded}}}{m_{\text{native}}} \times 100$$

where m_{refolded} is the slope of the graph for the refolded V_H and m_{native} is the slope of the graph for the native V_H.

When the original concentrations of refolded V_H are the same as those used for the standard curve, plot the steady state bindings of refolded against native ones (see Fig. 5f), and determine TRE:

$$\text{TRE} = m \times 100$$

where m is the slope of the graph.

8. Figure 5 shows the results of TRE experiments by SPR for HVHAm304 V_H at 0.5 and 5 µM concentrations. Native and refolded V_Hs were passed over the protein A surface at the indicated concentrations; higher concentrations were possible with native V_H and V_H refolded at 5 µM than with V_H refolded at 0.5 µM. The levels of steady state binding for the native versus the two refolded preparations were plotted at identical concentrations (Fig. 5f), with the slopes of the plots, m, indicating the refolding efficiency. The slope values are 1.0 and 0.87 translating to TRE values of 100% and 87% for refolding experiments at 0.5 and 5 µM V_H concentrations, respectively.

4. Notes

1. Use this DNA polymerase only for performing colony-PCR.
2. Use this for preparing sterile ampicillin and IPTG solutions.
3. Use the filtration system for filtering ddH_2O, buffers and solutions of sucrose, glucose, vitamin B1, and casamino acids.
4. This buffer can be purchased from GE Healthcare. The buffer should be filtered and thoroughly degassed before use.

5. In order to avoid introducing any unwanted mutations in V_H fragments, use a DNA polymerase with proof reading activity in all PCR experiments in Subheading 3.1, e.g., Expand High Fidelity Enzyme mix (*Hoffmann-La Roche Ltd.*).
6. The PCR products contain pSJF2H-BT32/A6 vector which, if not removed in a gel-purification step, will get carried over to the SOE/PCR step (step 3), and its embedded wild type V_H will be preferentially amplified over the spliced products (mutant V_Hs). We, however, avoided the gel-purification step, which is both tedious and low-yielding, by using a primer (VHFORBamHIMut3), which introduced a characteristic sequence at the 3′ end of the small sub-fragments (see Fig. 2). In the subsequent SOE/PCR step, the use of VHASSUF, which anneals at the characteristic sequence site present only in the spliced template, led to the amplification of spliced products and not the wild type V_H.
7. The amount of sub-fragments in the SOE/PCR reactions should be adjusted to give equimolar concentrations of the two sub-fragments.
8. Normally, the SOE/PCR step results in a single band of appropriate size and gives good yield. However, if the yield of amplified product is low, the amount of input sub-fragments in the PCR reaction mixture may need to be increased.
9. Avoid incubating LB/Amp plates at 37°C as it results in the appearance of TG1 satellite colonies.
10. With typical PCR yields, a 0.5 µL aliquot in typical sequencing reaction volumes gives clean sequencing profiles. In instances where the sequencing profile is not readable, purify the PCR product with a QIAquick PCR purification™ kit before proceeding with DNA sequencing.
11. Alternatively, commercially available antiHis antibodies can be used for detection, e.g., AntiHis Antibody from *GE Healthcare*. The V_Hs are frequently found in both the "shock" and "sucrose" fractions.
12. Commercial protein analysis softwares, e.g., Laser gene v6.0, DNASTAR, Inc., or online freewares, e.g., ExPASy ProtParam tool at http://us.expasy.org/tools/protparam.html, can be used for calculation of molar extinction coefficient.
13. There are typically some variations in the elution volumes of monomeric fractions (16, 18).
14. Unlike the monomeric peaks for the nonaggregating V_Hs, those for aggregating V_Hs may not be symmetrical but display tailing, which could be due to nonspecific interaction of the V_H with the column material.
15. Equilibrate the column with HBS-EP buffer, if proceeding to analyze the V_Hs by SPR (see Subheading 3.4.2, step 1).

16. As low as 50 µL of protein preparation may be sufficient to give a reliable signal (a mAU peak value of around 50 or higher). When the peak mAU values are too low, particularly when lower than 10, a reliable statement with regard to the V_H aggregation states cannot be made.
17. Use the same concentration when comparing different V_Hs as protein concentration is a factor in aggregation formation.
18. Avoid using buffers with high absorbance in the wavelength range used (e.g., buffers with high salt concentration) as they would increase the noise. This is particularly important for V_Hs as they are essentially β-strand proteins and as a result give only a modest CD signal change upon denaturation.
19. The choice of cuvette pathlength depends on the working V_H concentrations. Cuvettes with longer pathlengths (larger volumes) may need to be used with lower concentrations in order to obtain a significant CD signal.
20. The choice of V_H concentration may vary depending on the available concentration and the aggregation tendency of individual V_Hs. Lower concentrations should be used if high concentrations lead to V_H precipitation upon unfolding (CD values obtained from precipitated samples are not reliable). Conversely, higher V_H concentrations need to be used if no significant difference between the refolding efficiencies of the V_Hs is observed at lower V_H concentrations.
21. Always check the cuvettes at the end of the run to ensure that the sample solution is clear, an indication that the sample has not precipitated upon unfolding. Precipitated samples show reduced CD signals.
22. By the end of the run, a significant amount of protein may stick to the walls of the cuvette, resulting in a significantly higher background signal in the next round of CD measurements. Thus, if feasible, it is a good practice to soak the cuvettes in the 6 N HCl overnight and subsequently rinse them with water and methanol.
23. With V_Hs, elipticity at 235 nm has been used in a number of publications for TREs as well as melting temperatures (18, 19, 50). However, if other wavelengths, e.g., 205–215 nm, lead to a higher signal difference between the native and denatured V_H, they should preferentially be used.
24. A higher V_H concentration (10 µg/mL) was required to discriminate between V_Hs in terms of TREs. At 2 µg/mL, the TRE values of the V_Hs significantly increased, and BT32/A6.L1 and BT32/A6.L2 had essentially the same TREs.
25. Equilibrate the column with the identical batch of HBS-EP buffer to be used for SPR analysis. This reduces bulk change effects.

26. Theoretical surface capacity = surface density × (MW of V_H/ MW of ligand per binding site).
27. Higher V_H concentrations may need to be used in order to distinguish V_Hs in terms of refolding efficiencies.

Acknowledgments

We are indebted to Tomoko Hirama, for helping us prepare the manuscript, and to Tom Devecseri, for preparing publication quality figures.

References

1. Carter P (1986) Site-directed mutagenesis. Biochem J 237:1–7
2. Costa GL, Bauer JC, McGowan B, Angert M, Weiner MP (1996) Site-directed mutagenesis using a rapid PCR-based method. Methods Mol Biol 57:239–248
3. Crameri A, Cwirla S, Stemmer WP (1996) Construction and evolution of antibody-phage libraries by DNA shuffling. Nat Med 2:100–102
4. Jackel C, Kast P, Hilvert D (2008) Protein design by directed evolution. Annu Rev Biophys 37:153–173
5. Ling MM, Robinson BH (1997) Approaches to DNA mutagenesis: an overview. Anal Biochem 254:157–178
6. Lipovsek D, Plückthun A (2004) *In-vitro* protein evolution by ribosome display and mRNA display. J Immunol Methods 290: 51–67
7. Wark KL, Hudson PJ (2006) Latest technologies for the enhancement of antibody affinity. Adv Drug Deliv Rev 58:657–670
8. Yuan L, Kurek I, English J, Keenan R (2005) Laboratory-directed protein evolution. Microbiol Mol Biol Rev 69:373–392
9. Yuen CM, Liu DR (2007) Dissecting protein structure and function using directed evolution. Nat Methods 4:995–997
10. Davies J, Riechmann L (1995) Antibody VH domains as small recognition units. Biotechnology (N Y) 13:475–479
11. Holliger P, Hudson PJ (2005) Engineered antibody fragments and the rise of single domains. Nat Biotechnol 23:1126–1136
12. Holt LJ, Herring C, Jespers LS, Woolven BP, Tomlinson IM (2003) Domain antibodies: proteins for therapy. Trends Biotechnol 21:484–490
13. Revets H, De Baetselier P, Muyldermans S (2005) Nanobodies as novel agents for cancer therapy. Expert Opin Biol Ther 5:111–124
14. Saerens D, Ghassabeh GH, Muyldermans S (2008) Single-domain antibodies as building blocks for novel therapeutics. Curr Opin Pharmacol 8:600–608
15. Arbabi-Ghahroudi M, MacKenzie R, Tanha J (2009) Selection of non-aggregating V_H binders from synthetic V_H phage display libraries. Methods Mol Biol 525:187–216, xiii
16. Arbabi-Ghahroudi M, To R, Gaudette N, Hirama T, Ding W, MacKenzie R, Tanha J (2009) Aggregation-resistant V_Hs selected by *in vitro* evolution tend to have disulfide-bonded loops and acidic isoelectric points. Protein Eng Des Sel 22:59–66
17. Chen W, Zhu Z, Feng Y, Xiao X, Dimitrov DS (2008) Construction of a large phage-displayed human antibody domain library with a scaffold based on a newly identified highly soluble, stable heavy chain variable domain. J Mol Biol 382:779–789
18. Jespers L, Schon O, Famm K, Winter G (2004) Aggregation-resistant domain antibodies selected on phage by heat denaturation. Nat Biotechnol 22:1161–1165
19. Tanha J, Nguyen TD, Ng A, Ryan S, Ni F, MacKenzie R (2006) Improving solubility and refolding efficiency of human V_Hs by a novel mutational approach. Protein Eng Des Sel 19:503–509
20. Christ D, Famm K, Winter G (2007) Repertoires of aggregation-resistant human antibody domains. Protein Eng Des Sel 20:413–416

21. Famm K, Hansen L, Christ D, Winter G (2008) Thermodynamically stable aggregation-resistant antibody domains through directed evolution. J Mol Biol 376:926–931
22. Tanha J, Xu P, Chen ZG, Ni F, Kaplan H, Narang SA, MacKenzie CR (2001) Optimal design features of camelized human single-domain antibody libraries. J Biol Chem 276:24774–24780
23. To R, Hirama T, Arbabi-Ghahroudi M, MacKenzie R, Wang P, Xu P, Ni F, Tanha J (2005) Isolation of monomeric human V_Hs by a phage selection. J Biol Chem 280: 41395–41403
24. Hemsley A, Arnheim N, Toney MD, Cortopassi G, Galas DJ (1989) A simple method for site-directed mutagenesis using the polymerase chain reaction. Nucleic Acids Res 17:6545–6551
25. Higuchi R, Krummel B, Saiki RK (1988) A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res 16:7351–7367
26. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51–59
27. Nelson RM, Long GL (1989) A general method of site-specific mutagenesis using a modification of the *Thermus aquaticus* polymerase chain reaction. Anal Biochem 180:147–151
28. Vallette F, Mege E, Reiss A, Adesnik M (1989) Construction of mutant and chimeric genes using the polymerase chain reaction. Nucleic Acids Res 17:723–733
29. Coulon S, Pellequer JL, Blachere T, Chartier M, Mappus E, Chen SW, Cuilleron CY, Baty D (2002) Functional characterization of an anti-estradiol antibody by site-directed mutagenesis and molecular modeling: modulation of binding properties and prominent role of the V_L domain in estradiol recognition. J Mol Recognit 15:6–18
30. Dilsiz N, Crabbe MJ (1994) A high-yield modification of mutation by overlap extension using three primers. Anal Biochem 222:510–511
31. Dolk E, van der Vaart M, Lutje HD, Vriend G, de Haard H, Spinelli S, Cambillau C, Frenken L, Verrips T (2005) Isolation of llama antibody fragments for prevention of dandruff by phage display in shampoo. Appl Environ Microbiol 71:442–450
32. Hulett MD, Witort E, Brinkworth RI, McKenzie IF, Hogarth PM (1994) Identification of the IgG binding site of the human low affinity receptor for IgG Fc gamma RII. Enhancement and ablation of binding by site-directed mutagenesis. J Biol Chem 269: 15287–15293
33. Kusters-van Someren M, Kishi K, Lundell T, Gold MH (1995) The manganese binding site of manganese peroxidase: characterization of an Asp179Asn site-directed mutant protein. Biochemistry 34:10620–10627
34. Peng RH, Xiong AS, Yao QH (2006) A direct and efficient PAGE-mediated overlap extension PCR method for gene multiple-site mutagenesis. Appl Microbiol Biotechnol 73:234–240
35. Salins LL, Ware RA, Ensor CM, Daunert S (2001) A novel reagentless sensing system for measuring glucose based on the galactose/glucose-binding protein. Anal Biochem 294:19–26
36. Steenbergen SM, Vimr ER (2003) Functional relationships of the sialyltransferases involved in expression of the polysialic acid capsules of *Escherichia coli* K1 and K92 and *Neisseria meningitidis* groups B or C. J Biol Chem 278:15349–15359
37. Walker SL, Wonderling RS, Owens RA (1997) Mutational analysis of the adeno-associated virus Rep68 protein: identification of critical residues necessary for site-specific endonuclease activity. J Virol 71:2722–2730
38. Waschutza G, Dengler U, Villmann C, Bottinger H, Otto B (1998) Interferon-gamma variants with deletions in the AB surface loop. Flexibility is a critical point for receptor binding. Eur J Biochem 256:303–309
39. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene 77:61–68
40. Dan MD, Earley EM, Griffin MC, Maiti PK, Prashar AK, Yuan XY, Friesen AD, Kaplan HA (1995) Human monoclonal antibody BT32/A6 and a cell cycle-independent glioma-associated surface antigen. J Neurosurg 82:475–480
41. Sambrook J, Fritsch EF, Maniatis T (eds) (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
42. Tanha J, Dubuc G, Hirama T, Narang SA, MacKenzie CR (2002) Selection by phage display of llama conventional V_H fragments with heavy chain antibody V_HH properties. J Immunol Methods 263:97–109
43. Tung WL, Chow KC (1995) A modified medium for efficient electrotransformation of *E. coli*. Trends Genet 11:128–129
44. Anand NN, Dubuc G, Phipps J, MacKenzie CR, Sadowska J, Young NM, Bundle DR,

Narang SA (1991) Synthesis and expression in *Escherichia coli* of cistronic DNA encoding an antibody fragment specific for a *Salmonella* serotype B O-antigen. Gene 100:39–44

45. MacKenzie CR, Sharma V, Brummell D, Bilous D, Dubuc G, Sadowska J, Young NM, Bundle DR, Narang SA (1994) Effect of Cλ-Cκ domain switching on Fab activity and yield in *Escherichia coli*: synthesis and expression of genes encoding two anti-carbohydrate Fabs. Biotechnology (N Y) 12:390–395
46. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685
47. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. Protein Sci 4:2411–2423
48. Jespers L, Schon O, James LC, Veprintsev D, Winter G (2004) Crystal structure of HEL4, a soluble, refoldable human V_H single domain with a germ-line scaffold. J Mol Biol 337:893–903
49. van der Linden RHJ, Frenken LGJ, de Geus B, Harmsen MM, Ruuls RC, Stok W, de Ron L, Wilson S, Davis P, Verrips CT (1999) Comparison of physical chemical properties of llama V_{HH} antibody fragments and mouse monoclonal antibodies. Biochim Biophys Acta 1431:37–46
50. Davies J, Riechmann L (1996) Single antibody domains as small recognition units: design and *in vitro* antigen selection of camelized, human VH domains with improved protein stability. Protein Eng 9:531–537

Chapter 23

Phenotype Based Functional Gene Screening Using Retrovirus-Mediated Gene Trapping in Quasi-Haploid RAW 264.7 Cells

Sung Ouk Kim and Soon-Duck Ha

Abstract

In vitro random mutagenesis, followed by phenotype screening, provides a rapid and convenient tool for identifying novel genes involved in the phenotype of interest. However, the forward mutagenic approach in mammalian somatic cells is seriously limited by the diploidic nature of the genome. To overcome this impediment, we developed a method that allows functional screening for both haploid insufficient and sufficient genes involved in the phenotype of interest, utilizing a retrovirus gene trap mutagenesis in chemical mutagen-generated quasi-haploid cells. This method was used to identify novel host genes that are required for macrophage sensitivity to anthrax lethal toxin.

Key words: Retrovirus, Gene-trapping, Forward genetics, Macrophages, Cell death

1. Introduction

In vitro random mutagenesis, followed by phenotype screening, provides a rapid and convenient tool for identifying novel genes involved in the phenotype of interest (1). However, since only a single copy of each autosomal gene is destroyed by mutagens, this genetic approach is limited in its application for mammalian cells due to the diploid nature of the genome. In embryonic stem cells, biallelic mutations can be achieved through generating mice from embryonic stem cells and breeding (2, 3). However, considering costs, time, and effort required for screening in mice, in vitro screening approaches are preferred when addressing questions at cellular levels. In embryonic stem cells, loss-of-heterozygosity in vitro can be enhanced by promoting mitotic recombination

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_23,

through the Cre/loxP strategy (4, 5) or using a Bloom's syndrome gene (*Blm*) deficient background (6, 7). However, the Cre/loxP approach in somatic cells will complicate interpretation of any cell phenotypes because of choromosomal translocations induced by Cre expression (8). BLM is a family of RecQ helicase, which is crucial for genome stability, and mutant embryonic cells show an increased rate of sister chromatin exchanges and about 20-fold higher frequency rates of loss-of-heterozygosity (i.e., $2.3–4.2 \times 10^{-4}$) than wild-type cells. Approaches such as knocking down *Blm* using small interference RNAs may enhance loss-of-heterozygosity of a mutation. However, its practicality and feasibility in somatic cells are yet to be determined.

Mutagenesis can be induced by either chemical or insertional mutagens. Chemical mutagens such as ethyl-nitroso-urea (ENU) are potent mutagens causing high mutation frequency with multiple exposures (9); however, this approach requires extensive effort to identify genes whose mutations are responsible for a phenotype. In contrast, insertional mutagens, such as gene trapping retroviral vectors or transposons, allow rapid identification of the targeted genes through molecular tags from mutagens. However, their mutation frequency is generally low (10).

In principle, insertional mutagenesis would be very effective if haploid mammalian cells could be created in culture. Random mutations by powerful chemical mutagens can then render a pool of cells that can be regarded as quasi-haploid. Here, we describe a method that utilizes the power of chemical mutagens for generating quasi-haploid murine macrophages, followed by insertional mutagenesis using retroviral gene trapping vectors, to uncover genes involved in anthrax lethal toxin (LeTx)-induced cell death (11).

2. Materials

2.1. Cell Culture, Chemical Mutagenesis, and Hypoxanthine Phosphoribosyl Transferase Assay

1. Dulbecco's modified Eagle's medium (DMEM; Invitrogen) with high glucose supplemented with heat-inactivated 10% fetal bovine serum (FBS; Sigma), 10 mM MEM nonessential amino acids (Invitrogen), 1 mM sodium pyruvate, and 100 U/ml penicillin G sodium/streptomycin sulfate (Invitrogen).
2. All cells are cultured in a cell culture incubator at 37°C in 5% CO_2, unless otherwise indicated.
3. *N*-ethyl-*N*-nitrosourea (ENU; Sigma) is dissolved at 35 mg/ml (stock solution) in equal portions of 95% ethanol and phosphate-citrate buffer: 50 mM sodium citrate, 100 mM sodium phosphate, pH 5.0–6.0, and stored at –80°C in small aliquots.

4. O^6-benzylguanine (O^6-BG; Sigma) is dissolved in dimethyl sulfoxide (DMSO) at 20 mM (stock solution) and stored at −80°C in small aliquots.
5. 6-thioguanine (6-TG; Sigma-Aldrich) is dissolved at 1 mM in phosphate-buffered saline and stored at −80°C in small aliquots.
6. 100 mm cell culture dish.
7. 6-well and 96-well plates.
8. Phosphate Buffered Saline (10× PBS; Invitrogen).

2.2. Cell Viability Assay

1. Protective antigen (PA) and lethal factor (LF) are obtained from List Biological Laboratories (Campbell, CA, USA) and reconstituted in sterile distilled water containing 1 mg/mL of bovine serum albumin (BSA) at a concentration of 1 mg/mL, and are stored at −80°C in small aliquots.
2. Crystal violet (Sigma) solution is made up 0.5% (w/v) with 25% methanol in PBS and stored at room temperature (stable at room temperature).

2.3. Retrovirus Preparation and Infection

1. Phoenix Amphotropic cells (see Note 1) for virus amplification are cultured in the media (see Subheading 2.1, item 1). Maintain cell culture less than 90% in cell density.
2. pDisrup8 (see Note 2) and pCI-VSVG (see Note 3) are prepared in high purity using CsCl density gradient centrifugation, as described in detail in (12).
3. Polybrene (Sigma) is dissolved at 8 mg/mL (stock) in cell culture grade PBS and stored at −20°C in small aliquots.
4. Dexamethasone (Sigma) is dissolved in DMSO at 1 mM (stock) and stored at −20°C in small aliquots.
5. Sodium butyrate (EMD Chemicals) is dissolved in PBS at 1 M (stock) and stored at −20°C in small aliquots.
6. Blasticidin (EMD Chemicals) is dissolved in PBS at concentration 5 mg/mL (stock) and stored at −20°C in small aliquots.
7. Cell culture grade water.
8. 0.45 μm filter and syringe.

2.4. 3′-Rapid Amplification of cDNA Ends (RACE) and Sequencing

1. TRIzol reagent (Invitrogen; see Note 4).
2. Primers for 3′-RACE (see Subheading 3.4) are dissolved in DEPC-water at 10 μM (stock) and stored at −20°C.
3. Reverse transcription buffer: 10×; 500 mM Tris–HCl, pH8.3, 750 mM KCl, 30 mM $MgCl_2$, 10 mM DTT.

4. dNTP mixture: 10 mM each of dATP, dCTP, dGTP, and dTTP.
5. MMLV reverse transcriptase (New England Biolabs; NEB) is stored at –20°C.
6. *Taq*DNA polymerase (NEB) is stored at –20°C.
7. Gel extraction kit (MO Bio Laboratories, Carlsbad, CA, USA).

2.5. Identification of Mutations Induced by ENU

1. RT-PCR reagents including Oligo(dT)17 primer, MMLV reverse transcriptase (NEB), and *Pfu* polymerase (Promega).
2. Primers covering full length of a gene of interest.
3. Gel extraction kit (MO Bio Laboratories, CA, USA).

3. Methods

3.1. Generation of Quasi-Haploid RAW 264.7 Cells (See Note 5)

1. RAW 264.7 macrophages (ATCC) of 50% confluence in a 100 mm dish are treated with O^6-BG (10 μM) for 12 h, and ENU (0.35 mg/mL) is then added for the next 2 h. The cells are washed and incubated with fresh medium containing O^6-BG for another 24 h. Approximately 1,000 well-dispersed cells are plated on a 100 mm dish for mutation frequency analysis. The rest of the cells are used for the next round of mutagenesis, repeating the same procedures.
2. To estimate cumulative mutation frequencies of each ENU treatment, 96 individual clones are picked from the 100-mm dish 10–14 days after plating and added to separate wells of a 96-well plate and grown for 24 h.
3. Cells in each well are then exposed to 10 μM 6-TG for 7 days, and 6-TG-resistant clones are scored. The number of surviving clones indicates mutation frequency at the hypoxanthine phosphoribosyl transferase (*Hprt*) locus, which renders loss-of-function. The *hprt* gene product catalyzes transformation of purine analog 6-TG to a cytotoxic metabolite. Cells with deleterious mutation in the *hprt* cannot phosphoribosylate the analog and survive in the presence of 6-TG (see Fig. 1 and Note 6).

3.2. Selecting Anthrax Lethal Toxin Sensitive Clones

1. After the sixth round of ENU treatments, the frequency of loss-of-function mutation in the *Hprt* locus reaches about 0.1. About 1,000 well-dispersed cells are plated in a 100 mm dish, cultured for about 2 weeks, and 50–150 medium size individual cell colonies are replated into individual wells of 96-well plates.

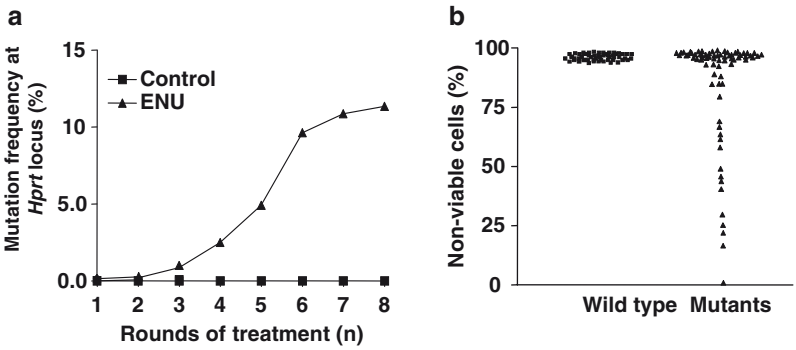


Fig. 1. Determination of ENU mutagenesis frequency and selection of lethal toxin-sensitive clones. (**a**) Mutation frequency in RAW264.7 cells treated with ENU. Cells were treated with different rounds of ENU (0.35 mg/ml) for 2 h. The mutation frequency was determined by calculating the percentage of cells that had a recessive mutation in the *Hprt* gene. (**b**) The sensitivity to lethal toxin-induced cell death of individual clones treated with or without ENU. The viability of each individual clone of 50 parental and 72 chemically mutated clones treated with 1 μg/mL LF and PA was measured (Reproduced from ref. 11 © 2007 BioTechniques. Used by permission).

2. Each clone in 96-well plates is then divided into triplicates of 96-well plates. To select LeTx-sensitive clones, cells in one of the plates are treated with anthrax lethal toxin (LeTx: 1 μg/mL of PA and LF) for 5 h.
3. Cell culture media are removed, and plates are immersed in 2 L of PBS. After washing the plates three times, cells are gently tapped on the top of paper towels to remove excess PBS and stained with 100 μl/well of 0.5% (w/v) crystal violet in PBS containing 25% (v/v) methanol for 5 min.
4. Excess crystal violet is removed by immersing plates in fresh PBS three times.
5. 100 μL of 50% (v/v) acetic acid in distilled water is added to each well to dissolve crystal violet stained cells.
6. Optical densities of the wells are analyzed using an automatic microtiter plate reader (UV MAX, Molecular Devices, Sunnyvale, CA) at wavelength of 590 nm. Live cells are attached to the plate and stained by crystal violet, and the higher optical density (dark blue) represents the most cells that survived. Cell death is calculated based on the optical density of wells from nontreated cells as 100% survival.
7. LeTx-sensitive clones are culled from one of the triplicate 96-well plates for subsequent retrovirus gene trapping (see Fig. 2).

3.3. Estimating Number of ENU Mutant Clones Required for Obtaining a Pool of Quasi-Haploid Culture

The number of cells that are needed to create a pool of cells with loss-of-function mutations in one allele of most or all genes depends upon the percentage of monoallelic mutations introduced within each cell. Given that six rounds of ENU renders random mutations of 10% of genes in each cell (see Fig. 1), a pool of 40 cells should contain monoallelic mutations of 98.5% of all

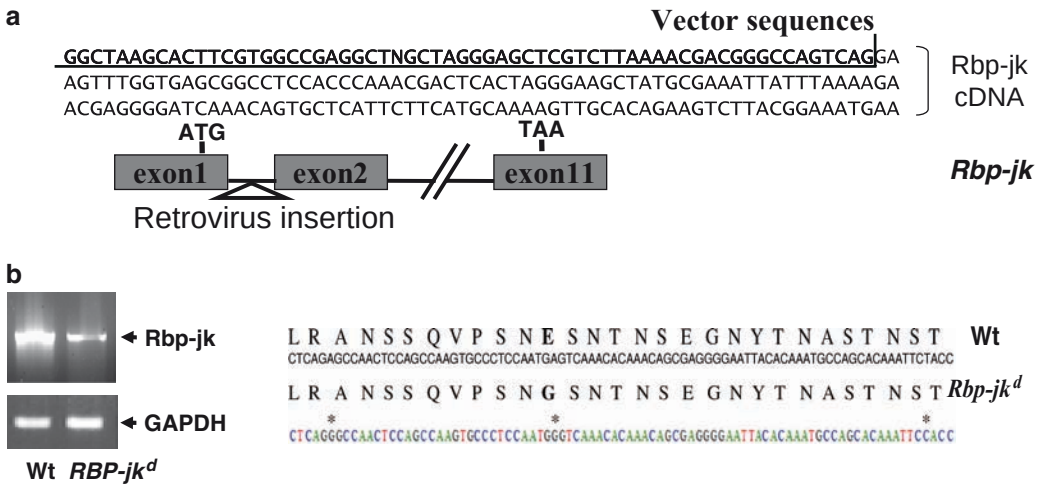


Fig. 2. The fused mRNA of blasticidin resistant gene (*blasticidin*$^{+}$) and an endogenous gene in a LeTx-resistance RAW264.7 clone were amplified by 3′-RACE and analyzed by DNA sequencing. (**a**) *Blasticidin*$^{+}$ encoded by retroviral vector was fused with *Rbp-jk* mRNA. The insertion occurred in an intron between exon 1 and exon 2, which results in a disruption of one allele of *Rbp-jk*. (**b**) RT-PCR was performed using a primer pairing with exon 1 sequence of *Rbp-jk* and another targeting the sequence around stop codon of *Rbp-jk*. The PCR product derived from mRNA of wild-type and Rbp-jk^{d} cells were analyzed on an agarose gel and sequenced. Stars show the mutations sites (Reproduced from ref. 11 © 2007 BioTechniques. Used by permission).

genes. In other words, the probability of finding a gene mutated at one allele (P) within the 40 culled cell clones is 0.985.

$P = 1 - (1 - X)^{n}$, where X = mutation rate, n = number of cells required to be culled.

3.4. Retrovirus Insertion-Induced Mutation in Quasi-Haploid Cells

1. Retroviruses containing pDisrup8 are produced in Phoenix Amphotropic cells using the calcium phosphate transfection method, as described in detail in (13). pDisrup8 retroviral vector(15 μg) and VSVG (5 μg) plasmids are transfected into Phoenix cells plated on 100 mm culture dish at about 70% confluence and incubated for 16–20 h.
2. Remove media from Phoenix Amphotropic cells, add fresh culture media containing 1 μM dexamethasone and 1 mM sodium butyrate, and further incubate at 32°C and 5% CO_2 for another 24 h.
3. Retrovirus-containing media are collected and filtered through a 0.45-μm filter to remove cells or cell debris.
4. Quasi-haploid RAW 264.7 cells (40 LeTx-sensitive clones culled in Subheading 3.2) are plated in a 6-well plate at 5×10^{5} cells/well 16 h prior to exposing to retrovirus.
5. Remove quasi-haploid RAW 264.7 cell media, and add retrovirus prepared in step 3.

6. Add polybrene (4 μg/ml at final concentration) to enhance retroviral infection efficiency.
7. Centrifuge the plate at 2,000 rpm for 45 min at room temperature.
8. Transfer the plate to a cell culture incubator and incubate for 16 h.
9. Cells in each well are transferred into a new 100 mm culture dish and incubated in a cell culture incubator for another 24 h.
10. Blasticidin resistant RAW 264.7 cells are then selected by treating cells with blasticidin (10 μg/mL) for 7 days. Quasi-haploid RAW 264.7 (nonretrovirus exposed) cells are used as a negative control for blasticidin resistance.
11. After counting the number of blasticidin resistant colonies, these cells are pooled.
12. The number of blasticidin resistant colonies is the minimum number of cells required to represent all the colonies to be tested for LeTx sensitivity. However, to minimize the chance of plating multiple numbers of cells originated from the same colony, cells at twice the number of colonies are replated in a 15 cm plate in the presence of LeTx and blasticidin (10 μg/mL). Therefore, in an ideal condition, the plate should contain two cells representing one colony.
13. After 2 weeks, each survival cell clone is picked using a 20 μl pipette and plated into 96-well plates (see Note 7).
14. Each isolated clone is rechallenged with LeTx to confirm LeTx-resistance.

3.5. Identifying and Confirming of Genes Disrupted by Retrovirus

1. Total RNAs are isolated from each LeTx-resistant clone (~1.5×10^6 cells) using the TRIzol reagent following manufacturer's instructions.
2. The 1–5 μg of total RNAs are denatured along with 20 pmoles of 3′-RACE cDNA synthesis primer (3′-CDS, 10 μM, 5′-CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGC $[T]_{17}$-3′) at 70°C for 2 min, and then kept on ice for 5 min.
3. 3′-RACE PCR products are amplified first by P1/Q1 primers using first strand cDNA, followed by a nested PCR using P2/Q2 primers and the first PCR product (0.5 μL). P1 and P2 primers are located in the blasticidin resistance gene, and Q1 and Q2 primers are the nested universal primers on the 3′-CDS as the anchor sequence. [P1; 5′-AAAGCGATAGTGAAGGACAGTG A-3′,Q1;5′-CCAGTGAGCAGAGTGACG-3′][P2;5′-TGCTGCC CTCTGGTTATGTGTGG-3′,Q2;5′-GAGGACTCGAGCTCAA GC-3′]. The PCR reaction is performed on the scale of 50 μL using the *Taq* DNA polymerase (NEB). Conditions for the

1st PCR with P1 and Q1 primers are 2 min at 94°C for initial denature, followed by 30 cycles of at 94°C for 30 s (denature), 68°C for 30 s (annealing) and 72°C for 3 min (extension), and final extension at 72°C for 10 min. Conditions for the 2nd PCR with P2 and Q2 primers are the same as the 1st PCR except performing amplification using annealing temperature at 64°C.

4. 3′-RACE (step 2–3) can be performed using commercial products such as SMART™ RACE cDNA amplification kit (Clontech Lab.)
5. 3′-RACE PCR products are run on an agarose gel and purified by gel extraction kit. Purified DNAs are used for sequencing using P2 primer or the blasticidin resistance gene specific primer downstream of the P2 sequence for identifying genes disrupted by retrovirus.
6. The cDNA sequences fused with blasticidin-resistant genes are aligned with GenBank® database using Basic Local Alignment Search Tool (BLAST).
7. An example of a gene targeted by retrovirus and the cDNA sequences are shown in Fig. 2a.

3.6. Identifying Mutations Induced by ENU

Once the retrovirus insertion-induced mutation responsible for the phenotype is identified, ENU-induced mutagenesis on the other allele can be revealed through sequencing cDNA of the gene identified. To specifically amplify the mRNA transcribed from the allele that is not hit by retrovirus, cDNA should be amplified by using primers, the products of which cover the full-length mRNA, or by using 5′ primers upstream of retrovirus insertion site. The PCR product of the cDNA is sequenced and aligned with known sequences (see Fig. 2b). This approach will reveal any mutations in exons; however, extensive genomic sequencing is required to identify mutations occurring in enhancer, promoter, or intron regions. Also, to determine whether mutations induced by ENU result in a defect in function requires further functional analysis of the mutated gene.

3.7. Comments on the Use of the Method

One of the major impediments of genetic screening in cultured mammalian cells is diploidy, since only a single copy of each autosomal gene is generally destroyed by mutagenesis. The method described in this chapter allows identifying both haplo-sufficient and -deficient genes involved in anthrax lethal toxin-induced cell death. The gene identified can be essential to the cell death process either in any genetic background (e.g., monogenic phenotype) or in the presence of nonspecific mutations scattered throughout the genome (e.g., polygenic phenotype). Whether the gene identified is responsible for a monogenic or polygenic

phenotype can be determined by the knocking down of the gene using small interference RNA or another method. One of the drawbacks in this method is that multiple rounds of ENU treatment can induce genetic instability, resulting in polyploidy or aneuploidy, and selection of tolerant cells to ENU-induced damages, which can obscure the effects of the subsequent insertional mutagenesis. Therefore, the fact that disruption of a given gene creates a phenotype within a quasi-haploid cell line must be interpreted with some qualification. The plasticity of cultured cells is another problem associated with the use of genetic approaches in cultured cells, and this may be compounded by increased genome instability after repeated ENU treatments. Therefore, it is recommended to use the cells that have as few passages as possible.

4. Notes

1. The Phoenix cells can be obtained by sending email request to: safedep@atcc.org.
2. Retroviral vector pDisrupt8 required for gene trapping is available by request from the authors. Send request to: Sung O. Kim, University of Western Ontario, Department of Microbiology and Immunology, Siebens-Drake Research Institute, Room 119, 1400 Western Road, London, Ontario, N6G 2V4, Canada. In order to produce gene deletion and avoid gene truncation, a self-cleavage ribozyme sequence was incorporated into the retroviral vector pDisrup8 (see Fig. 3a). The sequence of the ribozyme is shown in Fig. 3b. The vector was also designed to have a poly(A) trapping feature because

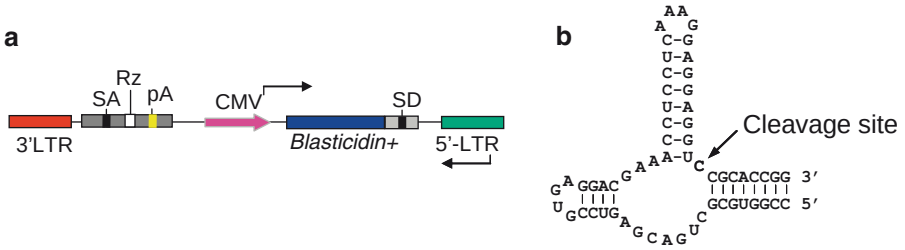


Fig. 3. Representation of pDisrup8 retroviral vector-mediated insertional mutagenesis. (**a**) Diagram of the retroviral vector pDirup 8 used for gene knockout. Rz: ribozyme sequence, SD: splicing donor, SA: splicing acceptor, pA: poly A signal sequence, *CMW* CMV promoter, *blasticidin+* blasticidin resistant gene; (**b**) Sequence of the ribozyme in the vector (Reproduced from ref. 11 © 2007 BioTechniques. Used by permission).

marker gene expression in poly(A) trapping does not depend on the activity of the endogenous promoter, and 3′-RACE, performed after poly(A) trapping to identify the disrupted gene, is much easier to apply than 5′-RACE used after promoter trapping. Figure 4 shows the consequences of all the possible retroviral insertion events. CMV-derived blasticidin resistance gene (*blasticidin+*) is expressed only when retroviral insertion occurred in either the intron or exon of a gene.

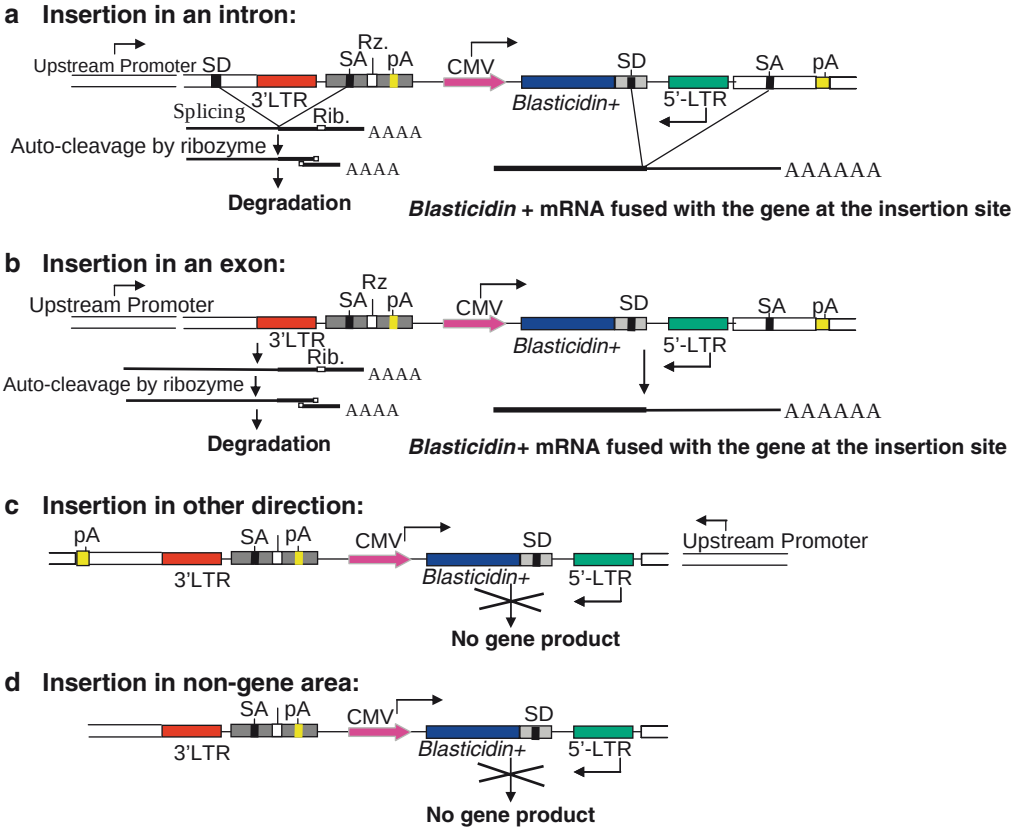


Fig. 4. Representation of possible pDisrup8 retroviral vector insertions. (**a**) Viral insertion in an intron. Open bars flank the sequence of the genome at the insertion site. The upstream SD of the endogenous gene splices with a vector-encoded SA and the transcript ends at the vector pA. The Rz located 3′ to the SA sequence cleaves the transcript and results in no gene product. The SD downstream of the *blasticidin*$^+$ coding sequence splices with endogenous SA to generate fused mRNA capable of expressing full-length *blasticidin*$^+$ gene product. (**b**) Viral insertion in an exon. The transcript initiated by the endogenous gene directly fused with the viral gene at the end of the vector pA. The ribozyme cleaves this transcript, and no gene product will be generated. The *blasticidin*$^+$ transcript will join exons downstream of the insertion site by fusion or splicing events, depending on the presence of a downstream SA. Full-length *blasticidin*$^+$ product is produced. (**c** and **d**) Insertion occurs in an inverted orientation (**c**) or non-gene area (**d**). The CMV promoter driven transcript of *blasticidin*$^+$ gene lacks a pA, which results in degradation of the transcript. No *blasticidin*$^+$ product is made (Reproduced from ref. 11 © 2007 BioTechniques. Used by permission).

3. pCI-VSVG plasmid can be obtained through Addgene Inc. (Cambridge, MA).
4. TRIzol is stable at 4°C for a long time. Other reagents (chloroform, isopropanol, 75% ethanol, and diethylpyrocarbonate treated water) are used to obtain high grade RNA. 100% ethanol is diluted with diethylpyrocarbonate (DEPC)-water for 75% ethanol. Purified total RNAs are used directly for the first strand cDNA synthesis or stored at −80°C for the next steps.
5. This protocol can be applied to various mammalian somatic cell lines. However, one of the inherent limits of this method is that the cell lines should contain normal numbers of chromosomes. Immortalized cell lines often harbor an abnormal number of chromosomes (polyploidy or aneuploidy). The karyotype of a cell can be examined by chromosome-spreading and counting analysis. Another limit of this method is the plasticity of cultured cells. Multiple rounds of chemical mutagen treatments could increase genome instability. At present, the most effective way to minimize plasticity in cultured cells is to use cells having as few passages as possible.
6. Estimating the number of ENU mutant clones required for obtaining a pool of quasi-haploid cell culture is based on defective functional mutation frequency at the *hprt* locus, which is located on the X-chromosome. Therefore, the cells to be examined should originate from a male.
7. Pretreatment of LeTx renders macrophages resistant to subsequent LeTx-cytotoxicity for about 1 week (14). Therefore, it is important to confirm LeTx-susceptibility at least a week after the initial treatment.

References

1. Carlson CM, Largaespada DA (2005) Insertional mutagenesis in mice: new perspectives and tools. Nat Rev Genet 6:568–80
2. Forrai A, Robb L (2005) The gene trap resource: a treasure trove for hemopoiesis research. Exp Hematol 33:845–56
3. Georgel P, Du X, Hoebe K, Beutler B (2008) ENU mutagenesis in mice. Methods Mol Biol 415:1–16
4. Koike H, Horie K, Fukuyama H, Kondoh G, Nagata S, Takeda J (2002) Efficient biallelic mutagenesis with Cre/loxP-mediated inter-chromosomal recombination. EMBO Rep 3:433–7
5. Liu P, Jenkins NA, Copeland NG (2002) Efficient Cre-loxP-induced mitotic recombination in mouse embryonic stem cells. Nat Genet 30:66–72
6. Guo G, Wang W, Bradley A (2004) Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells. Nature 429:891–5
7. Yusa K, Horie K, Kondoh G et al (2004) Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene. Nature 429:896–9
8. Yu Y, Bradley A (2001) Engineering chromosomal rearrangements in mice. Nat Rev Genet 2:780–90
9. Concepcion D, Seburn KL, Wen G, Frankel WN, Hamilton BA (2004) Mutation rate

and predicted phenotypic target sizes in ethylnitrosourea-treated mice. Genetics 168: 953–9

10. Farley AH (2004) ET, Kazazian HH, Jr. More active human L1 retrotransposons produce longer insertions. Nucleic Acids Res 32:502–10
11. Kim SO, Ha SD, Lee S, Stanton S, Beutler B, Han J (2007) Mutagenesis by retroviral insertion in chemical mutagen-generated quasi-haploid mammalian cells. Biotechniques 42:493–501
12. Heilig JS, Elbing KL, Brent R. Large-scale preparation of plasmid DNA. Curr Protoc Mol Biol 2001;Chapter 1:Unit1 7
13. Kingston RE, Chen CA, Okayama H. Calcium phosphate transfection. Curr Protoc Immunol 2001;Chapter 10:Unit 10 3
14. Ha SD, Ng D, Lamothe J, Valvano MA, Han J, Kim SO (2007) Mitochondrial proteins Bnip3 and Bnip3L are involved in anthrax lethal toxin-induced macrophage cell death. J Biol Chem 282:26275–83

Chapter 24

Site-Directed Disulfide Cross-Linking to Probe Conformational Changes of a Transporter During Its Functional Cycle: *Escherichia coli* AcrB Multidrug Exporter as an Example

Yumiko Takatsuka and Hiroshi Nikaido

Abstract

Many proteins, especially transporters, are thought to undergo large conformational alterations during their functional cycle. Since X-ray crystallography usually gives only the most stable conformation, other methods are needed to probe this conformational change. Site-directed disulfide cross-linking is often very useful for this purpose. We illustrate this by using the *Escherichia coli* AcrB, a proton-motive-force-dependent multidrug efflux transporter. Crystallographic studies of the asymmetric trimer of AcrB suggest that each protomer in the trimeric assembly goes through a cycle of conformational changes during drug export (functional rotation hypothesis). Site-directed disulfide cross-linking between those residues that come close to each other in only one stage in the cycle inactivated the transporter, showing that the conformational changes indeed occurred in vivo and that they are required for drug transport. A *dsbA* strain, which has a diminished activity to form disulfide bonds in the periplasm, was used to verify the conclusion by showing a restored transport activity in this strain. Furthermore, we describe "a real-time cross-linking experiment," in which rapidly reacting, sulfhydryl-specific cross-linkers, methanethiosulfonates, inactivate the AcrB double-cysteine mutant expressed in *dsbA* cells instantaneously.

Key words: Site-directed cross-linking, Conformational change, Disulfide bond, Sulfhydryl-specific cross-linker, Methanethiosulfonate, dsbA, Multidrug efflux transporter, AcrB, *Escherichia coli*

1. Introduction

Disulfide cross-linking is an attractive method for probing conformational changes in proteins, as was shown more than 20 years ago (1). When this method is applied to intact cells of *Escherichia coli*, however, it is important to realize that the formation of disulfide bonds is facilitated by an elaborate system in the periplasm (2),

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_24,

whereas such a system does not exist within the membrane bilayer or in the cytosol. Thus, the addition of oxidizing agents (1) or cross-linkers (3) is often necessary to cover the cysteine residues in the latter locations. It may also be possible to use genetically altered strains that are able to form disulfide bonds in the cytosol (4, 5). In the example to be described, cross-linking occurs in the periplasm, and we also describe the manipulation of the disulfide-forming machinery in this compartment.

Multidrug efflux transporters of the *r*esistance-*n*odulation-*d*ivision (RND) superfamily, such as AcrB of *E. coli*, play an important role in both intrinsic and elevated multidrug resistance of Gram-negative bacteria (6). These transporters associate with two other classes of proteins, the outer membrane channel such as TolC and the periplasmic membrane fusion protein such as AcrA (6). This tripartite complex spans the entire thickness of the cell envelope, containing the inner and outer membranes as well as the periplasm between them (7) and expels, driven by a proton-motive force, the drugs directly into the external medium. This process must obviously involve large conformational alterations of AcrB.

The AcrB exists as a homotrimer in which each subunit provides 12 transmembrane helices and a large periplasmic headpiece (8). The newly elucidated asymmetric crystal structure of AcrB trimer (9–11) shows that each protomer has a conformation significantly different from that of its neighbors, presumably representing each of the steps (open access, drug binding, and drug extrusion) in the conformational change cycle for drug export and suggests a functionally rotating mechanism.

We provided biochemical evidence supporting this hypothesis by site-directed disulfide cross-linking experiments (12), taking advantage of the observation that the cyclic conformational change of AcrB involves the opening and closing of the large external cleft in the periplasmic domain. The AcrB proteins with double-Cys mutations on both sides of the cleft, where a disulfide bond would be produced only in the extrusion protomer (Fig. 1), are inactive. The fact that the inactivation is due to the formation of disulfide bonds has been supported by the restoration of transport activity in a *dsbA* strain, which has a defective DsbA, a main enzyme that catalyzes disulfide bond formation in the periplasm (2). However, *dsbA* did not restore the activity of some mutant pairs, and formation of large aggregates of AcrB suggested that cross-linking may have occurred during the assembly of the trimeric complex. In view of these potential problems relying on the activity of naturally cross-linked proteins, it is important that the cross-linking and inactivation of AcrB be observed in "real rime" by using methanethiosulfonate (MTS) cross-linkers, which are known to act much more rapidly than the usual cross-linkers (13, 14). The addition of such cross-linkers to *dsbA* cells expressing double-Cys mutant AcrB proteins that are pumping out a fluorescent dye,

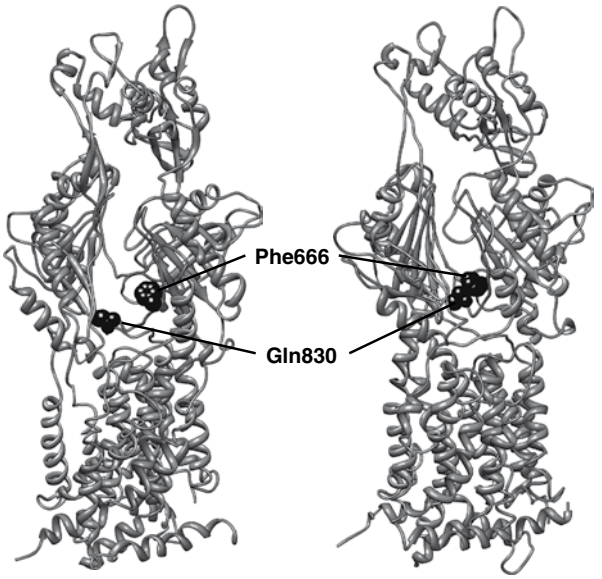


Fig. 1. Predicted conformational change of AcrB protomer during drug transport and the choice of the residues to introduce Cys mutations. The figure shows two of the protomers (the access protomer on the *left* and the extrusion protomer on the *right*) of the asymmetric AcrB structure (9), viewed from outside of the trimer. The large cleft in the periplasmic domain, wide open in the access protomer (*left*), is nearly completely closed in the extrusion protomer (*right*). The distance between the two residues, Phe666 and Gln830 (shown as dark space-filling models), on the opposing walls of the cleft becomes close enough to produce a disulfide bond only in the extrusion protomer (*right*, 4.4Å), as opposed to the access (*left*) and binding (not shown) protomers (>9Å). These two residues were therefore chosen as a pair of the positions to introduce the Cys mutations in order to test the conformational change of AcrB protomer during transport function. The figure was drawn by using the UCSF Chimera package (18), and PDB file 2DRD.

ethidium, stops the function of the pump instantaneously. (An alternative approach is the reduction of disulfide bonds with externally added dithiothreitol, which resulted in the recovery of the transport function (15)). All these results show that when the external cleft in the periplasmic domain of AcrB becomes closed to allow the formation of a disulfide bridge or a cross-linked structure, then the resultant loss of flexibility in the conformation inactivates the efflux pump.

2. Materials

2.1. Site-Directed Mutagenesis and Mutant Plasmids Preparation

1. For manipulation of recombinant DNA, we follow the standard methods (16).
2. Template plasmid (see Note 1): pSCLBH (12). This pSPORT1 (medium copy-number vector with an *amp* marker; Gibco BRL, Invitrogen Corp) based plasmid expresses the cysteineless

and hexahistidine-tagged AcrB protein, CL-AcrB^{His} , under the control of the *lac* promoter. The codons for two intrinsic cysteines of the *acrB* gene have been converted to serine codons. The CL-AcrB^{His} is fully functional.

3. *PfuUltra* high-fidelity (HF) DNA polymerase (Stratagene).
4. 10× *PfuUltra* HF reaction buffer (Stratagene).
5. dNTP mix: 2.5 mM each dNTP, store at −20°C.
6. Oligonucleotide primers to introduce Cys mutations (see Note 2).
7. DpnI restriction enzyme (New England Biolabs).
8. Competent cells of *E. coli* DH5α (16). Store at −80°C.
9. LB (Tryptone 1%; Yeast Extract 0.5%; NaCl 0.5%) agar plates, supplemented with ampicillin (100 µg/mL).
10. LB liquid medium, supplemented with ampicillin (100 µg/mL).
11. Plasmid miniprep kit (Fermentas Inc or Qiagen Inc).
12. Tris–EDTA (TE) buffer: 10 mM Tris–HCl, 1 mM EDTA, pH 8.0. Autoclaved and stored at room temperature.
13. Autoclaved distilled water. Store at room temperature.

2.2. Functional Characterization of Single- and Double-Cys Mutants and Effect of dsbA Mutation on Transport Activities

1. *E. coli* AG100YB (12): *acrB* deletion mutant (Δ*acrB*::Spc^r), generated from a K-12 strain AG100. In this strain, 90% of the *acrB* gene is deleted and replaced with the spectinomycin resistance (Spc^r) gene, *aadA*. Prepare as competent cells, and store at −80°C. (Bacterial strains and plasmids mentioned in this chapter are available from the authors.)
2. *E. coli* AG100YBD (12): Δ*acrB dsbA* mutant (*dsbA1*::*kan*), derived from AG100YB. Prepare as competent cells, and store at −80°C.
3. LB agar plates, supplemented with ampicillin (100 µg/mL).
4. LB liquid medium supplemented with ampicillin (100 µg/mL).
5. LB agar, autoclaved and kept at 50°C.
6. Cholic acid sodium salt: hydrate, from ox or sheep bile, ≥99% (Sigma)
7. Square dishes with grid, 100 × 100 × 15 mm (Fisher Scientific)
8. Disposable loops, 1 µL (Fisher Scientific).

2.3. Real-Time Inactivation of AcrB Double-Cys Mutant by the Use of Fast-Acting MTS Cross-Linker

1. AG100YBD (Δ*acrB dsbA*) strains transformed by the plasmids expressing wild-type cysteineless AcrB, or single- or double-Cys mutant.
2. LB liquid medium with ampicillin (100 µg/mL).
3. Assay buffer: 50 mM sodium phosphate, 0.1 M NaCl, 0.1% (vol/vol) glycerol, pH 7.0. Autoclaved and stored at room temperature.

4. Shimadzu RF-5301PC spectrofluorometer (Shimadzu Scientific Instruments, Inc.).
5. Quartz cuvette for fluorometry, 10 mm path length (Starna Cells, Inc.)
6. Micro stirring bar, 7 mm length × 2 mm diameter (Fisher Scientific).
7. Ethidium bromide, 2 mM stock solution in distilled water. Store at –20°C.
8. MTS reagents: 5 mM 1,2-ethanediyl bismethanethiosulfonate (MTS-2-MTS, approximately 5.2Å spacer), and 10 mM pentyl MTS (5-MTS) (Toronto Research Chemicals) (see Note 3). Freshly dissolved in dimethyl sulfoxide (DMSO)-ethyl acetate (3:1, vol/vol) (see Note 4), and kept on ice until use (see Note 5).
9. Carbonyl cyanide *m*-chlorophenylhydrazone (CCCP), as a proton conductor: 10 mM stock solution in ethanol. Store at –20°C. The solution is stable at least for a month.

3. Methods

3.1. Site-Directed Mutagenesis and Preparation of Mutant Plasmids

The method for introduction of Cys mutations follows the QuickChange site-directed mutagenesis protocol from Stratagene. A detailed instruction manual is available at their website.

1. Combine 5 µL 10× *PfuUltra* HF reaction buffer, 5–20 ng template plasmid, 4 µL dNTP mix, 2 µL each of two complementary oligonucleotide primers (10 pmol/µL in TE buffer). Adjust the volume to 49.2 µL with distilled water, then add 0.8 µL *PfuUltra* HF DNA polymerase (2.5 U/µL) (see Note 6), and mix gently.
2. Overlay the reaction mixture with approximately 30 µL of mineral oil if necessary.
3. Cycle the reaction using the PCR program of 3 min at 95°C followed by 16 cycles of amplification, and an appropriate time for the final extension at 68°C. An amplification cycle consists of: 45 s at 95°C, 45 s at an appropriate annealing temperature, and 1 min/kb of plasmid length at 68°C.
4. Remove the mineral oil, and treat the reaction mixture with 10 U of DpnI at 37°C for 1 h to digest the template DNA.
5. Transform *E. coli* DH5α with the DpnI-treated DNA. Up to 8 µL of the reaction can be used to 100 µL of the competent cells prepared with the $CaCl_2$ procedure. Incubate the transformation plates overnight at 37°C.

6. Prepare the plasmids from the culture of colonies using a plasmid miniprep kit, and sequence the *acrB* gene to ensure the presence of the desired mutation and also to confirm that there are no unintended nucleotide sequence alterations.

3.2. Functional Characterization of Single- and Double-Cys Mutants and Effect of dsbA Mutation on Transport Activities

Activity of each mutated AcrB protein can be evaluated by the drug susceptibility of the Δ*acrB* strain expressing each mutant protein. We use the gradient plate technique (17) (Fig. 2a, b), with a plate containing two agar layers that allow a gradual increase of the drug concentration along one horizontal axis.

This method, which relies on the formation of disulfide bonds between nearby Cys residues in the periplasm, works for many pairs of Cys residues. However, there is a possibility that disulfide bonds are formed during the formation of the multiprotein assembly, that is, in a way that has nothing to do with the conformational alterations during the functional cycle of the transporter (12).

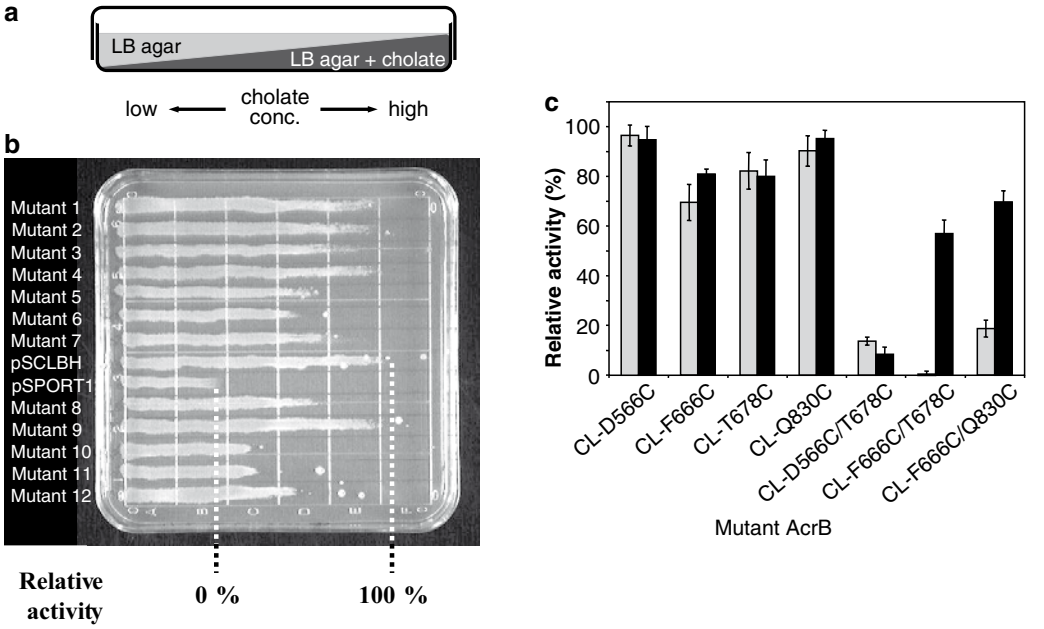


Fig. 2. Effect of *dsbA* mutation on the activities of AcrB Cys mutants evaluated by drug susceptibility of the host strains. (**a**) Side view of a gradient plate containing cholate in the lower layer. (**b**) An example of measurement of bacterial drug susceptibility using the gradient plate. Fully functional control, the Δ*acrB* strain harboring wild-type AcrB-expressing-plasmid, pSCLBH, grew well up to the region containing cholate at high concentrations, while the nonfunctional control, the host with vector only (pSPORT1), grew only for a short length. The relative activity of each mutant was calculated from the length of its growth compared to the two controls, pSCLBH (100%) and pSPORT1 (0%). (**c**) Activities of AcrB single- and double-Cys mutants and the effect of *dsbA* mutation. AcrB proteins containing double-Cys mutations on both sides of the cleft are strongly compromised in function in *dsbA*$^+$ host strain AG100YB (*gray bars*, three columns on the right). In contrast, in the *dsbA* host AG100YBD (*black bars*), the CL-F666C/T678C and CL-F666C/Q830C double mutants largely retain their transport activities, indicating that the inactivation of these two double-Cys mutants in the *dsbA*$^+$ host is due to the formation of disulfide bonds. This is one of the biochemical evidences that the conformational changes, including the closure of the cleft suggested from the asymmetrical trimer structure, indeed occur in vivo and that they are required for AcrB function (Reproduced from ref. (12)).

In order to exclude such a possibility, the real-time cross-linking of already assembled AcrB, described below, is very useful.

1. Transform competent cells of a pair of isogenic strains, AG100YB (Δ*acrB*), and AG100YBD (Δ*acrB dsbA*), with plasmids containing mutated *acrB*. As fully functional and non-functional AcrB controls, cells are also transformed with plasmids containing the wild-type cycteineless *acrB* and vector only. Incubate the transformation plates overnight at 37°C.
2. Pick a single colony from the fresh transformation plates (see Note 7) and grow in LB liquid medium with ampicillin until an optical density at 660 nm (OD_{660}) reaches approximately 0.6 (midexponential-phase) (see Note 8).
3. To prepare a linear concentration gradient of cholic acid in LB agar plates, add cholic acid to LB agar kept at 50°C, dissolve and mix well, and pour 20 mL into the square dishes. Slant the plates sufficiently so that the entire bottom is just covered. Wait until the agar hardens. It is necessary to adjust the concentration of cholic acid in the bottom layer depending on the susceptibility levels of the strains used (see Note 9).
4. Place the dishes in the normal horizontal position, and overlay ~28 mL LB agar without drugs to cover the bottom layer (adjust the volume so that the bottom layer is covered sufficiently as shown in Fig. 2a). Wait until the agar hardens, and the surface becomes dry. Use the plates in the same day.
5. Dilute midexponential-phase cultures to an OD_{660} of 0.1 with LB broth with ampicillin.
6. Streak the diluted cultures using Fisherbrand disposable 1 μL loops as a line across the plate, in parallel with the drug gradient (Fig. 2b, see Note 10). Each plate should also be inoculated with control strains harboring wild-type pSCLBH and the vector pSPORT 1 (see Note 11). Incubate the plates at 37°C for 24 h.
7. Measure bacterial growth across the plates, from low to high drug concentrations.
8. Calculate the relative activity of each mutated AcrB protein as follows. Divide the length of growth of each mutant strain minus that of the strain carrying the vector alone, by the length of growth of the strain containing pSCLBH minus that of the vector-containing strain, and multiply the result by 100. Thus, the full efflux activity and no activity should produce values of 100 and 0%, respectively (Fig. 2b). An example of the results obtained is shown in Fig. 2c.

3.3. Real-Time Inactivation of AcrB Double-Cys Mutant by the Use of Fast Acting MTS Cross-Linker

1. Pick fresh single colonies of AG100YBD transformants containing pSCLBH, single- and double-Cys mutant plasmids, and grow them in 2 mL LB with ampicillin overnight without shaking at room temperature (see Note 12).
2. Dilute the cultures into 10 mL of LB medium with ampicillin to an OD_{660} ~0.08 and grow with shaking at 37°C to an OD_{660} of 0.7–0.9 without isopropyl-β-D-thiogalactopyranoside (IPTG) induction.
3. Harvest the cells at room temperature, wash once with 10 mL of assay buffer, and resuspend in 10 mL of assay buffer. Measure the OD_{660}.
4. Dilute the cells with assay buffer to an OD_{660} of 0.2 (see Note 13), and place 2 mL of cells into the quartz 1 × 1 cm cuvette with micro stirring bar. Set the cuvette in the spectrofluorometer and stir gently.
5. Add 5 µL of 2 mM ethidium bromide (final concentration, 5 µM, see Note 13) into the cell suspension, and monitor the ethidium influx rate into the cells at room temperature with fluorescence increase at excitation and emission wavelengths of 520 nm and 590 nm, respectively.
6. After a few minutes of preincubation with ethidium bromide, add to the cells one of the following: 8 µL of solvent alone (DMSO-ethyl acetate [3:1, vol/vol]), 8 µL of 5 mM MTS-2-MTS (final concentration, 20 µM), or 8 µL of 10 mM 5-MTS (final concentration, 40 µM) (see Note 14). Follow the accumulation of ethidium for several minutes more.
7. As a positive control of AcrB inactivation, add 8 µL of 10 mM CCCP (final concentration, 40 µM) to the cells instead of MTS reagents.
8. An example of the data is shown in Fig. 3. It is seen that only in the mutant protein containing the double Cys residues, an instantaneous inactivation of the AcrB pump occurred, causing the increased influx of ethidium (the bottom right panel).

4. Notes

1. The template plasmid should be isolated from a *dam*$^+$ *E. coli* strain for the following removal step by DpnI digestion. The majority of the commonly used *E. coli* strains are *dam*$^+$. We use *E. coli* DH5α or DH10B (Invitrogen) for plasmid preparation.

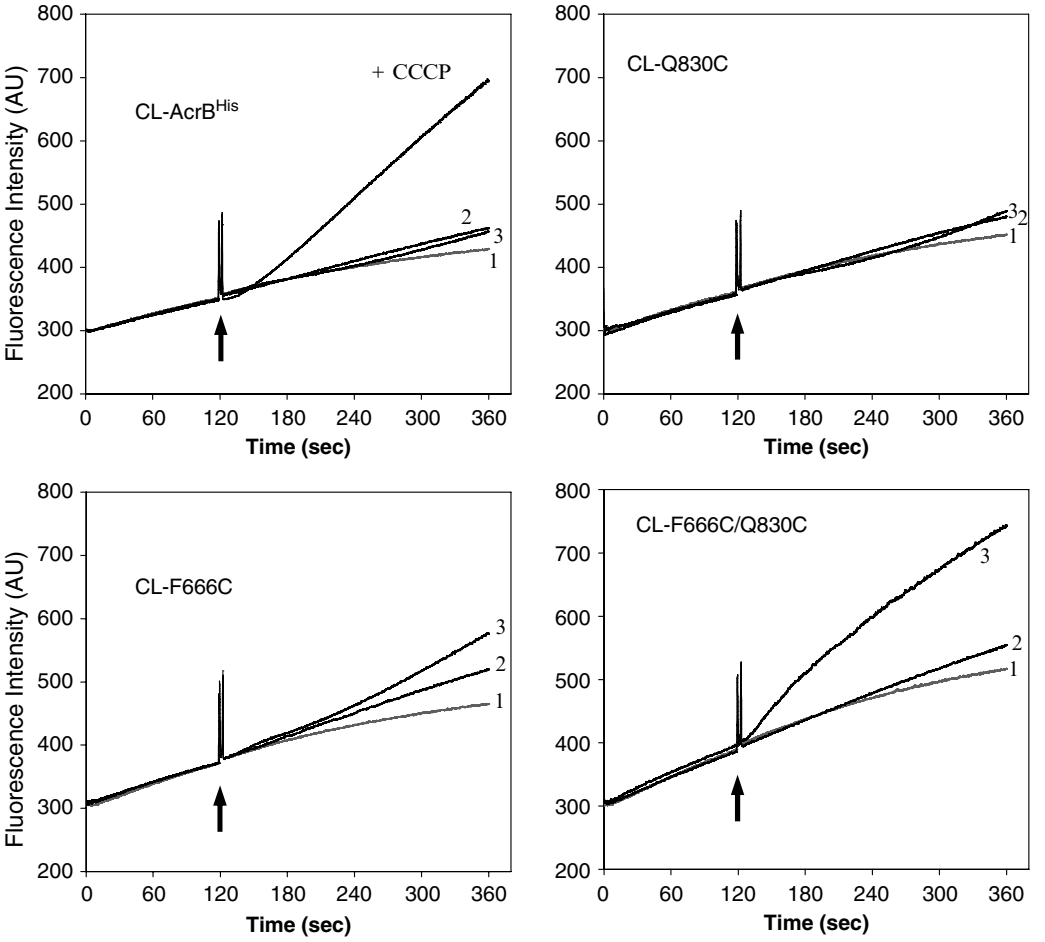


Fig. 3. Real-time inactivation of AcrB double-Cys mutant by cross-linking with MTS reagent. Cellular accumulation of ethidium by the Δ*acrB dsbA* host AG100YBD cells expressing each AcrB protein was monitored continuously by measuring the fluorescence of the ethidium–nucleic acid complex. After 2 min of incubation with ethidium bromide (*arrow*), solvent alone (curve 1) or MTS reagents (curve 2, 5-MTS; curve 3, MTS-2-MTS) were added to cell suspension. The AcrB with double-Cys mutation in the cleft, CL-F666C/Q830C, was active in the *dsbA* host in the efflux of ethidium, so that only very slow entry of ethidium was seen initially (*bottom right panel*). However, upon addition of MTS-2-MTS, cross-linking apparently occurred, so that AcrB became inactivated, inducing a rapid ethidium influx and increased fluorescence (curve 3, *bottom right panel*). In contrast, MTS-2-MTS had little effect on the ethidium entry rate in cells expressing no-Cys (CL-AcrB^{His}) or a single-Cys (CL-F666C and CL-Q830C) AcrB. A control reagent, 5-MTS (a noncross-linker), produced no AcrB inactivation in any of the mutants (curves 2). As a positive control, a proton conductor (CCCP) was added to the cells expressing CL-AcrB^{His}, resulting in a rapid influx of ethidium because of inactivation of AcrB due to the loss of the proton-motive force (*top left panel*) (Reproduced from ref. (12)).

2. We designed most of the primers in the length of 20–30 bases, with the calculated melting temperature of about 60°C or higher, and checked that there is no significant matching sequences in other parts of *acrB*.
3. We chose the MTS reagents that are likely to cross the outer membrane. MTS cross-linkers with different size spacers are

available from approximately 3.9Å (MTS-1-MTS) to approximately 10.4Å (MTS-6-MTS). We also examined MTS-5-MTS (approximately 9.1Å spacer) and obtained similar results as MTS-2-MTS. As a noncross-linker control reagent, we used 5-MTS whose length is similar to that of MTS-2-MTS.

4. MTS-2-MTS is soluble in DMSO, but 5-MTS is not. We found that ethyl acetate (good solvent for 5-MTS) alone affected significantly the ethidium influx rate into the cells. To minimize this effect and to obtain a reasonable solubility for both MTS-2-MTS and 5-MTS, DMSO-ethyl acetate (3:1, vol/vol) was the best solvent we found so far.
5. MTS reagents are sensitive to moisture and are hydrolyzed in water over a period of time. The reagents should be stored in a desiccator at −20°C and warmed up to room temperature before opening of the vial. We made up the solutions in organic solvent immediately prior to use and the solutions, kept on ice, were used within ~1.5 h.
6. The protocol from Stratagene recommends the use of 1 μL (2.5 U) of *PfuUltra* HF DNA polymerase. We found that 0.5–0.8 μL enzyme also provided enough products for obtaining plasmids with point mutation.
7. We observed that the levels of resistance of the Δ*acrB* cells expressing AcrB mutants from pSPORT1 derivatives often varied after storage of the transformed strains for several days at 4°C. Therefore, for these experiments, we only used freshly transformed cells.
8. To avoid the nonreproducible drug susceptibility patterns caused by strong overexpression of AcrB alone, *acrB* is expressed without isopropyl-β-D-thiogalactopyranoside (IPTG) induction.
9. To calculate the relative activity of each mutated AcrB, both controls (fully functional pSCLBH and nonfunctional vector only) should grow and show large differences on the lengths of growth zone on the same plate (see Fig. 2b). We tested several AcrB substrates with gradient plates and found that cholic acid satisfied this requirement. With some compounds such as novobiocin and rhodamine 6G, it was difficult to find the appropriate concentration in the bottom layer, because the difference in susceptibility to these compounds was too large between the two controls. Thus, concentrations at which the fully active control grew well, no growth was seen for the nonfunctional control, whereas at other concentrations that gave a short growth zone for the vector only control, cells expressing the fully active transporter gave growth covering the entire length of the plate. We also adjusted the

concentration of cholic acid in the lower layer to 10,000 μg/mL for AG100YB and to 16,000 or 18,000 μg/mL for AG100YBD, since the resistance levels of these strains to cholate were different.

10. We streak the cells twice on the same line without refilling the loop, from high to low drug concentrations using one side of the loop, and then low to high concentrations using the other side of the loop. Dry the surface of the plate well to avoid flooding of the plate.
11. We assay all mutants at least four times with controls and change the position of strains on each plate.
12. Overgrowth of the preculture sometimes caused nonreproducible results in ethidium influx assay. To avoid this, the preculture was prepared from a fresh transformant colony without shaking, either at room temperature or at 30°C. We also tested the frozen glycerol stock, which was made with a culture in midexponential phase (OD_{660} ~0.5) from a fresh colony and stored in aliquots at −80°C, as the inoculum for the preculture. This method produced similar results until up to 2 weeks of storage of the stock.
13. The concentrations of cell suspension and ethidium bromide can be modified depending on the conditions of experiments. Generally, higher concentrations of the cells and ethidium bromide show faster increase of the fluorescence. For our conditions (host strain and the expression levels of AcrB), the cell suspension of OD_{660} of 0.2 and 5 μM ethidium bromide produced good results.
14. MTS reagents at these final concentrations produced the best results for these strains. MTS reagents at higher concentrations caused increased influx of ethidium even with the cells expressing the wild-type cysteineless AcrB. On the other hand, MTS-2-MTS at lower concentrations sometimes failed to cause a detectable inactivation of AcrB double-Cys mutants. We also note that a very low concentration of MTS reagents (4 μM) gave the best results for another host strain derived from BL21.

Acknowledgments

This work was supported by research grant AI-09644 from U.S. Public Health Service.

References

1. Falke JJ, Koshland DE Jr (1987) Global flexibility in a sensory receptor: a site-directed cross-linking approach. Science 237: 1596–1600
2. Kadokura H, Katzen F, Beckwith J (2003) Protein disulfide bond formation in prokaryotes. Annu Rev Biochem 72:111–135
3. Moore KJ, Fillingame RH (2008) Structural interactions between transmembrane helices 4 and 5 of subunit *a* and the subunit *c* ring of *Escherichia coli* ATP synthase. J Biol Chem 283:31726–31735
4. Stewart EJ, Aslund F, Beckwith J (1998) Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. EMBO J 17:5543–5550
5. Masip L, Pan JL, Haldar S et al (2004) An engineered pathway for the formation of protein disulfide bonds. Science 303:1185–1189
6. Nikaido H (1996) Multidrug efflux pumps of gram-negative bacteria. J Bacteriol 178: 5853–5859
7. Eswaran J, Koronakis E, Higgins MK, Hughes C, Koronakis V (2004) Three's company: component structures bring a closer view of tripartite drug efflux pumps. Curr Opin Struct Biol 14:741–747
8. Murakami S, Nakashima R, Yamashita E, Yamaguchi A (2002) Crystal structure of bacterial multidrug efflux transporter AcrB. Nature 419:587–593
9. Murakami S, Nakashima R, Yamashita E, Matsumoto T, Yamaguchi A (2006) Crystal structure of a multidrug transporter reveals a functionally rotating mechanism. Nature 443:173–179
10. Seeger MA, Schiefner A, Eicher T, Verrey F, Diederichs K, Pos KM (2006) Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. Science 313: 1295–1298
11. Sennhauser G, Amstutz P, Briand C, Storchenegger O, Grütter MG (2007) Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors. PLoS Biol 5:e7
12. Takatsuka Y, Nikaido H (2007) Site-directed disulfide cross-linking shows that cleft flexibility in the periplasmic domain is needed for the multidrug efflux pump AcrB of *Escherichia coli*. J Bacteriol 189:8677–8684
13. Akabas MH, Stauffer DA, Xu M, Karlin A (1992) Acetylcholine receptor channel structure probed in cysteine-substitution mutants. Science 258:307–310
14. Kenyon GL, Bruice TW (1977) Novel sulfhydryl reagents. Methods Enzymol 47:407–430
15. Seeger MA, von Ballmoos C, Eicher T et al (2008) Engineered disulfide bonds support the functional rotation mechanism of multidrug efflux pump AcrB. Nat Struct Mol Biol 15:199–205
16. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
17. Bryson V, Szybalzski W (1952) Microbial selection. Science 116:45–51
18. Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF Chimera – a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612

Chapter 25

Site-Specific Incorporation of Extra Components into RNA by Transcription Using Unnatural Base Pair Systems

Michiko Kimoto and Ichiro Hirao

Abstract

The expansion of the genetic alphabet, by an unnatural base pair system, provides a powerful tool for the site-specific incorporation of extra, functional components into nucleic acids by replication and transcription. We developed several unnatural base pairs that function in PCR and in vitro transcription. Among them, a hydrophobic, unnatural base pair between 7-(2-thienyl)-imidazo[4,5-b]pyridine (denoted by **Ds**) and pyrrole-2-carbaldehyde (denoted by **Pa**) exhibits high fidelity in PCR and T7 transcription. Modified **Pa** bases linked with functional groups of interest can also be site-specifically incorporated into RNA opposite **Ds** in DNA templates, by T7 RNA polymerase. Here, we describe the methods for the site-specific biotinylation of RNA molecules by transcription using the **Ds**–**Pa** pair with biotinylated **Pa**TP (Biotin-**Pa**TP).

Key words: Unnatural base pair, PCR, Transcription, Biotinylation

1. Introduction

An artificially created third base pair (unnatural base pair), compatible with the natural A-T(U) and G-C base pairs in replication, transcription, and translation, and having sites where functional groups of interest can be attached, could enable site-specific incorporation of additional components into nucleic acids and proteins (1–10). Many unnatural base pairs have been developed and tested as candidates for the third base pair (2–14). Here, we describe site-specific biotinylation at a desired position, using a hydrophobic, unnatural base pair between 7-(2-thienyl)-imidazo[4,5-b]pyridine (denoted by **Ds**) and pyrrole-2-carbaldehyde (denoted by **Pa**), by in vitro transcription (10). Although some methods for biotinylation of RNA have been available so far, such as chemical synthesis of biotinylated RNA, posttranscriptional

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_25,

modification, and nonsite-specific enzymatic incorporation by transcription, these methods are tedious and time consuming. Chemical synthesis is difficult to prepare RNA fragments longer than 100-mer. Posttranscriptional biotinylation using biotin-conjugated hydrazide requires multiple steps. Enzymatic incorporation by transcription using biotin-linked natural base substrates results in its random incorporation, decreasing the original RNA functions. On the other hand, the **Ds**–**Pa** pair system provides the one-step site-specific biotinylation of any RNA molecules by direct transcription.

The **Ds**–**Pa** pair functions in in vitro replication and transcription (Fig. 1). DNA fragments containing the **Ds**–**Pa** pair can be amplified by PCR with triphosphate substrates G, C, T, and **Pa**, and modified γ-amidotriphosphate substrates of A and **Ds**. The γ-amidotriphosphates are used for increasing the base pair selectivity by preventing the undesired pairings of A-**Pa** and **Ds**–**Ds**. In transcription, each unnatural base substrate can be incorporated into RNA, opposite the complementary unnatural base in DNA templates, by T7 RNA polymerase without γ-amidotriphosphates. Importantly, a variety of modified **Pa** substrates can be synthesized chemically, by attaching a functional group through an amino propyne linker. These modified **Pa** substrates are also site-specifically incorporated into RNA, opposite **Ds** in DNA templates, by T7 transcription (10). The **Ds** amidite reagent for chemical synthesis of DNA templates and the biotin-linked **Pa** substrate (Biotin-**Pa**TP) for transcription are commercially available for site-specific biotinylation of RNA molecules.

Biotin-labeled RNAs are routinely used for immobilization and detection with avidin or streptavidin. The transcription system using the **Ds**–**Pa** pair enables site-specific biotinylation without decreasing intrinsic RNA functional activity. For this system, the DNA templates containing **Ds** are synthesized by two methods: one is enzymatic ligation of a set of short DNA fragments (Fig. 1b), and the other is PCR amplification using 3′-primers containing **Ds** (Fig. 1c). The former is useful for introduction of **Ds** into any position of a DNA template, and the latter is convenient for biotinylation at a specific position in the 3′-region of an RNA molecule. The **Ds**-containing DNA fragments are chemically synthesized by an automated DNA synthesizer using a d**Ds**-amidite. The transcription is performed using Biotin-**Pa**TP with commercially available T7 kits. The incorporation selectivity of Biotin-**Pa** into the desired position of transcripts is 90–96%, depending on the ratio of Biotin-**Pa**TP to the natural base NTPs in transcription (10). We demonstrated the biotinylation (Fig. 2) and streptavidin binding (Fig. 3) of two RNA fragments: one is an example of biotinylation in the middle of an RNA 152-mer, and the other is the biotinylation in the 3′-region of an RNA aptamer (aptamer 9A; (15)) that binds to human Raf-1 protein.

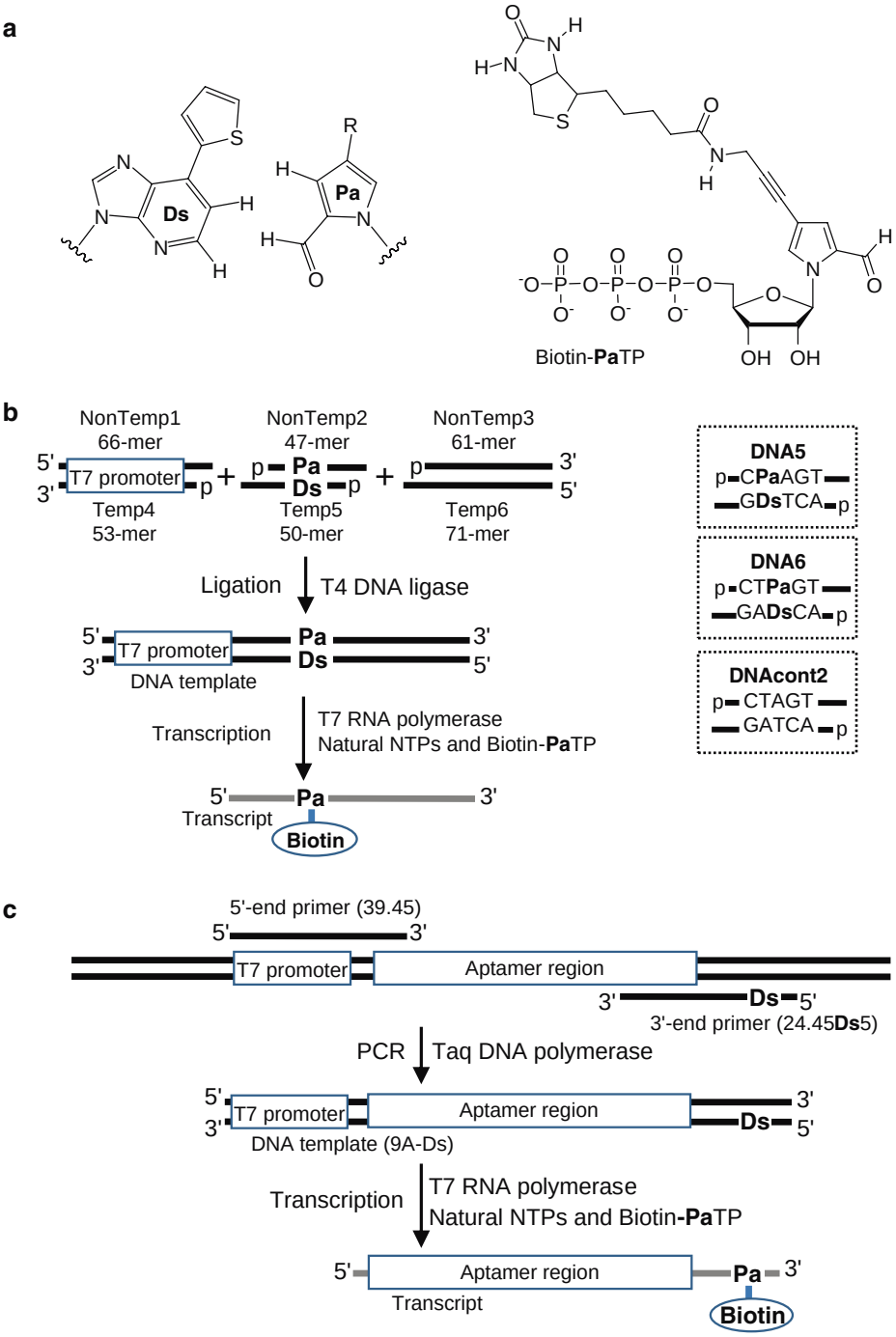


Fig. 1. Specific transcription system using the **Ds**–**Pa** pair. (**a**) Structures of the **Ds**–**Pa** pair and Biotin-**Pa**TP. Preparation of **Ds**-containing templates for T7 transcription by enzymatic ligation (**b**) and by PCR amplification (**c**). For the present demonstration, the sequences of the 174-bp DNA fragments were reported elsewhere (10). The T7 promoter is encoded in NonTemp1, Temp4, and 39.45. The aptamer region is for aptamer 9A that binds to human Raf-1 protein.

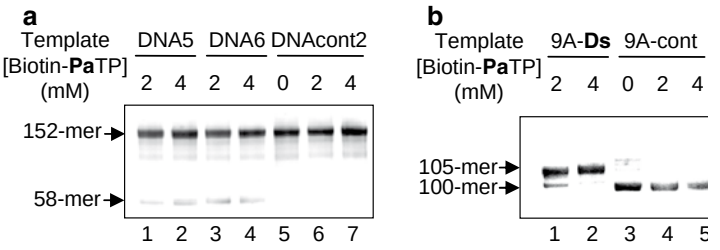


Fig. 2. Gel electrophoresis of the transcripts using the indicated templates, after an incubation at 37°C for 6 h with the natural NTPs (2 mM) in the presence (2 or 4 mM) or absence of Biotin-**Pa**TP. (**a**) The templates were prepared by enzymatic ligation. DNA5 and DNA6 contain **Ds** at positions 58 and 59 in the template strands, respectively. DNAcont2 has no **Ds** in the sequence. (**b**) The templates were prepared by PCR amplification with the T7 promoter-primer (5′-primer) and a 3′-primer containing **Ds** (9A-**Ds** for the 105-mer transcript) or without **Ds** (9A-cont for the 100-mer transcript). Template 9A-**Ds** contains **Ds** at position 103 in the template strand. The gels were stained with SYBR Green II, and the images were obtained with a FujiFilm LAS-4000 bio-imager.

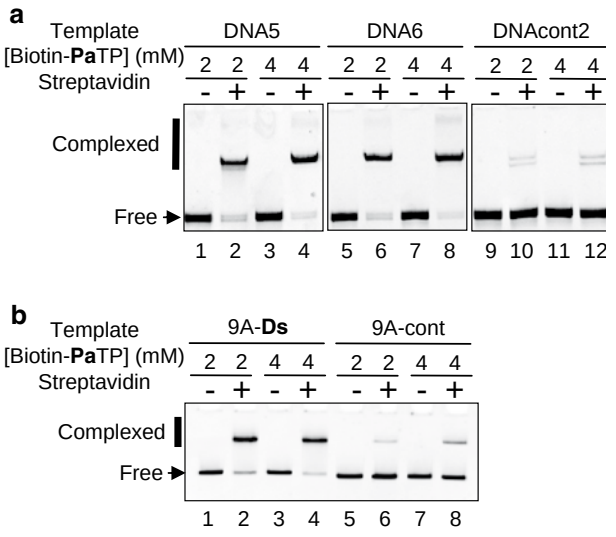


Fig. 3. Gel-mobility shift assays for the binding of biotinylated transcripts with streptavidin. The 152-mer RNAs (**a**) and 105-mer RNAs (**b**), transcribed with Biotin-**Pa**TP, from the templates, were mixed with streptavidin. Biotinylated RNA-streptavidin complexes were separated from free RNAs on a 10% polyacrylamide – 7 M urea gel. The gels were stained with SYBR Green II, and the images were obtained with a FujiFilm LAS-4000 bio-imager. The transcripts (lanes 6 and 7 in Fig. 2a and lanes 4 and 5 in Fig. 2b) obtained from the control DNA templates, DNAcont2 and 9A-cont, in the presence of Biotin-**Pa**TP, only slightly bind to streptavidin (lanes 10 and 12 in Fig. 3a and lanes 6 and 8 in Fig. 3b), indicating that Biotin-**Pa**TP is rarely incorporated into RNA, opposite the natural bases in templates, by T7 transcription.

2. Materials

2.1. Common Materials

It is important to take care to avoid nuclease contamination from conceivable sources, as in the usual handing of RNA molecules. Use gloves during experiments, and ElectroZap or RNaseZap (Ambion, Austin, TX) is effective to remove RNases from laboratory instruments.

1. Apparatus for gel electrophoresis (see Note 1).
2. Nuclease-free deionized, sterile water.
3. Forty percent acrylamide/bis solution (19:1 with 5%C) (see Note 2). Store at 4°C with protection from light.
4. Nuclease-free 10× TBE (see Note 1).
5. Urea (ultra pure grade).
6. Acrylamide/bis solution with 7 M urea in 1× TBE (see Notes 3 and 4).
7. APS: 10% Ammonium persulfate (w/v). This can be stored at 4°C for 1–2 months (see Note 4).
8. TEMED: *N,N,N′,N′*-tetramethyl-1,2,-diaminoethane. Store at 4°C (see Note 4).
9. Polyacrylamide gel containing 7 M urea (see Note 4).
10. TLC plate containing a UV fluorescent indicator (Merck 60F254, Germany) (see Note 5).
11. Hand-held UV lamp (254 nm) (see Note 5).
12. SYBR Green II (see Note 6).
13. Filter Units (Millipore, Billerica, MA): Ultrafree-MC and Steriflip (pore size 0.22 μm) (see Note 7).
14. 3 M sodium acetate, pH 5.2 (see Note 8).
15. Ethanol (see Note 8).
16. 1 μg/μL glycogen (see Note 8).
17. 2× denaturing sample solution: 10 M urea solution in 1× TBE. Store in aliquots at −10 to −30°C (see Note 9).
18. 10× annealing solution: 100 mM Tris–HCl (pH 7.6) and 100 mM NaCl.

2.2. Synthesis and Purification of Ds-Containing DNA Fragments

1. Chemically synthesized **Ds**-containing DNA fragments (0.2 μmol scale) using d**Ds**-amidite (dDs-CE Phosphoramidite, Catalog No: 10-1521-90 for 100 μmol, and 10-1521-02 for 0.25 g, Glen Research, Sterling, VA) (see Note 10).
2. Polyacrylamide gel with 7 M urea (20 cm×40 cm, 2 mm thickness).

2.3. Preparation of Ds-Containing DNA Templates (174 bp) for the 152-mer RNA by Ligation (Fig. 1b)

1. Oligonucleotides for ligation (NonTemp1: 66-mer, Non Temp2: 47-mer, NonTemp3: 61-mer, Temp4: 53-mer, Temp5: 50-mer, and Temp6: 71-mer) (see Note 11).
2. Rapid DNA Ligation Kit (Roche Applied Biosciences, Germany), including vial 1 (T4 DNA Ligation Buffer, 2× conc.), vial 2 (DNA Dilution Buffer, 5× conc.), and vial 3 (T4 DNA Ligase) (see Note 12).
3. 8% polyacrylamide gel containing 7 M urea.

2.4. Preparation of Ds-Containing Template for Aptamer 9A (105-mer) by PCR Amplification (Fig. 1c)

1. 5′-End primer (39.45): 5′-GGTAATACGACTCACTATA*GG GAGTGGAGGAATTCATCG* (39-mer; The T7 promoter region is underlined, and the coding region of aptamer 9A is *italic*).
2. 3′-End primer containing **Ds** (24.45**Ds**5): 5′-ACDsTC *GCAGAAGCTTGCTGTCGCTAAGGC* (29-mer). The underlined sequence is added to introduce **Ds** into the DNA template (see Note 13).
3. 3′-End primer for a control (24.45): 5′-*GCAGAAGCTT GCTGTCGCTAAGGC* (24-mer).
4. The plasmid DNA encoding the 9A aptamer (the original 100-base sequence is 5′-GGGAGUGGAGGAAUUCAUCG AGGCAUAUGUCGACUCCGUCUUCCUUCAAACCA GUUAUAAAUUGGUUUUAGCAUAUGCCUUAGCGA CAGCAAGCUUCUGC, in which the PCR primer annealing sites are underlined).
5. Premix Taq (Ex Taq Version, TaKaRa Bio Inc., Japan) (see Note 14).
6. 10% polyacrylamide gel containing 7 M urea.

2.5. T7 Transcription Using Biotin-PaTP for RNA Molecules (Fig. 2)

1. 10 mM Biotin-**Pa**TP (Catalog No: 81-3525-02, Glen Research). Store at −10 to −30°C.
2. 25 mM each NTP: mix equal volumes of 100 mM ATP, 100 mM GTP, 100 mM CTP, and 100 mM UTP. Store at −10 to −30°C.
3. 10× T7 RNA Polymerase Buffer (TaKaRa): 400 mM Tris–HCl (pH 8.0), 80 mM $MgCl_2$, and 20 mM spermidine (see Note 15).
4. 50 mM DTT (TaKaRa) (see Note 15).
5. T7 RNA polymerase (50 units/μL) (TaKaRa) (see Note 15).
6. 160 mM $MgCl_2$ (see Note 16).
7. 0.2% Triton X-100 (see Note 16).
8. A DNA template containing **Ds** (**Ds**-template).
9. A DNA template without **Ds** (control-template) (see Note 17).
10. 20 mM EDTA (pH 8.0) (see Note 18).

11. 8 and 10% polyacrylamide gel with 7 M urea for the 152-mer and 105-mer RNAs.

2.6. Analysis of the Biotinylated RNA–Streptavidin Complexes by Gel Electrophoresis (Fig. 3)

1. SA buffer: 10 mM Tris–HCl (pH 8.0) and 50 mM NaCl.
2. 1 mg/mL streptavidin: dissolve 1 mg of lyophilized streptavidin (Promega, Madison, WI) in 1 mL of SA buffer. Store aliquots at −10 to −30°C.
3. Purified RNA transcripts.
4. 10% polyacrylamide gel with 7 M urea.

3. Methods

3.1. Synthesis and Purification of Ds-Containing DNA Fragments

1. Prepare DNA fragments containing **Ds**: Purchase from custom suppliers, or chemically synthesize on a 0.2 μmol scale by a conventional phosphoramidite method using the **Ds**-amidite (Glen Research) (see Note 10).
2. Set up the electrophoresis apparatus. Attach the gel plate to the electrophoresis apparatus, fill the upper and lower reservoirs with 1× TBE, and connect to a power supply. Prior to loading the sample, run the gel for 30–60 min until the temperature of the gel reaches 45–50°C.
3. Prepare the loading sample: Dissolve DNA fragments in 200 μL of water and mix with 2× denaturing sample solution. Heat the solution at 75°C for 3 min.
4. Turn off the power supply, flush out the urea leached from the wells with 1× TBE, and then load the sample solution into the slots.
5. Run the gel until the marker dyes reach the predetermined positions. Generally, stop the electrophoresis run when the bromophenol blue dye reaches at near the bottom of the gel.
6. After electrophoresis, remove the gel plate from the apparatus. Detach one glass plate, and cover the gel with Saran Wrap. Turn the glass plate over, and remove the glass plate from the gel. Place the gel on a fluorescent TLC plate. Check the shadow band corresponding to the DNA fragments on the gel by illumination with a handy UV lamp (see Note 5).
7. Excise the bands with a clean razor blade, transfer the gel slices to a 50-mL conical tube, and crush the slices with a spatula. Add 5.4 mL of sterile water, and incubate the tube for 10 h at 37°C with gentle agitation.
8. Pass the eluted solution through a 0.22 μm Steriflip filter (see Note 7).

9. Divide the filtrate (about 4.8 mL) into 1.2 mL aliquots in four 5 mL round-bottom tubes. Add 120 μL of 3 M sodium acetate, 1 μL of glycogen, and 2.5 mL of ethanol per tube. After mixing, store them at –20°C for 1 h. Recover the eluted oligonucleotide by centrifugation (see Note 8).
10. Wash the precipitate with 70% EtOH, and dissolve it in 400 μL of water. Determine the oligonucleotide concentration from its UV absorbance in an appropriate buffer, such as TE buffer.

3.2. Preparation of Ds-Containing DNA Templates for the 152-mer RNA by Ligation (Fig. 1b)

1. Heat a mixture of 900 pmol each of DNA fragments (NonTemp1, NonTemp2, NonTemp3, Temp4, Temp5, and Temp6) (Fig. 1b) in 100 μL of the 1× annealing solution, at 95°C for 3 min, and then gradually cool it to 25°C.
2. Mix with 25 μL of 5× DNA Dilution Buffer (vial 2).
3. Mix with 125 μL of 2× T4 DNA Ligation Buffer (vial 1). Remove a portion (1 μL) of the mixture, as an original sample for the gel analysis to compare with the ligated products in step 4.
4. Add 10 μL of T4 DNA Ligase (vial 3) to the mixture and incubate at 15°C for 1 h. Remove a portion (1 μL) of the mixture, and analyze the reaction products on a denaturing polyacrylamide gel (see Note 19).
5. Incubate the residual mixture at 65°C for 10 min, and then collect the reaction products by ethanol precipitation (see Note 8).
6. Purify the full-length ligated products by electrophoresis on an 8% polyacrylamide gel containing 7 M urea (see Note 19). For this, dissolve the precipitated products in 60 μL of 1× denaturing sample solution, and incubate the solution at 75°C for 3 min before gel loading. After electrophoresis, excise the desired gel band, transfer the slices to a 1.5 mL tube, crush them, and add 500 μL of water.
7. After an incubation at 37°C for 4 h or more with gentle agitation, filter the supernatant with an Ultrafree-MC filter to remove the small gel slices (see Note 7).
8. Collect the eluted fragments by ethanol precipitation (see Note 8).
9. Dissolve the collected DNA fragments in 40 μL of water, and determine their concentration from their UV absorbance.
10. Dilute the DNA solution to a final concentration of 300 nM in the 1× annealing solution. Anneal the DNA fragments by heating a portion of the solution at 95°C for 3 min and gradually cooling it to 25°C, before use as a template in T7 transcription.

3.3. Preparation of Ds-Containing DNA Templates for Aptamer 9A by PCR Amplification (Fig. 1c)

1. Set up two reactions to generate the **Ds**-containing DNA template (PCR#1 for the 105-mer transcript) and the DNA template without **Ds** (PCR#2 for the 100-mer transcript) as a control. In a 100 μL scale reaction, mix 50 μL of Premix Taq, 100 pmol of 5′-end primer (39.45), 100 pmol of 3′-end primer (24.45**Ds**5 for PCR#1 and 24.45 for PCR#2), 100 ng of 9A-coding plasmid DNA, and sterile water (adjust to 100 μL), on ice. The final concentrations of the components are 0.2 mM each dNTP, 1× ExTaq Buffer, 0.025 U/μL polymerase, and 2 mM $MgCl_2$ (see Note 20).
2. Amplify by PCR using the following cycle profiles: 18 cycles of 94°C for 0.5 min (denaturation), 45°C for 0.5 min (annealing), 72°C for 1 min (extension), and finally 72°C for 5 min.
3. Run a portion of the PCR products on a 10% polyacrylamide gel with 7 M urea, to check the products. Usually, a 2.5 μL aliquot of the PCR solution is mixed with 2.5 μL of the 2× denaturing sample solution, heated at 75°C for 3 min, and then loaded on a denaturing gel at 50°C. Stain the gel with SYBR Green II, and verify that there is one major amplified product.
4. Precipitate the total PCR products with ethanol, and dissolve the precipitate in 25 μL of the 1× annealing solution (see Note 21). There is no need to perform the annealing procedure before T7 transcription.

3.4. T7 Transcription Using Biotin-PaTP for RNA Molecules (Fig. 2)

1. Mix the following materials in a microcentrifuge tube on ice (20-μL scale) (see Note 22).

Biotin-**Pa**TP final conc.	0 mM	2 mM	4 mM
Water	8.4 μL	4.4 μL	0.4 μL
10× T7 RNA Polymerase Buffer	2.0 μL	2.0 μL	2.0 μL
50 mM DTT	2.0 μL	2.0 μL	2.0 μL
0.2% Triton X-100	1.0 μL	1.0 μL	1.0 μL
160 mM $MgCl_2$	2.0 μL	2.0 μL	2.0 μL
25 mM each NTP	1.6 μL	1.6 μL	1.6 μL
10 mM Biotin-**Pa**TP	–	4.0 μL	8.0 μL
DNA template	2.0 μL	2.0 μL	2.0 μL
T7 RNA polymerase	1.0 μL	1.0 μL	1.0 μL

2. Incubate at 37°C for 2–6 h.
3. Add 25 μL of 20 mM EDTA to the reaction mixture and mix well (see Note 18).

4. Add 45 μL of 2× denaturing sample solution and mix well.
5. Heat at 75°C for 3 min.
6. Analyze the transcribed products and purify the desired full-length products by gel electrophoresis (Fig. 2) (see Note 23).

3.5. Analysis of the Biotinylated RNA–Streptavidin Complexes by Gel Electrophoresis (Fig. 3)

1. Prepare a 200 nM RNA solution, by dissolving purified RNA transcripts in SA buffer.
2. Mix 2 μL of streptavidin or SA buffer (as a control) with 5 μL of 200 nM RNA and incubate at 25°C for 30 min.
3. Add 7 μL of 2× denaturing sample solution and mix well.
4. Load the sample solution on a 10% polyacrylamide gel containing 7 M urea, and run the gel (see Note 24).
5. Stain the gel with SYBR Green II, and analyze the biotinylated RNA–streptavidin complexes, detected as low-mobility bands as compared with that of the free-RNA (Fig. 3) (see Note 25).

4. Notes

1. We use a commercially supplied vertical slab electrophoresis system: the sizes of the glass plates are 10 cm × 10 cm (thickness: 1 mm), 16 cm × 16 cm (1 mm), and 20 cm × 40 cm (0.5, 1, and 2 mm). For higher resolution to separate a one-base difference in oligonucleotides, a longer plate size is recommended. A power supply with a temperature probe allows automatic and precise control of the temperature during electrophoresis. We usually set the temperature at 45–50°C and use 1× TBE for the running buffer.
2. Acrylamide is a neurotoxin, and care should be taken to avoid exposure. Filtration of the solution is recommended to remove impurities.
3. To prepare the polyacrylamide gel, it is useful to make a 400 mL stock solution, which is kept at 4°C. For example, to make a 10% acrylamide/bis solution with 7 M urea, mix 100 mL of 40% acrylamide/bis solution (5%C), 40 mL of 10× TBE, and 168 g of urea, and bring the solution to 400 mL with water. Filtration is recommended to remove impurities. For >120-mer oligonucleotides, an 8% denaturing polyacrylamide gel is generally used, with 10% for 45 to 120-mer, 15% for 25 to 50-mer, and 20% for less than 30-mer.
4. For example, to prepare a 16 cm × 16 cm (1 mm) gel, mix 20 mL acrylamide/bis solution with 100 μL of APS. Just before pouring, add 20 μL of TEMED, and mix the solution by gentle swirling. Immediately pour the gel mix between

the gel plates and insert the gel comb. For purification of oligonucleotides, making the gel on the previous day is recommended to ensure complete polymerization, although the gel usually polymerizes in about 30 min. After removing the comb, promptly wash out the wells thoroughly.

5. High concentrations (>0.1 OD units) of oligonucleotides can be visualized by ultraviolet (UV) shadowing, without staining. Place the gel, covered with Saran Wrap, onto the fluorescent TLC plate, and shine UV light from a handy lamp (254 nm). Regions with high concentrations of nucleotides leave a "shadow" on the plate, due to their absorbance of the illuminated UV.
6. Staining a denaturing gel with SYBR Green II (excitation: 497 nm; emission: 513 nm) can detect ssDNA, dsDNA, and RNA with higher sensitivity than that with ethidium bromide, even with a 254 nm or 300 nm transilluminator. Detection by a bio-imaging analyzer, such as an LAS4000 or FLA 7000 (FujiFilm, Japan), is also acceptable.
7. Before ethanol precipitation of the purified fragments eluted from the gel slices, the solution should be filtered through a filter unit, to remove small gel slices. These impurities are also precipitated; inhibit the concentration determination by UV absorbance.
8. For ethanol precipitation, mix the solution with 1/10 vol 3 M sodium acetate and 2–3 vol ethanol, and store the solution at −10 to −30°C (at least for 30 min). Sometimes, 1–2 μL of glycogen is added to the solution, as a coprecipitating agent. After centrifugation (10,000–12,000 × *g* for 40 min) at 4°C, the precipitate is washed with 500 μL of prechilled 70% ethanol. Evaporate the ethanol residue with a centrifugal evaporator.
9. 10 M urea is difficult to dissolve. Incubation at 37°C is helpful, and the solution should be filtered. Incubate the aliquots at 75°C to completely dissolve the urea before use. A denaturing solution with a dye, such as 0.05% (w/v) xylene cyanol (XC) or 0.05% (w/v) bromophenol blue (BPB) as marker dyes, is also alternatively used. Take care that the migration of the dyes does not correspond with that of the desired fragments on the gel.
10. DNA fragments containing **Ds** are chemically synthesized with an automated DNA synthesizer using d**Ds**-amidite. Removal of protecting groups in the synthetic reactions (deprotection) is usually performed by heating the preparation at 55°C for 6 h, after the fragment is released from the CPG column by an incubation at room temperature for an hour in a concentrated NH_4OH solution. The NH_4OH is evaporated to dryness with a centrifugal evaporator.

11. "NonTemp" fragments stand for a nontemplate strand and "Temp" fragments for a template strand in a double-stranded DNA template. Phosphorylation of DNA fragments for ligation can be done enzymatically, by using T4 polynucleotide kinase in the presence of ATP, or chemically, by modification in the DNA chemical synthesis step. As for the base complementary to **Ds** in the fragment, T or C can be used instead of **Pa**, if it is not amplified by PCR. Keep in mind that a **Ds** base located near a ligated site might reduce the ligation efficiency.
12. Other comparable reagents and kits are available from various commercial sources.
13. The **Ds** position in the primer affects the annealing between the target template and the primer. A position closer to the 3′-end of the primer will destabilize the annealing, because of the mismatching of **Ds** with a natural base, and may result in low PCR amplification efficiency. The addition of an extra natural base sequence (two or three natural bases before the **Ds** position) to the 5′-end of the primer (as seen in 24.45**Ds**5) is recommended.
14. In this case, DNA polymerase lacking 3′→5′ exonuclease activity, such as Taq DNA polymerase, is recommended for PCR. Since the 3′→5′ exonuclease-proficient DNA polymerases, such as Pfu, Vent, and DeepVent DNA polymerases, trim the 3′-end of the primers during extended reactions, the **Ds** nucleotide might be removed if it is located near the 3′-end of the primer.
15. Other comparable reagents and kits are available from other commercial sources. Keep in mind that the use of buffers with unusual components and compositions or a mutated T7 RNA polymerase may affect the incorporation efficiency of Biotin-**Pa**TP.
16. Higher concentration $MgCl_2$ and Triton X-100 solutions are recommended, because of the limitation of the reaction volumes. These components affect the transcript yields, and their optimal concentration may be different when using different templates.
17. Comparing the transcripts from DNA templates containing **Ds** and without **Ds** will provide an indication of how efficiently Biotin-**Pa** is incorporated into RNA, depending on the presence of **Ds** in templates.
18. The EDTA is added after the reaction to chelate the magnesium ions. A white precipitate of magnesium pyrophosphate, a by-product generated in the reaction mixture, may spoil the band patterns of the transcripts on the electrophoresis gel.

19. The band of the fully ligated product exhibits the lowest mobility on the gel.
20. No triphosphates of the unnatural base are included in the reaction mixtures for PCR, and thus, some of the natural dNTPs are mis-incorporated opposite **Ds** in the primer during the reaction. However, this is not a problem because this extended fragment without unnatural bases is a nontemplate strand and is not used as a template strand in T7 transcription (see Fig. 1c).
21. Note that estimating the concentration of the template from its UV absorbance may cause overestimation, due to the presence of excess amounts of unused dNTPs and primers. These impurities can be removed by a commercially available purification kit for PCR products.
22. As an alternative method, the addition of 2–4 μCi [α-^{32}P] NTP or [γ-^{32}P] GTP in the reaction mixture yields internally labeled or 5′-labeled transcripts. The transcript yield is dependent on many factors, such as the sequence of the template (independent of **Ds** or dependent on **Ds**), the position of **Ds** in the template, the NTP to $MgCl_2$ concentrations, and the ratio of Biotin-**Pa**TP to each natural NTP. Avoid the introduction of the unnatural base in the aborting and early transcribed region (around the first 10–12 bases) as much as possible. For the highest yields, pilot studies to optimize the component concentrations are recommended.
23. Before loading the sample solution onto the gel, desalting by filtration with a Microcon filter (Millipore) or ethanol precipitation can reduce the loading volume on a gel, which results in better separation of the products. For analysis on a gel, a portion of the solution is sufficient to detect the bands. In Fig. 3, a 5 μL aliquot of the solution was loaded for analysis on the gel. If a higher purity of the biotinylated transcript is required, then use a longer gel enough to separate the biotinylated and nonbiotinylated transcripts. For example, the separation of the biotinylated 105-mer transcript from the shorter, non-biotinylated fragments was done on a 10% denaturing gel with a size of 20 cm × 40 cm (1 mm thickness), and the electrophoresis was stopped when the XC dye marker reached the bottom of the gel.
24. The streptavidin (SA) binding to biotin is very strong, and the presence of urea does not inhibit the binding. Denaturation by heating is also applicable, but the band patterns on the gel, corresponding to the complex between SA and the biotinylated RNA, may be affected. Actually, the shifted band patterns in the biotinylated 152-mer RNA did not change with or without heating the sample before gel loading,

but those in the biotinylated 105-mer RNA changed (the shifted bands were smeared).

25. The quantification of the yield for the incorporation of Biotin-**Pa** can be accomplished by measuring the fluorescent intensities of the bands. Keep in mind that the staining patterns may be different for different-mobility bands. It is helpful to compare the intensities of nonshifted bands (free RNA) in the presence and absence of streptavidin (Fig. 3). For more precise quantification, the use of the isotope-labeled RNA is recommended (10). The misincorporation of Biotin-**Pa**TP opposite each natural base is low (0.06–0.12% per position in the 152-mer transcript) (10), but much longer transcripts may increase the nonspecific biotinylation in total. In that case, adjusting the ratio of Biotin-**Pa**TP relative to the natural NTPs (generally by lowering the concentration of Biotin-**Pa**TP in the reaction mixture) is helpful. To confirm the biotinylation site, an RNA sequencing method using the end-labeled RNA and Streptavidin Magnetic Beads (New England Biolabs, Ipswich, MA) is available (10).

Acknowledgments

This work was supported by Grants-in-Aid for Scientific Research (KAKENHI 19201046 (I.H.), and 20710176 (M.K.)) from the Ministry of Education, Culture, Sports, Science, and Technology and by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses, Ministry of Education, Culture, Sports, and Science and Technology of Japan.

References

1. Benner SA, Burgstaller P, Battersby TR, Jurczyk S (1999) Did the RNA world exploit an expanded genetic alphabet? In: The RNA world, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y, pp 163–181
2. Kool ET (2000) Synthetically modified DNAs as substrates for polymerases. Curr Opin Chem Biol 4:602–608
3. Henry AA, Romesberg FE (2003) Beyond A, C, G and T: augmenting nature's alphabet. Curr Opin Chem Biol 7:727–733
4. Hirao I (2006) Unnatural base pair systems for DNA/RNA-based biotechnology. Curr Opin Chem Biol 10:622–627
5. Hirao I (2006) Placing extra components into RNA by specific transcription using unnatural base pair systems. Biotechniques 40:711–717
6. Tor Y, Dervan PB (1993) Site-specific enzymatic incorporation of an unnatural base, N^6-(6-aminohexyl)isoguanosine, into RNA. J Am Chem Soc 115:4461–4467
7. Johnson SC, Marshall DJ, Harms G, Miller CM, Sherrill CB, Beaty EL, Lederer SA, Roesch EB, Madsen G, Hoffman GL, Laessig RH, Kopish GJ, Baker MW, Benner SA, Farrell PM, Prudent JR (2004) Multiplexed genetic analysis using an expanded genetic alphabet. Clin Chem 50:2019–2027

8. Moriyama K, Kimoto M, Mitsui T, Yokoyama S, Hirao I (2005) Site-specific biotinylation of RNA molecules by transcription using unnatural base pairs. Nucleic Acids Res 33:e129
9. Kawai R, Kimoto M, Ikeda S, Mitsui T, Endo M, Yokoyama S, Hirao I (2005) Site-specific fluorescent labeling of RNA molecules by specific transcription using unnatural base pairs. J Am Chem Soc 127:17286–17295
10. Hirao I, Kimoto M, Mitsui T, Fujiwara T, Kawai R, Sato A, Harada Y, Yokoyama S (2006) An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. Nat Methods 3:729–735
11. Yang Z, Sismour AM, Sheng P, Puskar NL, Benner SA (2007) Enzymatic incorporation of a third nucleobase pair. Nucleic Acids Res 35:4238–4249
12. Kimoto M, Mitsui T, Harada Y, Sato A, Yokoyama S, Hirao I (2007) Fluorescent probing for RNA molecules by an unnatural base-pair system. Nucleic Acids Res 35: 5360–5369
13. Hirao I, Mitsui T, Kimoto M, Yokoyama S (2007) An efficient unnatural base pair for PCR amplification. J Am Chem Soc 129: 15549–15555
14. Leconte AM, Hwang GT, Matsuda S, Capek P, Hari Y, Romesberg FE (2008) Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. J Am Chem Soc 130:2336–2343
15. Kimoto M, Shirouzu M, Mizutani S, Koide H, Kaziro Y, Hirao I, Yokoyama S (2002) Anti-(Raf-1) RNA aptamers that inhibit Ras-induced Raf-1 activation. Eur J Biochem 269:697–704

Part VI

Random Mutagenesis

Chapter 26

Mutagen™: A Random Mutagenesis Method Providing a Complementary Diversity Generated by Human Error-Prone DNA Polymerases

Philippe Mondon, David Grand, Nathalie Souyris, Stéphane Emond, Khalil Bouayadi, and Hakim Kharrat

Abstract

Random mutagenesis is one of the most effective methodologies to generate variant libraries for directed protein evolution. Indeed, this approach requires no structural or mechanistic information and can uncover unexpected beneficial mutations. Here, we describe a new random mutagenesis method based on the use of human error-prone DNA polymerases (pol *beta*, pol *eta* and pol *iota*). This approach allows the random introduction of mutations through a single replication step followed by a selective PCR amplification of the replicated mutated sequences. The libraries generated using this methodology display different mutation rates and complementary mutational spectra. By taking advantage of the mutation bias of naturally highly error-prone DNA polymerases, MutaGen™ thus appears as a very useful tool for gene and protein randomization.

Key words: Directed molecular evolution, Random mutagenesis, DNA polymerase, Pol *eta*, Pol *beta*, Pol *kappa*, Pol *iota*, DNA replication

1. Introduction

DNA polymerases catalyze the template-directed incorporation of deoxynucleotides into a growing primer terminus to generate a new complementary daughter strand. These enzymes are classified on the basis of their structural similarities into seven polymerase families (A, B, C, D, X, Y, and RT) (1). The fidelity of DNA polymerases, defined as the inverse of mis-insertion frequency, is ensured by base pair complementarity, by substrate-induced conformational changes and, in some cases, by proofreading catalyzed by a 3′→5′ exonuclease domain (2). Such fidelities span a wide

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_26,

range. Replicative DNA polymerases from structural families A, B, and C are high fidelity enzymes exhibiting error frequencies of around 10^{-6} (one error per million nucleotides incorporated). The least accurate DNA polymerases belong to the X and Y families. Mammalian DNA polymerase *beta* (pol *beta*), which is involved in the base excision repair (BER) pathway (3), is the best known member of the X-family and exhibits an error frequency ranging from 10^{-3} to 10^{-4} (4, 5). Y-family polymerases, also known as *t*rans*l*esion *s*ynthesis (TLS) polymerases, replicate DNA in a distributive manner and lack any exonucleolytic activity for proofreading (6). These enzymes display the highest error rates measured among DNA polymerases (10^{-1}–10^{-3}). In human, known members of this family are DNA polymerases *eta*, *iota*, or *kappa* (7).

The hypermutation profile of these repair DNA polymerases has been used to design a new in vitro random mutagenesis technique, MutaGen™ (8). This method consists of a first single mutagenic replication step catalyzed by low-fidelity human DNA polymerases, which is followed by a selective PCR amplification of the replicated mutated sequences (see Fig. 1). MutaGen™ was recently reported as an efficient method to enlarge the initial diversity of human fragment antibody libraries (9, 10). This new methodology was also applied on amylosucrase, an enzyme of industrial interest, to generate libraries displaying different mutation rates and complementary mutational spectra (11). The screening of these amylosucrase variant libraries for thermostability yielded three variants showing 3.5- to 10-fold increased half-lives at 50°C compared to the wild-type enzyme (12, 13).

By taking advantage of the mutational bias of several error-prone DNA polymerases, the MutaGen™ approach thus provides different and complementary base substitution spectra according to the replication conditions (see Table 1 and Fig. 2). The variants obtained through this approach combine all types of mutations, including deletions of codons. The widest diversity is reached by pooling libraries generated using different DNA polymerases in a final library. This chapter describes the cloning, expression, and purification of human DNA polymerases *beta*, *eta*, and *iota* and their use to generate random mutant libraries.

2. Materials

2.1. Amplification and Cloning of cDNA of Human DNA Polymerase beta, eta, and iota

1. Testis human cDNA libraries (Clontech).
2. Brain human cDNA libraries (Clontech).
3. *Taq* platinum HF DNA polymerase (Invitrogen).
4. PCR Thermocycler (PT100 MJ Research).

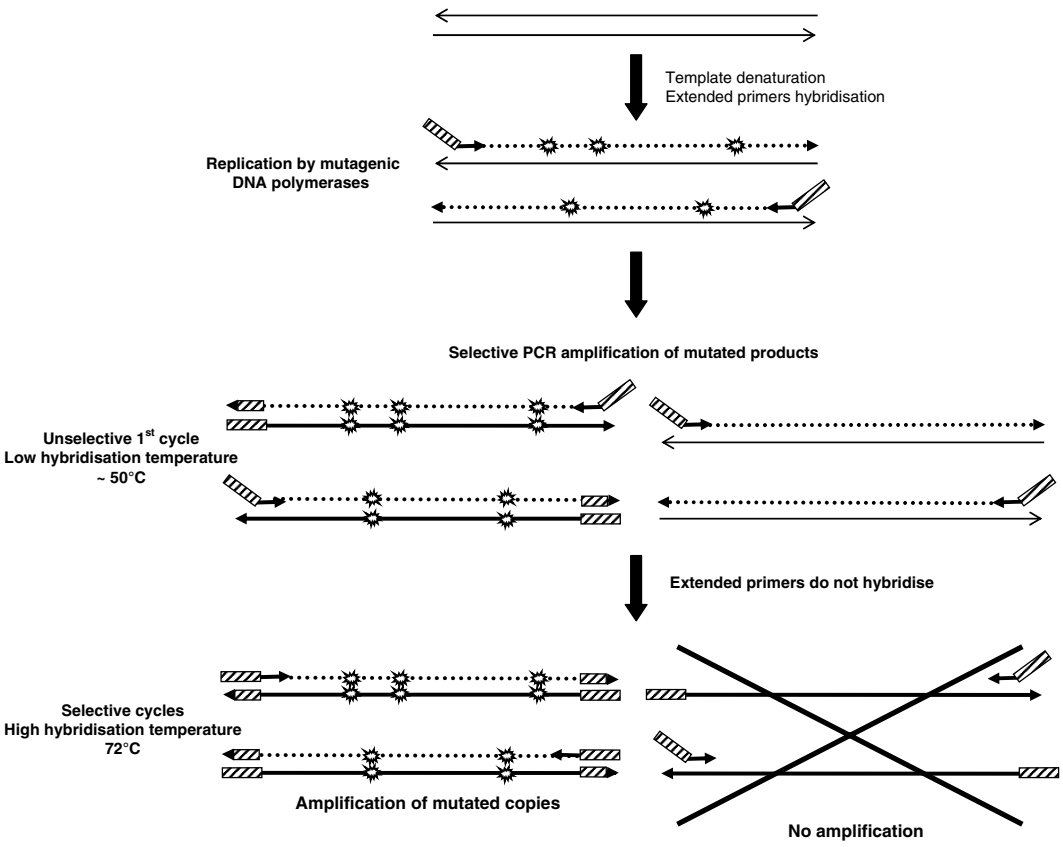


Fig. 1. The MutaGen™ process. MutaGen™ consists of a first single mutagenic replication step followed by selective PCR amplification of the replication products. This technique uses primers possessing a 5′ tail that is not complementary to the template. After template denaturation and primers hybridization, replication is carried out by a low-fidelity DNA polymerase. The mutated copies are then selectively amplified by a PCR procedure consisting of a first cycle with a low hybridization temperature followed by selective cycles (up to 25) with high hybridization temperature disabling template amplification.

Table 1
Range of mutation frequencies using different replication conditions (base mutation frequency/kb)

	Pol *beta*	Pol *eta*	Pol eta-*iota*
Condition A	2–4	ND	ND
Condition B	15–20	ND	ND
Condition E	6–10	ND	ND
Condition N	ND	7–9	10–12

ND not determined

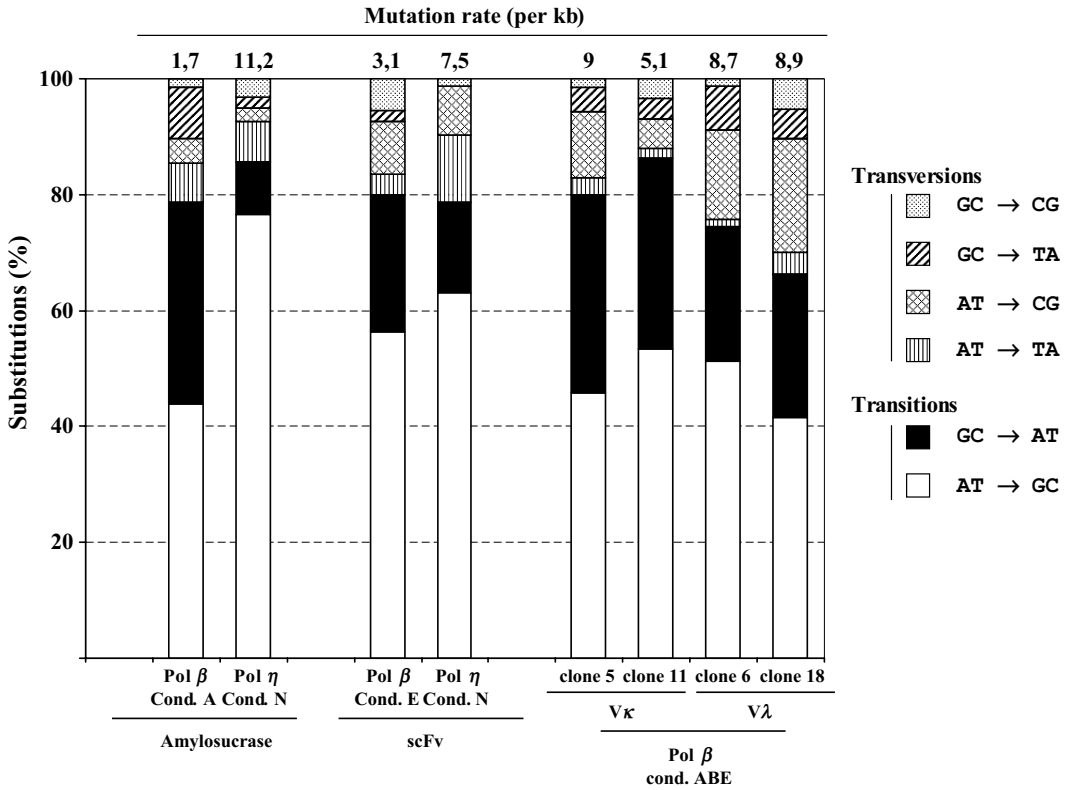
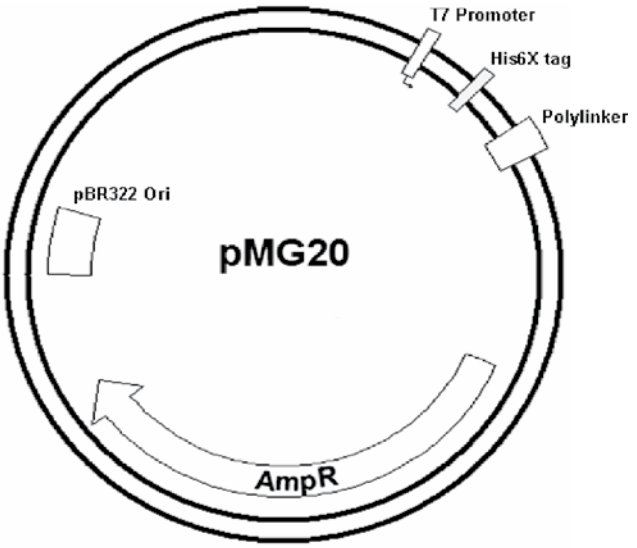


Fig. 2. Complementary diversity generated using MutaGen™ on several DNA templates encoding either an amylosucrase, antibody single chain variable fragment (scFv), or antibody light chain variable domains (Vκ and Vλ). *Cond* mutagenesis condition, *ABE* pool of libraries generated using A, B, and E conditions.

5. dNTPs (New England Biolabs).
6. 1% agarose gel: 2% (w/v) agarose in 1× TAE buffer.
7. Ethidium bromide solution (Euromedex).
8. 50× TAE buffer: dissolve 242 g of Tris(hydroxymethyl) aminomethane, 57.1 mL of acetic acid, and 7.43 g of Na_2EDTA in distilled water; make up to 1 L.
9. Plasmid miniprep kit (Qiagen).
10. TOP10 chemically competent bacteria (Invitrogen).
11. LB agar plates with 100 µg/mL ampicillin.
12. SOC medium: 2% tryptone, 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose.
13. Vectors: pMG20A and pMG20B (see Fig. 3).
14. Klenow polymerase (New England Biolabs).
15. Restriction endonucleases: *Nde*I, *Hind*III, *Eco*RI and *Bam*HI, and restriction buffers (New England Biolabs).



MCS pMG20A

```
 BamHI     XhoI             KpnI  NdeI  EcoRI  HindIII
GGGATCCGAGCTCGAGATCTGCAGCTGGTACCATATGGGAATTCGAAGCTTCGATCCGGC
 G   S   E   L   E   I   C   S   W   Y   H   M   G   I   R   S   F   D   P   A
```

MCS pMG20B

```
BamHI     XhoI             KpnI NdeI   EcoRI  HindIII
GGATCCGAGCTCGAGATCTGCAGCTGGTACCATATGGGAATTCGAAGCTTCGATCCGGCT
 D   P   S   S   R   S   A   A   G   T   I   W   E   F   E   A   S   I   R   L
```

MCS pMG20C

```
BamHI     XhoI             KpnI NdeI   EcoRI  HindIII
GGATCCGACCTCGAGATCTGCAGATGGTACCATATGGGAATTCGAAGCTTCG
 I   R   P   R   D   L   Q   M   V   P   Y   G   N   S   K   L   R
```

Fig. 3. Map and multiple cloning site (MCS) sequence of pMG20A, B, and C expression vectors.

16. Ultraclean gel spin DNA purification kit (MO BIO Laboratories).
17. Calf Intestinal Phosphatase (CIP; New England Biolabs).
18. Primers for amplification of human DNA polymerase *eta*, *beta*, and *iota* (see Table 2).
19. T4 DNA ligase and ligation buffer (New England Biolabs).

2.2. Human DNA Polymerases Expression and Purification

1. *E. coli* BL21 (DE3) (Stratagene).
2. 2× YT agar medium: dissolve 15 g of agar, 16 g of bacto-tryptone, 10 g of yeast extract, and 5 g of NaCl in 1 L of distilled water; sterilize by autoclaving.

Table 2
Primers used for the cloning of DNA polymerase *beta*, *eta*, and *iota* cDNAs and for the construction of random mutagenesis libraries

Primer name	Sequence	Restriction sites
ETAS1	5′-AATAGGATCCATGGCTACTGGACAGGATCG-3′	*Bam*HI
ETAR1	5′-AATAGAATTCCTAATGTGTTAATGGCTT-AAAAAATGATTCC-3′	*Eco*RI
BetaS1	5′-TAGATCATATGAGCAAACGGAAGGCGCCG-3′	*Nde*I
BetaR1	5′-GACTAAGCTTAGGCCTCATTCGCTCCGGTC-3′	*Hind*III
IOTAS1	5′-ATATGGATCCATGGAACTGGCGGACGTGGG-3′	*Bam*HI
IOTAR1	5′-TAATAAGCTTTTATTTATGTCCAATGTGG-AAATCTGATCC-3′	*Hind*III
MutpUC18_S1	5′-*TCTGACGAGTACTAGCTGCTACA*TGCAGGT-CGACTCTAGAGGATCC-3′	*Bam*HI
MutpUC18_R1	5′-*ACAGCTACGTGATACGACTCACA* CTATGA-CCATGATTACGAATTCC-3′	*Eco*RI

3. 2× YT/amp/1% glu plates: 2× TY agar media containing 100 µg/mL ampicillin and 1% glucose. The antibiotic and glucose are added only when the autoclaved 2× YT agar solution cools to below 50°C.
4. Isopropyl-beta-D-thiogalactopyranoside (IPTG): 1 M stock solution stored at −20°C.
5. Refrigerated rotating incubator shaker (Sartorius).
6. Sterile glass Erlenmeyer flask 1 L.
7. 50 mL conical centrifuge tubes (Falcon).
8. Lysis buffer: 50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 10 mM Imidazole, 0.05% TritonX100, 1 mM EDTA, 1 mM DTT, and 1 mg/mL Lysozyme.
9. Washing buffer: 50 mM NaH_2PO_4/Na_2HPO_4, pH 8.0, 300 mM NaCl, 20 mM Imidazole, 0.05% TritonX100, 1 mM EDTA, and 1 mM DTT.
10. Elution buffer: 50 mM NaH_2PO_4/Na_2HPO_4, pH 8.0, 300 mM NaCl, and 250 mM Imidazole.
11. Dialysis buffer: 40 mM Tris–HCl, pH 8.0, 2 mM DTT, 0.2 mM EDTA, and 200 mM NaCl.
12. Ni-NTA resin (Qiagen).
13. Anti protease cocktail (SIGMA).
14. Ultra sonic homogenizer (Bandelin SONOPULS HD2200, probe TT13 flat tip).

15. 0.45 µm syringe filter (Luer Lock).
16. 50 mL Syringe (Terumo).
17. Empty chromatography polyprep column (Bio-Rad).
18. Bradford protein assay (Bio-Rad).
19. Dialysis membrane (Visking).
20. Glycerol (Euromedex).
21. SDS–Polyacrylamide gel Electrophoresis (SDS–PAGE): 40% acrylamide/bis 29:1 solution (Bio-Rad), TEMED (Euromedex), ammonium persulfate 10% (Sigma), 20% SDS (Euromedex).

2.3. Mutant Library Construction

1. QIAprep Spin Midiprep Kit (Qiagen).
2. One Shot *E. coli* TOP10 strain (Invitrogen).

2.3.1. Preparation of Template

3. LB medium: 5 g of NaCl, 10 g of bacto-tryptone, and 5 g of yeast extract per 1 L; sterilize by autoclaving.
4. LB/amp medium: LB media containing 100 µg/mL ampicillin (stock at 50 mg/mL in distilled water).
5. LB/amp plates: Prepare LB media containing 1.5% (w/v) agar and 100 µg/mL ampicillin (stock at 50 mg/mL in distilled water). The antibiotic is added only when the autoclaved LB-agar solutions cools to below 50°C.
6. Plasmid vector harboring the gene X to be mutated: pUC18-X (cloned into *Bam*HI, *Eco*RI restriction sites).

2.3.2. DNA Replication Assay with Error Prone Polymerases

1. Human DNA polymerases *beta*, *eta*, and *iota*.
2. Human DNA polymerase *beta* replication buffer: 50 mM Tris–HCl, pH 8.8, 10 mM $MgCl_2$, 100 mM KCl, 1 mM DTT, and 10% glycerol.
3. Human DNA polymerase *eta* and *iota* replication buffer: 25 mM Tris–HCl, pH 7.2, 1 mM DTT, 5 mM $MgCl_2$, and 2.5% (v/v) Glycerol.
4. Replication conditions (see Table 3) (8, 9).
5. Plasmid DNA templates (pUC18-X).

Table 3
Replication conditions used for random mutagenesis

	dATP (µM)	dCTP (µM)	dTTP (µM)	dGTP (µM)	Mn^{2+} (mM)
Condition A	50	50	100	100	
Condition B	20	100	100	100	0.5
Condition E	20	100	100	100	0.25
Condition N	100	100	100	100	

6. dNTPs: 2.5 mM each of deoxyadenosine 5′ triphosphate (dATP), deoxythymidine 5′ triphosphate (dTTP), deoxycytidine 5′ triphosphate (dCTP), and deoxyguanosine 5′ triphosphate (dGTP).
7. Forward primer MutpUC18_S1 and reverse primer Mutp UC18_R1 (Table 2).
8. Sterile distilled water.
9. Phenol-chloroform (Sigma-Aldrich).
10. 3 M sodium acetate, pH 5.2.
11. Absolute and 70% (v/v) ethanol (Prolabo).
12. ISS110 Speed Vac system (Thermo Savant).

2.3.3. Selective Amplification of Replication Products

1. Platinum Taq DNA polymerase (Invitrogen).
2. 10× Platinum *Taq* polymerase buffer: 200 mM Tris–HCl, pH 8.4, and 500 mM KCl (Invitrogen).
3. 50 mM $MgCl_2$ (Invitrogen).
4. dNTPs (see Subheading 2.3.2, item 6).
5. Forward primer MutpUC18_S1 and reverse primer Mutp UC18_R1 (see Table 2).
6. Sterile distilled water.
7. Centrifugal Filter Devices Microcon PCR (Millipore).

2.3.4. Cloning of Mutant Libraries

1. *Bam*HI and *Eco*RI restriction enzymes, NEB 2 buffer, and 100× BSA solution (New England Biolabs).
2. 50× TAE buffer (see Subheading 2.1, item 8).
3. 1% agarose gel: 1% (w/v) agarose in 1× TAE buffer.
4. Ultraclean gel spin DNA purification kit (MO BIO Laboratories).
5. Sterile distilled water.
6. T4 DNA ligase and ligation buffer (New England Biolabs).
7. Microcon PCR spin-column device (Millipore).
8. *E. coli* XL1-Blue electrocompetent cells (Stratagene).
9. Electroporator 2510 (Eppendorf) and 0.2 cm-gap electroporation cuvettes (Cell Projects, UK).
10. SOC medium (see Subheading 2.1, item 13).
11. 2× YT agar medium (see Subheading 2.2, item 2).
12. 2× YT/amp/1% glu plates (see Subheading 2.2, item 3).
13. 2× TY/amp/1% glu/15% glycerol: 16 g of bacto-tryptone, 10 g of yeast extract, and 5 g of NaCl in 1 L of distilled water; sterilize by autoclaving. The antibiotic, glucose, and glycerol are added only when the autoclaved 2× YT agar solution cools to below 50°C.

2.3.5. Analysis of Mutant Library

1. Montage Plasmid Miniprep HTS 96 Kit (Millipore).
2. LB/amp medium: LB media containing 100 µg/mL ampicillin (stock at 50 mg/mL in distilled water).
3. Incubator shaker for 96 well plates (TiMix and Th15 Edmund Buhler GmbH).
4. Integrated robotic platform (TECAN Genesis RSP-200).
5. Sequencing primers M13 (-21) 5′-TGTAAAACGACGGCCAG-3′ and M13R (-29) 5′-CAGGAAACAGCTATGACC-3′.
6. V3.1 Cycle Sequencing Kit (Applied Biosystems).
7. DNA analyzer 96-capillary 3730xl (Applied Biosystems).
8. MutAnalyse Software (MilleGen).

3. Methods

3.1. Human DNA Polymerases Cloning

1. cDNA coding for the polymerase *beta* (Pubmed GeneID: 5423) is amplified by PCR from a brain cDNA library according to Clontech recommendations using HF *Taq* DNA polymerase and the specific primers BETAS1 and BETAR1 (see Table 2). The PCR cycles are (94°C for 2 min) 1 cycle, (94°C for 20 s, 64°C for 20 s, 68°C for 1 min) 30 cycles, and then the reaction is stored at 4°C. The final amplified cDNA of 1,008 bp is digested with 40 units of restriction enzyme *Nde*I, blunt ended with 5 units of Klenow followed by *Hind*III digestion (40 units). The *Nde*I (blunt end)-*Hind*III fragment is gel-purified using the Ultraclean gel spin DNA purification kit. The plasmid pMG20B is digested with 40 units of *Eco*RI, blunt ended with 5 units of Klenow, followed by *Hind*III digestion (40 units). pMG20B is dephosphorylated with 10 units of CIP and gel-purified. The *Nde*I (blunt end)-*Hind*III fragment is ligated with 1 unit of T4 DNA ligase using molar ratio of 1(vector):3(insert) into the pMG20B vector.
2. cDNA coding for polymerase *eta* (Pubmed GeneID: 5429) is PCR amplified from a testis cDNA library according to Clontech recommendation using HF *Taq* DNA polymerase and the specific primers ETAS1 and ETAR1 (see Table 2). The PCR cycles are (94°C for 2 min) 1 cycle, (94°C for 30 s, 64°C for 30 s, 68°C for 2 min 30 s) 30 cycles, and then the reaction is stored at 4°C. The final amplified cDNA of 2,142 bp is digested with 40 units of *Bam*HI and 40 units *of Eco*RI. The *Bam*HI–*Eco*RI fragment is then gel-purified and ligated using 1 unit of T4DNA ligase with a molar ratio of 1(vector):3(insert) into the pMG20A vector. pMG20A was previously digested using 40 units of *Bam*HI and 40 units

*Eco*RI and dephosphorylated with 10 units of CIP and gel-purified.

3. cDNA coding for polymerase *iota* (Pubmed GeneID: 11201) is PCR amplified from a testis cDNA library according to Clontech recommendation using HF *Taq* DNA polymerase and the specific primers IOTAS1 and IOTAR1 (see Table 2). The PCR cycles are (94°C for 2 min) 1 cycle, (94°C for 30 s, 62°C for 30 s, 68°C for 2 min 30 s) 30 cycles, and then the reaction is stored at 4°C. The final amplified cDNA of 2,148 bp is digested with 40 units of *Hind*III and 40 units of *Bam*HI. The *Hind*III–*Bam*HI fragment is gel-purified and ligated using 1 unit of T4 DNA ligase with a molar ratio of 1(vector):3(insert) into pMG20A vector. The pMG20A vector was previously digested using 40 units of *Hind*III and 40 units of *Bam*HI and dephosphorylated with 10 units of CIP and gel-purified before the ligation step.
4. Chemically competent *E. coli* TOP10 cells are thawed on ice and transformed with the ligation products of pol *beta*, *eta,* and *iota* in their respective vectors.
5. Plasmid minipreps are done for clones validated by double strand DNA sequencing at MilleGen's sequencing service.

3.2. Expression and Purification of Human DNA Polymerases: beta, eta, and iota

1. Each of the vectors allowing expression of human DNA polymerases is transformed into *E. coli* BL21(DE3). Cells are then plated onto 2YT agar, ampicillin 100 µg/mL and grown overnight at 37°C.
2. Fresh single colony transformants are picked and precultured in 10 mL 2× YT medium (ampicillin 100 µg/mL). Incubate overnight at 37°C under shaking at 230 rpm in 50 mL Falcon tubes.
3. OD 600 nm of each overnight culture is measured, and the appropriate dilution is used to inoculate the expression culture into 1 L Erlenmeyer flasks containing 300 mL 2YT + ampicillin (100 µg/mL) at an OD 600 nm equal to 0.1. The cultures were incubated at 37°C with rotary shaking at 230 rpm.
4. When OD 600 nm reaches 0.8, each culture is shifted to 15°C by incubation in a cold water bath. Inductions are done by adding 0.2 mM of IPTG and incubation for 5 h at 15°C with rotary shaking at 230 rpm. Cultures are centrifuged at 3,000 × *g* for 15 min. The cell pellet obtained is stored at –20°C for purification.
5. A cell pellet from each expression culture is homogenously resuspended into 20 mL lysis buffer complemented with anti proteases (following manufacturer recommendations).
6. Cells are lysed in a beaker immersed in cold water and ice, using an ultrasonic homogenizer setting the power at 25% and applied four times for 2 min each.

7. Lysates are subjected to centrifugation at 16,200 × *g* for 30 min at 4°C.
8. Supernatants are filtered through 0.45 µm filters with 50 mL syringes and each passed through 1 mL Ni-NTA resin columns (50% v/v) preequilibrated with lysis buffer.
9. Resins are washed with at least 25 mL of washing buffer.
10. Elutioned fractions of 500 µL are recovered and each fraction's protein quantity and quality respectively evaluated by a Bradford assay and SDS–PAGE followed by Coomassie blue staining.
11. According to the quantification results, the fractions corresponding to each of the DNA polymerases are separately pooled and subjected to overnight dialysis into dialysis buffer.
12. Glycerol is added to the dialyzed fractions until 50% is achieved and the final protein quantity measured by Bradford protein assay.
13. Store pooled fractions at −80°C.

3.3. Mutant Library Construction

A mutant library is generated by double replication of the parental DNA template (see Note 1) using a human DNA polymerase (pol *beta*, *eta*, or *iota*) and a pair of primers (see Note 2). The DNA polymerase replicates both DNA strands starting from the primer (see Fig. 1). The replication products are then selectively amplified by PCR amplification with tail primers (see Note 3). The primers are designed with a tail (20–25 bp) that is not specific to the template and allows specific amplification (higher PCR stringency) of DNA fragments synthesized by the polymerases. The selectively amplified PCR products are finally cloned in pUC18 (in this example) but can also be cloned into any expression vector suitable for phage display (pHEN, pComb), yeast display (pYD1), or High-Throughput Screening.

3.3.1. Preparation of DNA Template

1. Inoculate a single colony of *E. coli* TOP10 harboring the pUC18-X vector (containing the gene X to be mutated) picked from an LB/amp plate into 5 mL LB/amp media and incubate with shaking at 200 rpm and 37°C for 16 h.
2. Prepare plasmid DNA using Midiprep kit according to manufacturer's instructions.
3. Store the purified pUC18-X vector at −20°C.

3.3.2. DNA Replication Assay with Error Prone Polymerases

1. Perform the replication reaction in a total volume of 20 µL containing 1 µg of the plasmid DNA (pUC18-X) as template, 200 nM of each primer, and 2 µL of 10× replication buffer and dNTP's according to the chosen condition (see Table 2).
2. Template DNA is denaturated by incubation for 5 min at 95°C followed by an immediate switch to 4°C.

3. Add 10 µL of a solution containing 4 units (see Note 4) of polymerase (*beta*, *eta*, or *iota*) and 1 µL of 10× replication buffer (see Note 5).
4. Incubate the DNA replication mixture at 37°C for 1 h.
5. Inactivate the DNA polymerase, and purify DNA from the reaction mixture by phenol/chloroform extraction.
6. Precipitate the DNA by adding 0.1× volume of 3 M sodium acetate and 2.5× volume of absolute ethanol. Incubate at least 2 h at –20°C, and centrifuge the precipitate for 20 min at 13,000 × *g* at 4°C. Wash the pellet two times with 500 µL of 70% ethanol, dry the pellet 3 min in a Speedvac rotary evaporator, and dissolve the dry pellet in 20 µL sterile milliQ water.

3.3.3. Selective Amplification

1. Perform the selective PCR in a total volume of 30 µL containing 1 µL of 5 ng of the purified replication product, 3 µL of 10× Platinum *Taq* polymerase buffer, 200 µM of dNTP, 0.2 µM of each tail primer MutpUC18_S1 and MutpUC18_R1, and 1 U of Platinum *Taq* polymerase.
2. Selective PCR process cycles are (94°C for 2 min) 1 cycle, (94°C for 2 min, 58°C for 10 s, 72°C for 2 min) 1 cycle, and then 25 higher stringency cycles: 20 s at 94°C and 1 min 30 s at 72°C (see Note 6).
3. Purify the PCR products using the Microcon filtration device following the manufacturer instructions.

3.3.4. Cloning of Mutant Libraries

1. Digest the totality of the purified PCR products and 2 µg of pUC18 vector with 60 units of *Bam*HI and 60 units *Eco*RI in the supplied 1× NEB 2 buffer and 100 µg/mL of BSA solution.
2. Incubate the reactions at 37°C for 6 h.
3. Purify the digested DNA (PCR products and pUC18) on a 1% agarose gel. Extract the DNA using ultraclean gel spin DNA purification kit. Resuspend the product in water, and determine the DNA concentration by analysis on a 1% agarose gel with markers of known size and concentration.
4. Ligate the digested PCR products and the digested vector with a molar ratio of 1(vector):3(insert) with T4 DNA ligase (see Note 7).
5. Incubate at 16°C overnight.
6. Purify the ligated product using the Microcon filtration device following the manufacturer instructions.
7. Add 2 µL of the ligation product to 50 µL of XL1-Blue electrocompetent cells. Transfer to a 0.2 cm-gap electroporation cuvette.

8. Electroporate cells at 1.8 kV, 25 µF, and 200 W.
9. Add 450 µL of SOC media to each transformation and incubate with shaking at 37°C for 1 h.
10. Pool the transformants and plate on solid 2× TY medium containing 100 µg/mL ampicillin and 1% glucose. Plate serial dilutions onto additional plates containing the same medium to determine the library size.
11. Incubate the plates overnight at 37°C.
12. After counting colonies on the dilution plate, scrape the colonies from the plates with 2× TY/amp/1% glu media containing 15% glycerol (v/v), and store the bacterial library stock at –80°C.

3.3.5. Analysis of the Library

1. From the serial dilution plates, pick 4×96 clones randomly chosen and inoculate them into four 96 deep well plates (Montage Plasmid Miniprep HTS 96 Kit) containing LB liquid medium (100 µg/mL ampicillin).
2. Incubate at 37°C overnight under agitation (800 rpm) using an Incubator shaker for 96 well plates.
3. Perform the plasmid preparation using Montage Plasmid Miniprep HTS 96 Kit on an integrated robotic TECAN Genesis RSP-200 platform.
4. Sequence the mutated portions of the DNA using the primers M13 (-21) and M13R (-29) and MilleGen's high-throughput sequencing service.
5. Analyze the sequences using the MutAnalyse software: mutagenesis frequency, library quality (frequency of deletions, insertions and substitutions), frequency of wild-type sequence, number of mutations per variant, and the repartition of substitutions along the mutagenized DNA (see Note 8).

4. Notes

1. The X gene used as a template in the replication reaction can be any cDNA coding for a gene or a fragment of a gene up to 3 kb.
2. The selective tail primers have to be adapted to your plasmid if your template is not cloned into *Bam*HI and *Eco*RI of pUC18.
3. Primers MutpUC18_S1 and MutpUC18_R1 are designed to anneal to pUC18 at the 5′ part upstream of the *Bam*HI site (27 nucleotides) and at the 3′ part downstream of the *Eco*RI site, respectively. They also contain tails of 23 nucleotides that are not complementary with the vector.

4. One unit is defined as the amount of enzyme required to catalyze the incorporation of 1 nmole of dNTP into an acid-insoluble form in 1 h at 37°C.
5. Each DNA polymerase has its own catalytic properties and provides specific mutational spectra. To obtain the widest diversity, the final library should be the result of the compilation of the libraries generated using each DNA polymerase.
6. The selective PCR process is based on two different annealing temperatures. The annealing temperature of the first cycle (58°C) allows the hybridization of the 3′ end of the primers, and then the following higher stringency cycles use an annealing temperature of 72°C.
7. The final size of the mutant library depends on the total ligation volume and the number of transformations performed. The size of the final library can be defined according to your screening capabilities.
8. The mutagenesis frequency depends on the DNA polymerase used and the composition of the DNA. The range of mutagenesis frequency generally obtained is indicated in Table 1.

References

1. Rothwell PJ, Waksman G (2005) Structure and mechanism of DNA polymerases. Adv Protein Chem 71:401–440
2. Kunkel TA (2004) DNA replication fidelity. J Biol Chem 279:16895–16898
3. Matsumoto Y, Kim K (1995) Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. Science 269:699–702
4. Kunkel TA (1985) The mutational specificity of DNA polymerase-beta during in vitro DNA synthesis. Production of frameshift, base substitution, and deletion mutations. J Biol Chem 260:5787–5796
5. Osheroff WP, Jung HK, Beard WA, Wilson SH, Kunkel TA (1999) The fidelity of DNA polymerase beta during distributive and processive DNA synthesis. J Biol Chem 274: 3642–3650
6. Yang W (2005) Portraits of a Y-family DNA polymerase. FEBS Lett 579:868–872
7. Kunkel TA, Pavlov YI, Bebenek K (2003) Functions of human DNA polymerases eta, kappa and iota suggested by their properties, including fidelity with undamaged DNA templates. DNA Repair (Amst) 2:135–149
8. Bouayadi K, Kharrat H, Louat T, Servant L, Cazaux C, Hoffmann JS (2002) Use of mutagenic polymerase for producing random mutations. Patent WO/2002/0238756
9. Mondon P, Bouayadi K, Kharrat H (2007) Highly diversified antibody libraries. Patent WO/2007/137616
10. Mondon P, Souyris N, Douchy L, Crozet F, Bouayadi K, Kharrat H (2007) Method for generation of human hyperdiversified antibody fragment library. Biotechnol J 2:76–82
11. Emond S, Mondon P, Pizzut-Serin S, Douchy L, Crozet F, Bouayadi K, Kharrat H, Potocki-Veronese G, Monsan P, Remaud-Simeon M (2008) A novel random mutagenesis approach using human mutagenic DNA polymerases to generate enzyme variant libraries. Protein Eng Des Sel 21:267–274
12. Emond S, Andre I, Jaziri K, Potocki-Veronese G, Mondon P, Bouayadi K, Kharrat H, Monsan P, Remaud-Simeon M (2008) Combinatorial engineering to enhance thermostability of amylosucrase. Protein Sci 17: 967–976
13. Emond S, Potocki-Veronese G, Mondon P, Bouayadi K, Kharrat H, Monsan P, Remaud-Simeon M (2007) Optimized and automated protocols for high-throughput screening of amylosucrase libraries. J Biomol Screen 12:715–723

Chapter 27

Random-Scanning Mutagenesis

Robert A. Smith

Abstract

Oligonucleotide-mediated mutagenesis is a useful tool for engineering nucleotide changes at defined positions in a DNA sequence. Oligonucleotide-based approaches are commonly used to introduce missense mutations at individual codons in a gene or gene segment, thereby revealing the functional importance of specific amino acid residues in a protein. For mutagenesis studies involving tracts of polypeptide sequence, investigators typically change each successive residue to alanine or to a limited number of alternative amino acids. Although these strategies can provide useful information, it is sometimes desirable to test a broader spectrum of amino acid changes at the targeted positions. This article describes a facile, oligonucleotide-based method for generating all 19 possible replacements at individual amino acid sites within a protein. This technique is known as "random-scanning mutagenesis" and is illustrated herein using examples from our studies of a conserved polymerase motif in HIV-1 reverse transcriptase.

Key words: Random-scanning mutagenesis, Protein structure-function, Motif, Human immunodeficiency virus type 1 (HIV-1), Reverse transcriptase

1. Introduction

In vitro mutagenesis encompasses a broad array of strategies for dissecting the functional importance of DNA sequences and the products they encode (1). Among the techniques described in the literature, oligonucleotide and PCR-mediated mutagenesis have emerged as popular approaches for altering the nucleotide sequence of a gene (2–5). These methods are easily adapted to a particular target sequence and have greatly facilitated the studies of gene regulation, protein structure-function relationships, and enzymatic catalysis.

To examine the importance of specific residues in a protein or enzyme, investigators typically introduce nucleotide changes that encode a limited sampling of amino acid replacements at the target codon site. In cases where the area of interest spans a

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_27,

moderate to large stretch of amino acids, oligonucleotide or PCR-directed mutagenesis is often used to change each residue in the target region to alanine (i.e., alanine-scanning mutagenesis) (6), although replacements with other residues have also been described (7, 8). While these mutagenesis approaches often provide valuable information, the decision to use a specific subset of amino acid replacements is arbitrary and can potentially misrepresent the structural or functional importance of the residue(s) in question. For this reason, it is sometimes desirable to introduce random mutations that encode the full spectrum of amino acid replacements at a single codon or stretch of codons (11). This strategy avoids the bias inherent in smaller samplings of replacements and can be used to reveal the range of allowable substitutions in the region of interest.

In recent studies of HIV-1 reverse transcriptase (RT), I devised an approach called "random-scanning mutagenesis" (9, 10) and used this strategy to examine the functional importance of individual amino acid residues in a conserved structural motif. This method is an extension of scanning-saturating mutagenesis (11) with additional provisions for improving library diversity and limiting the frequency of wild-type sequences at the codons of interest. The following sections describe the random-scanning mutagenesis procedure using specific examples from our analysis of HIV-1 RT (9, 10).

2. Materials

1. Luria–Bertani (LB) broth: 10 g/l Bacto-tryptone, 5 g/l Bacto-yeast extract, and 10 g/l NaCl. Adjust pH to 7.5 with 5 N NaOH. Sterilize by autoclaving.
2. 2× yeast–tryptone broth (2× YT): 16 g/l Bacto-tryptone, 10 g/l Bacto-yeast extract, and 5 g/l NaCl. Adjust to pH 7.0 with 5 N NaOH. Sterilize by autoclaving.
3. SOC broth: 20 g/l Bacto-tryptone, 5 g/l Bacto-yeast extract, and 0.5 g/l NaCl. Dissolve solutes, and then add 10 ml/l of 250 mM KCl. Adjust pH to 7.5 with 5 N NaOH. Divide into 100-ml aliquots and sterilize by autoclaving. Immediately before use, add 0.5 ml of sterile 2 M $MgCl_2$ and 2 ml of sterile 1 M glucose solution to each 100 ml bottle of medium.
4. LB-ampicillin (100 µg/ml) agar plates: Prepare LB broth as described above, and then add 15 g/l of Bacto-agar. Sterilize by autoclaving. Allow the agar to cool to 55°C, and then add 4 ml/l of 25 mg/ml ampicillin solution (prepared in dH_20 and filter-sterilized). Mix and pour into petri plates. Cool the agar plates at room temperature and store at 4°C.

5. LB-ampicillin–chloramphenicol agar plates: Prepare LB broth as described above, and then add 15 g of Bacto-agar. Sterilize by autoclaving. Allow the agar to cool to 55°C, and then add 2 ml of 25 mg/ml ampicillin solution and 882 µl of 34 mg/ml chloramphenicol (prepared in 100% ethanol). The final drug concentrations are 50 µg/ml ampicillin and 30 µg/ml chloramphenicol. Mix and pour into petri plates. Cool the agar plates at room temperature and store at 4 C.
6. *Escherichia coli* strains DH10B and CJ236: Electrocompetent ("ElectroMAX") DH10B *E. coli* – Invitrogen. Chemically competent CJ236 cells – TaKaRa Bio Inc.
7. VCSM13 interference-resistant helper phage: Stocks containing approximately 1×10^{11} plaque-forming units per ml (pfu/ml) – Stratagene.
8. PEG/NaOAc solution: 20% (w/v) polyethylene glycol 8000 in 2.5 M sodium acetate. Filter-sterilize.
9. 10× annealing buffer: 200 mM Tris–HCl (pH 7.4), 20 mM $MgCl_2$, and 500 mM NaCl. Sterilize by autoclaving and store at –20°C.
10. 10× synthesis buffer: 5 mM each of dATP, dCTP, dGTP, and dTTP, 10 mM ATP, 100 mM Tris–HCl (pH 7.4), 50 mM $MgCl_2$, and 20 mM dithiothreitol. Divide into 100 µl aliquots and store at –80°C.
11. T7 DNA polymerase (10 units/µl): New England Biolabs.
12. 1× DNA polymerase dilution buffer: 20 mM KH_2PO_4 (pH 7.4), 1.0 mM dithiothreitol, 0.1 mM EDTA, and 50% glycerol. Filter-sterilize and store at –20°C.
13. T4 DNA ligase: 400 cohesive end ligation units/µl, 6 Weiss units/µl – New England Biolabs.
14. Oligonucleotide primers: HPLC-purified, 5′-phosphorylated oligonucleotides – Operon Biotechnologies, Inc.
15. pBluescript II KS(-): Stratagene.
    All other reagents and solutions were prepared from molecular biology-grade chemicals as described elsewhere (12).

3. Method

Initially, the nucleotide sequence of the gene or gene segment is modified by introducing a unique restriction enzyme site within the target region (Fig. 1a). This step can be accomplished via conventional oligonucleotide-directed mutagenesis or by using a PCR-based strategy (2–5). For our analysis of HIV-1 RT, we used oligonucleotide-mediated mutagenesis to replace the

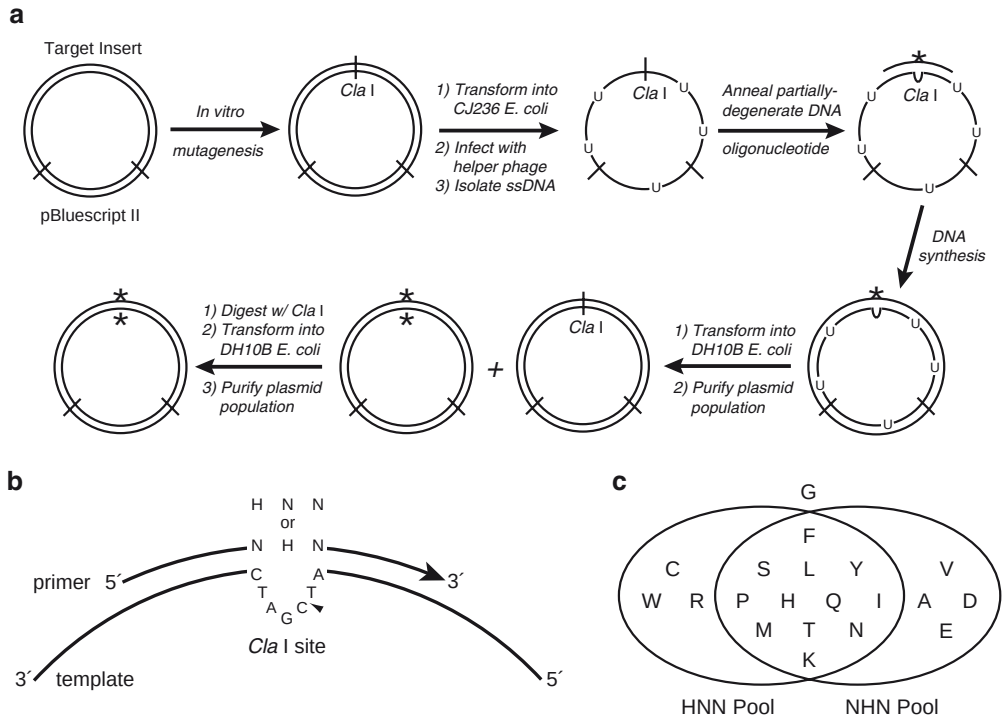


Fig. 1. Random-scanning mutagenesis. (**a**) Overview of the mutagenesis procedure. In this example, the second and third nucleotides of codon 152 of HIV-1 RT (GGA, glycine) were replaced with a short sequence encoding a unique *Cla* I restriction site (5′-ATCGAT-3′). The mutated construct was used as a template for producing single-stranded, uracil-containing DNA (ssDNA), which was then annealed to oligonucleotides encoding "NHN" or "HHN" at codon 152 (*, where H represents A, C, or T and N represents A, C, T, or G; each randomized position was synthesized using equimolar concentrations of nucleoside phosphoramidites). Following in vitro DNA synthesis and transformation into *E. coli*, the plasmids were isolated, digested with *Cla* I, and reintroduced into the *E. coli* host. The resultant plasmid libraries were greatly enriched for amino acid-altering mutations at the target codon site. (**b**) Closeup view of the mutagenic oligonucleotide primer annealed to the ssDNA template. The *small arrowhead* indicates the location of the *Cla* I cleavage site in the template strand. Note that in the initial pBluescript plasmid clone, the RT-encoding sequence was inserted in the opposite orientation relative to the f1 phage origin. As a result, the ssDNA template encoded the antisense strand of HIV-1 RT. (**c**) Venn diagram showing the amino acids encoded by each random oligonucleotide pool at position 155 of RT. The glycine residue found in wild-type RT was excluded from both populations.

second and third nucleotides of codon 152 (GGA, glycine) with a six-nucleotide sequence encoding a unique *Cla* I restriction site (Fig. 1a, b). The retention of a single guanosine (G) immediately 5′ of the *Cla* I site disrupted the *pol* reading frame, thereby ensuring that functional HIV-1 RT was not expressed from the plasmid sequence.

In the second phase of random-scanning mutagenesis, the modified plasmid is introduced into an *E. coli* host that is deficient in deoxyuridine triphosphatase (dUTPase) and uracil DNA glycosylase (UDG) activity (Fig. 1a). Plasmid-transformed cells are then infected with a defective helper phage that induces the synthesis of single-stranded, uracil-containing copies of the plasmid construct (ssDNA). The ssDNA is preferentially packaged

into phage capsids and released into the culture supernatant. Following phage precipitation and DNA extraction, the purified ssDNA template is hybridized to oligonucleotide primers that restore the reading frame of the target gene sequence while simultaneously introducing random nucleotide changes at the target codon position. For example, to randomize codon 152 of HIV-1 RT, we designed pools of oligonucleotides that introduced random nucleotides opposite the *Cla* I site in the template strand (Fig. 1b). We also designed pools of mutagenic oligonucleotides that restored the wild-type nucleotide sequence at codon 152 (GGA) and introduced random nucleotides at positions up to four codons 5′ or seven codons 3′ of the engineered restriction site (9). In this manner, we were able to randomize the sequence spanning codons 148 through 156 of HIV-1 RT using a single preparation of ssDNA template.

To further restrict the frequency of wild-type sequences in the plasmid libraries, we excluded specific bases from one or more nucleotide positions in the randomized portion of the mutagenic oligonucleotides. In the case of RT codon 152, we designed two separate oligonucleotide pools: one that excluded guanosine at the first nucleotide position and one that excluded guanosine at the second nucleotide position of the target codon (sequences "HNN" and "NHN," respectively; Fig. 1b). Together, the resultant HNN and NHN pools encoded all 19 possible amino acid replacements and excluded the wild-type glycine at codon 152 (Fig. 1c). Similar coverage was achieved at neighboring codons in HIV-1 RT using other mixtures of partially randomized oligonucleotide primers (9).

It is important to note that the approach described above generates libraries of plasmids that encode an uneven distribution of variant amino acids at the target site (i.e., some variants are present at relatively high frequencies, while others are relatively rare in the mutant populations). This variation is attributable to the intrinsic asymmetry of the genetic code as well as differences in the annealing efficiencies of individual DNA molecules in the random oligonucleotide pools. The bias toward particular codons can be partially overcome by adjusting the ratios of nucleoside phosphoramadites that are used to synthesize the randomized positions (13). Alternatively, oligonucleotides that encode low-frequency amino acids can be synthesized separately and either mixed into the random pools at defined ratios or used in separate mutagenesis reactions.

The remaining steps in the protocol are technically straightforward and are familiar to anyone who has used the Kunkel approach to site-directed mutagenesis (3). Briefly, the oligonucleotide-ssDNA hybrids are used as primer-templates for in vitro DNA synthesis, and the reaction products are introduced into UDG^+ *E. coli* to preferentially degrade the uracil-containing template strand (Fig. 1a). Next, the plasmids are isolated from the transformants,

treated with an enzyme that recognizes the engineered restriction site (i.e., *Cla* I in Fig. 1a, b) and retransformed into UDG^+ cells. This step provides an additional level of selection against the parental template sequence. Finally, the plasmids are purified en masse from the second set of UDG^+ transformants. The net result is a library of clones in which >90% of the plasmid copies encode an amino acid replacement at the targeted codon position.

3.1. Design of the Mutagenesis Template

The following points should be considered when designing the plasmid template for random-scanning mutagenesis:

1. *The choice of plasmid backbone*. Because the mutagenesis procedure requires ssDNA, the plasmid backbone must contain an f1 origin of replication. For our HIV-1 studies, we ligated a portion of the viral cDNA into pBluescript II KS(-). Other f1-containing vectors such as pGEM-3Zf± and pTZ18U should also, in principle, produce sufficient quantities of ssDNA following helper phage infection.
2. *The overall size of the ssDNA template*. The efficiency of oligonucleotide-directed mutagenesis depends in part on the length of the ssDNA template. In our studies, we ligated a 3.7 kilobase (Kb) insert into the 3.0 Kb pBluescript II KS(-) vector, for an overall plasmid length of 6.7 Kb. When we used this template and the protocol described below to perform conventional site-directed mutagenesis (i.e., to introduce one or two consecutive nucleotide changes into an otherwise wild-type gene segment), we typically achieved mutagenesis efficiencies of 60–90%. Although the use of larger plasmid constructs may diminish the mutagenesis efficiency, the subsequent digestion of the plasmid library with an enzyme that cleaves the DNA at the engineered restriction site (i.e., *Cla* I in Fig. 1a) will increase the frequency of mutants to >90%. DNA sequence analysis of at least 50 individual clones from the final plasmid library is strongly recommended to ensure that the resultant mutant frequency and diversity is sufficient for downstream applications.
3. *The placement of the engineered restriction site*. In situations where a single codon is randomized, the restriction site should be placed at the target codon as shown in Fig. 1b. If the mutagenesis target is a contiguous stretch of amino acid residues, the restriction site should be located in the center of the target region.

3.2. Design of the Partially Degenerate Oligonucleotide Primers

The mutagenic oligonucleotides must contain a terminal 5′-phosphate group for extension by T7 DNA polymerase; this can be accomplished via chemical synthesis or by treating the oligonucleotides with T4 polynucleotide kinase as previously described (12). The overall length of the oligonucleotide should be sufficient

to ensure that, when hybridized to the ssDNA template, the primers are anchored by 20–25 complementary base pairs at both the 5′and 3′ ends of the DNA duplex. All of the random oligonucleotide pools used in our analysis of HIV-1 RT were purified by high pressure liquid chromatography (9).

3.3. Production of Single-Stranded, Uracil-Containing DNA Template

1. Transform chemically competent CJ236 *E. coli* with the plasmid to be used as the mutagenesis template. CJ236 cells are dUTPase$^-$ UDG$^-$ and, as a result, uracil will be incorporated into the plasmid during DNA replication. Recover the transformants in 1 ml of SOC medium for 1 h at 37°C, and then spread several dilutions of cells onto LB-ampicillin–chloramphenicol plates. Invert the plates and incubate them at 37°C overnight.
2. Pick a well-isolated colony from the LB-ampicillin–chloramphenicol plates and inoculate the cells into 2 ml of 2× YT medium containing ampicillin (50 µg/ml) and chloramphenicol (15 µg/ml) (see Note 1). Incubate the culture for 12 h at 37°C with shaking (300 rpm).
3. Transfer the entire 2 ml starter culture to a 2 l flask containing 300 ml of 2× YT with ampicillin (50 µg/ml) and chloramphenicol (15 µg/ml). Immediately add VCSM13 helper phage at a final concentration of 2×10^8 pfu/ml (see Note 2).
4. Incubate the flask for 2 h at 37°C with shaking (300 rpm), then add kanamycin to a final concentration of 70 µg/ml and continue the incubation at 37°C, 300 rpm for 12 h (see Note 3).
5. Harvest the culture supernatant (which contains ssDNA packaged into phage particles) by centrifugation at 7,000 × *g* for 15 min at 4°C (see Note 4).
6. Taking care not to disturb the cell pellet, decant the supernatant to a new set of centrifuge tubes and repeat the centrifugation as described in step 5.
7. Using a pipette, transfer the upper 80% of the supernatants to a new set of centrifuge tubes and add ¼ volume of PEG/NaOAc solution to each tube. Gently mix the solution well and chill on ice for at least 1 h. This step will precipitate the phage particles. The tubes also can be stored at 4°C overnight if necessary.
8. Centrifuge the supernatants at 9,000 × *g* for 25 min at 4°C (see Note 5).
9. Decant the supernatants and resuspend the phage pellets in a total of 2 ml of TE pH 8. Aliquot 1 ml of each phage suspension into two separate 1.5-ml microcentrifuge tubes, then add 2 µl of 25 mg/ml DNAse-free RNAse A to each tube. Mix and incubate the suspensions at 37°C for 1 h.

10. Spin the tubes in a microcentrifuge at 15,000 g for 5 min at 4°C to pellet any remaining cells or cell debris and transfer the resultant supernatants to new 1.5 ml microcentrifuge tubes. Reprecipitate the phage by adding 250 µl of PEG/NaOAc solution to each tube and placing the tubes on ice for 15 min. Then, spin the tubes again at 14,000 rpm for 5 min at 4°C.
11. Drain the pellets well, pulse the tubes briefly in a microcentrifuge, remove any last traces of supernatant and resuspend the pellets in 300 µl TE pH 8.
12. Extract the ssDNA with phenol (equilibrated to pH 7.0), phenol/chloroform (1:1), and then with phenol/chloroform/isoamyl alcohol (25:24:1) (see Note 6). A step-by-step protocol for phenol–chloroform extraction of DNA is provided in ref. (12).
13. Precipitate the ssDNA in 1.5 ml microcentrifuge tubes by adding 2× volume of ice-cold 100% ethanol and 1/10 volume of 3 M NaOAc pH 5. Mix by inversion and incubate the precipitate on ice for 20–30 min. Spin the tubes at 14,000 rpm in a microcentrifuge for 15 min to pellet the ssDNA. Remove the supernatants carefully with a pipette, spin them again briefly and remove any remaining traces of ethanol from the tubes.
14. Allow the DNA pellets to air dry, and then resuspend them in a final volume of 500 µl TE pH 8. Determine the DNA concentration by spectrophotometry and load 5 µl into an agarose gel to visualize ssDNA yield and purity (see Note 7). Typical recovery is approximately 500 µg of ssDNA, which is sufficient for many mutagenesis reactions.

3.4. Mutagenesis Reactions

1. Place one 250 µl microcentrifuge tube on ice for each mutagenesis reaction and include an additional tube that will contain a negative control (no primer) reaction. To each tube, add 0.6 pmole of ssDNA template. Then, add 1 µl from a 6.6 pmole/µl stock of the random oligonucleotide pool(s) to the appropriate tube(s). Finally, add sterile distilled H_2O to bring the volume of each tube to 9 µl.
2. Add 1 µl of 10× annealing buffer to each tube and mix the contents by pipetting. Anneal the primers to the template DNA by heating a beaker containing 300 ml of dH_2O to 80°C, then add the reaction tubes to the bath and allow the beaker to cool to 30°C (~3 to 4 h). When cooling is complete, place the tubes on ice.
3. Dilute T7 DNA polymerase to a final concentration of 0.5 units/µl in 1× DNA polymerase dilution buffer. Mix by pipetting and place the diluted polymerase mixture on ice.

4. Add 1 µl of 10× synthesis buffer, 1 µl (6 Weiss units) of T4 DNA ligase, and 1 µl (0.5 units) of diluted T7 DNA polymerase to the annealing reactions. Mix the contents thoroughly by pipetting and place the reaction tubes in a 37°C water bath for 2 h. If necessary, store the completed reactions at –20°C.
5. Transform 1 µl of each mutagenesis reaction (including the no-primer control reaction) into chemically competent or electroporation-competent *E. coli*. The host strain must be dUTPase$^+$ UNG$^+$ to select against the uracil-containing template DNA. Place the transformed cells in sterile culture tubes containing 1 ml of SOC medium and incubate the cultures at 37 C for 1 h with shaking (225 rpm). Store the remainder of the mutagenesis reactions at –20°C.
6. Following the recovery period, remove 100 µl of cell suspension from each 2 ml culture and spread the sample of cells onto LB-ampicillin agar plates. Invert the plates and place them in a 37°C incubator overnight. In addition, transfer 200 µl of each recovery culture (not including the no-primer control) to a 1 l culture flask containing 100 ml of LB-ampicillin (50 µg/ml). Incubate the liquid cultures at 37°C for 12–16 h with shaking (250 rpm).
7. Harvest the cells from the 100 ml cultures by centrifugation at 6,000 × *g* for 15 min at 4°C. Decant the culture medium and store the cell pellets at –20°C.
8. Determine the number of colonies on each LB-ampicillin plate from step 6. At least threefold more colonies should be observed from the reactions that contained mutagenic oligonucleotide primers relative to the no-primer control reaction (see Note 8).
9. If favorable colony counts are observed, purify plasmid DNA from the *E. coli* pellets prepared in step 6 and store the resultant libraries at –20°C.

3.5. Enriching for Mutant Plasmids by Restriction Enzyme Digestion

At this stage in the protocol, the plasmid libraries consist of two subpopulations: plasmids in which the reading frame of the gene segment has been restored and mutations have been introduced at the target codon, and plasmids that retain the engineered restriction site (Fig. 1a). To increase the proportion of plasmids that contain mutations at the target codon, the DNA libraries are digested with an enzyme that recognizes the engineered restriction sequence (i.e., *Cla* I in Fig. 1a, b). An aliquot of each digest is then transformed into *E. coli*, resulting in the preferential uptake and replication of circular DNA molecules. Finally, the transformed cells are expanded en masse in liquid cultures as described in Subheading 3.4 step 6, and individual clones derived from the plasmid libraries are characterized by DNA sequencing. Once the level of diversity

in the libraries has been deemed sufficient, the plasmids can be purified from the liquid cultures and used in functional assays or in additional subcloning operations (see Note 9).

4. Notes

1. I have observed substantial variation in the yield and the purity of ssDNA obtained from different preparations of phage-infected cultures, including cultures that were infected on the same day with the same preparation of helper phage (see also Note 7). For this reason, I typically establish four cultures from four independent colonies of CJ236 transformants and maintain the culture products separately throughout the ssDNA isolation procedure.
2. Based on multiple trials, infection of the CJ236 transformants with MV1190 helper phage does not produce a sufficient yield and the purity of ssDNA for oligonucleotide-mediated mutagenesis.
3. Never grow the phage-infected cultures for more than 12 h, as this results in large amounts of contaminating *E. coli* RNA and genomic DNA in the final preparation of ssDNA template (see also Note 8).
4. For this step, I use a Sorvall JA-10 rotor and 500 ml polypropylene bottles.
5. A smear of phage particles ending in a large, translucent pellet at the bottom of the tube should be observed following centrifugation.
6. Follow appropriate safety precautions for handling liquid phenol and phenol-containing solutions, which can burn skin on contact.
7. Specifically, check for contaminating products such as *E. coli* chromosomal DNA (>>13 Kb) and RNA (diffuse bands <1 Kb). The predominant product (by at least 100-fold) should be a diffuse band corresponding to the ssDNA template. A second diffuse band corresponding to single-stranded helper phage DNA is normal and will not interfere with the mutagenesis procedure. Note that the migration of ssDNA on agarose gels is strongly influenced by secondary structure as well as the presence of ethidium bromide. As a result, the apparent size of the ssDNA template may differ from the expected result.
8. Failure to see a threefold or greater increase in colony counts may be indicative of spurious priming of DNA synthesis by *E. coli* DNA or RNA. These contaminating products are the

result of lysis of the CJ236 host in culture. Contamination by bacterial nucleic acids is minimized by using freshly transformed CJ236 cells and by carefully following the incubation times provided in Subheading 3.3.

9. During subcloning operations, it is important to avoid "bottlenecks" that diminish the diversity of the random plasmid libraries. If a segment that spans the target site is to be cloned into a larger plasmid, then following ligation and transformation of *E. coli*, an aliquot of the recovery culture should be used to directly inoculate a large-scale culture as in Subheading 3.4 step 6. A second aliquot should also be removed from the recovery culture and spread to LB-ampicillin plates to assess the number of colony-forming units at the time of sampling. In this way, the number of independent "founder" clones that were used to establish each of the final random plasmid libraries can be determined.

References

1. Botstein D, Shortle D (1985) Strategies and applications of in vitro mutagenesis. Science 229:1193–1201
2. Hutchison CA III, Phillips S, Edgell MH, Gillam S, Jahnke P, Smith M (1978) Mutagenesis at a specific position in a DNA sequence. J Biol Chem 253:6551–6560
3. Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci U S A 82:488–492
4. Higuchi R, Krummel B, Saiki RK (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res 16:7351–7367
5. Sarkar G, Sommer SS (1990) The "megaprimer" method of site-directed mutagenesis. Biotechniques 8:404–407
6. Cunningham BC, Wells JA (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science 244:1081–1085
7. Schulman BA, Kim PS (1996) Proline scanning mutagenesis of a molten globule reveals non-cooperative formation of a protein's overall topology. Nat Struct Biol 3:682–687
8. Sharp LL, Zhou J, Blair DF (1995) Features of MotA proton channel structure revealed by tryptophan-scanning mutagenesis. Proc Natl Acad Sci U S A 92: 7946–7950
9. Smith RA, Anderson DJ, Preston BD (2004) Purifying selection masks the mutational flexibility of HIV-1 reverse transcriptase. J Biol Chem 279:26726–26734
10. Smith RA, Anderson DJ, Preston BD (2006) Hypersusceptibility to substrate analogs conferred by mutations in HIV-1 reverse transcriptase. J Virol 80:7169–7178
11. Burks EA, Chen G, Georgiou G, Iverson BL (1997) In vitro scanning saturation mutagenesis of an antibody binding pocket. Proc Natl Acad Sci U S A 94:412–417
12. Fritsch EF, Sambrook J, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
13. Airaksinen A, Hovi T (1998) Modified base compositions at degenerate positions of a mutagenic oligonucleotide enhance randomness in site-saturation mutagenesis. Nucleic Acids Res 26:576–581

Chapter 28

Easy Two-Step Method for Randomizing and Cloning Gene Fragments

Vivian Q. Zhang and Holly H. Hogrefe

Abstract

Random mutagenesis is widely used in protein engineering to improve or alter protein function. Creating random mutant libraries typically requires cloning of randomly mutagenized fragments into an expression vector, which is laborious and often hampered by lack of unique and convenient restriction sites. Here, we report an easy two-step method that produces a more balanced mutational spectrum and simplifies the cloning of randomly mutagenized genes or gene fragments for constructing high titer random mutant libraries.

Key words: Random mutagenesis, Domain mutagenesis, Balanced mutational spectrum, Megaprimer, EZClone, GeneMorph

1. Introduction

Random mutagenesis is a powerful tool for studying protein structure–function relationships and for improving or altering protein function. Error-prone PCR, the most commonly used random mutagenesis technique, randomly introduces mutations during PCR cycling with low fidelity thermostable DNA polymerases (1). The mutated PCR products are then cloned into an expression vector, and the resulting mutant library can be screened for changes in function. Random mutagenesis allows researchers to identify beneficial mutations in the absence of structural information, or when such mutations are difficult to predict from protein structure.

There are several factors that need to be considered when generating a random mutant library. First, to ensure representation, researchers should choose conditions that produce an unbiased spectrum of mutations. That is, all four bases A, C, G, and T should have equal opportunity to replace and be replaced. Second, researchers

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_28,

should use conditions that provide the most appropriate mutation frequencies for a particular application. For example, a mutation rate of one amino acid change per gene is desired to assess the contribution of an individual amino acid to protein function (2), while mutation rates of 2–6 amino acid changes per gene are considered the most effective for directed evolution studies (3–6). Third, researchers should consider the method used to clone randomly mutagenized fragments, as fragments produced by error-prone PCR can be difficult to clone due to low yield, incomplete extension of 3′ ends (resulting from low mispair extension efficiency), and/or the introduction of mutations into restriction sites designed for cloning. Improved cloning methods should not only speed up library construction, but also allow researchers to create higher titer libraries.

Creating random mutant libraries basically involves two steps: random mutant strand synthesis and cloning of mutated products into an expression vector. Error-prone PCR with *Taq* DNA polymerase is widely used in the first step. But there are several drawbacks with *Taq*. The mutational spectrum of *Taq* exhibits high bias, strongly favoring mutations of A and T over G and C (2). To achieve useful mutation rates, error-prone PCR using *Taq* are conducted with unbalanced dNTPs and Mn^{2+}, which generates low product yields and directly impacts library size (7). The conventional method for cloning randomized PCR products is to digest both vector and insert with unique restriction enzymes, followed by in vitro ligation with DNA ligase. Although restriction cloning can be applied to a whole gene, it is hardly applicable in situations where one is interested in randomizing a portion of a gene, such as a region encoding a functional protein domain. In most cases, it is difficult to introduce unique restriction sites at the desired location (e.g., at the 5′ and 3′ ends of a mutated domain) without altering amino acid sequence, which can have deleterious effects on protein function or expression. In addition, restriction-based cloning requires careful attention to vector/insert ratios to minimize the percentage of clones with no insert or multiple inserts.

The GeneMorph II EZClone Domain Mutagenesis Kit (Fig. 1) uses Mutazyme II DNA polymerase, a novel blend of a *Taq* mutant and Mutazyme I (8), in the first mutant strand synthesis step. The *Taq* mutant is more error-prone than wild-type *Taq*, and useful error rates can be achieved under optimal cycling condition (no need for unbalanced dNTPs and Mn^{2+}). Using a combination of DNA polymerases produces a more uniform mutational spectrum, and product yields are much higher than with conventional (see Note 7) *Taq*-based methods. Additionally, low, medium, or high mutation rates can be achieved simply by varying input DNA template amount and/or PCR cycle number. The PCR products synthesized in the first step serve as megaprimers in the second EZCloning step (based on the Quik Change method). This method eliminates the traditional cloning step, does not require any restriction sites, and is applicable to any portion of a gene

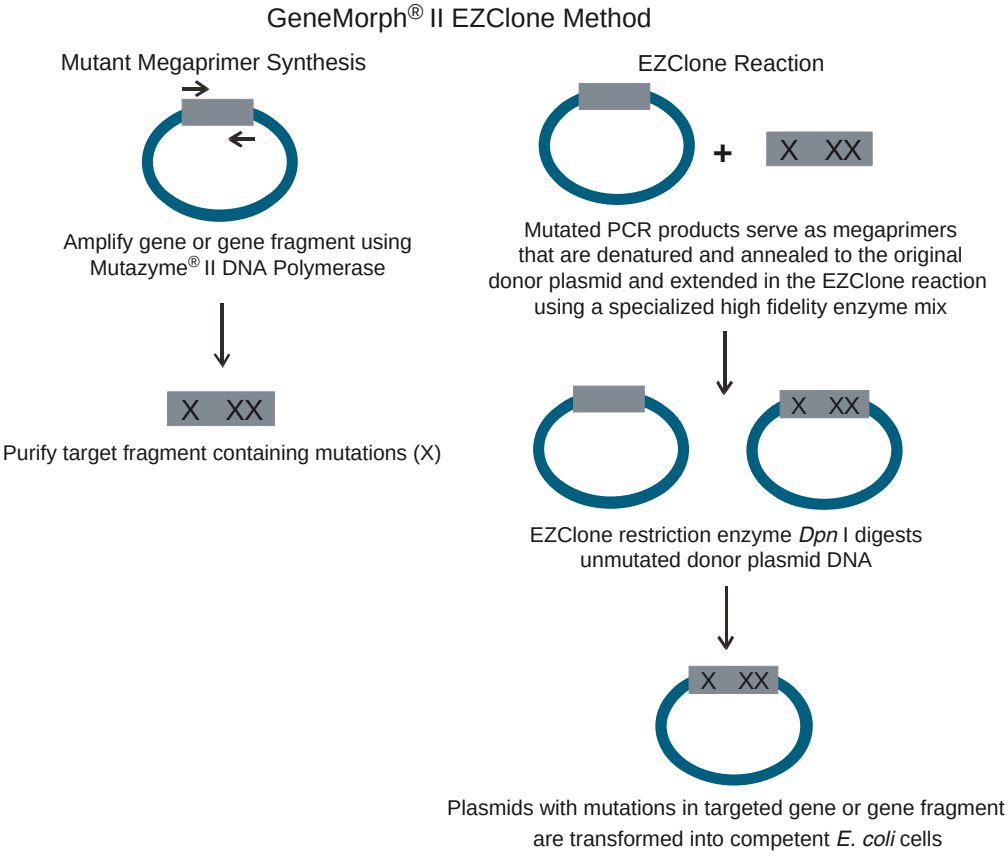


Fig. 1. GeneMorph II EZClone domain mutagenesis kit method. Reprinted by permission of Agilent Technologies, Stratagene Products Division.

(see Note 1–3). Moreover, random fragments can be cloned accurately and efficiently, with minimal mutations in the vector backbone and virtually all clones contain a single insert. Libraries prepared by the method exhibit sufficiently high titers for screening.

2. Materials

2.1. The GeneMorph II EZClone Domain Mutagenesis Kit (Agilent Technologies, Stratagene Products Division, La Jolla, CA 92037)

1. 10× Mutazyme II reaction buffer.
2. 40 mM dNTP mix.
3. Mutazyme II DNA polymerase (Agilent Technologies, Stratagene Products Division, La Jolla, CA 92037).
4. 1.1 kb DNA gel standard.
5. 2× PfuUltra Hotstart master mix (Agilent Technologies, Stratagene Products Division, La Jolla, CA 92037).
6. *Dpn*I restriction enzyme (Agilent Technologies, Stratagene Products Division, La Jolla, CA 92037).

7. EZClone solution (Agilent Technologies, Stratagene Products Division, La Jolla, CA 92037).
8. Positive control plasmid (what is it?) (Agilent Technologies, Stratagene Products Division, La Jolla, CA 92037) (see Note 8).
9. Positive control primer mix (what is it?) (Agilent Technologies, Stratagene Products Division, La Jolla, CA 92037) (see Note 8).
10. XL10-Gold ultracompetent cells (Agilent Technologies, Stratagene Products Division, La Jolla, CA 92037).
11. XL10-Gold β-mercaptoethanol mix (Agilent Technologies, Stratagene Products Division, La Jolla, CA 92037).
12. pUC18 control plasmid.

2.2. Purification and Quantification of PCR Products

1. StrataPrep PCR purification kit (Agilent Technologies, Stratagene Products Division, La Jolla, CA 92037).
2. StrataPrep DNA gel extraction kit (Agilent Technologies, Stratagene Products Division, La Jolla, CA 92037).
3. Agarose gel.

2.3. Transformation

1. 14 ml Falcon polypropylene round-bottom tube (BD Biosciences, San Jose, CA).
2. NZY+ broth (1 l): 10 g NZ amine, 5 g yeast extract, 5 g NaCl, 12.5 ml of 1 M $MgCl_2$, 12.5 ml of 1 M $MgSO_4$, and 20 ml of 20% (w/v) glucose.
3. Luria–Bertani (LB) media: 10 g/L NaCl, 10 g/L tryptone, and 5 g/L yeast extract.
4. LB agar (2%).

3. Methods

3.1. Mutant Megaprimer Synthesis by Error-Prone PCR

1. Error-prone PCR setup: combine 5 µl of 10× Mutazyme II buffer, 50 ng of the DNA template, 125 ng of each PCR primer, 1 µl of 40 mM dNTP, 1 µl of Mutazyme II DNA polymerase, and deionized H_2O to a final volume of 50 µl (see Note 6).
2. Thermal cycling: 1 cycle of 95°C for 2 min, 30 cycles of 95°C for 30 s, primer Tm – 5°C for 30 s, 72°C for 1 min (<1 kb target) or 1 min/kb (>1 kb target), 1 cycle of 72°C for 10 min (see Note 6).

3.2. Quantification of the PCR Product

1. Perform electrophoresis of 1–10 µl of each amplification reaction along with 2.5 µl (50 ng) of the 1.1 kb DNA gel standard on a 1% agarose gel.

2. Estimate the PCR product yield by comparing the intensities of the PCR product bands with the 1.1 kb gel standard by visual inspection or by using an imaging system. The total yield is used for calculating mutation frequency (see Note 6).

3.3. Purification of the PCR Product

1. Use the PCR purification kit when the amount of the initial DNA template is less than 50 ng per 50 µl reaction, and the same plasmid DNA is used in the following EZClone reaction (see Subheading 3.5).
2. Use the DNA gel extraction kit: when the amount of the initial DNA template is more than 50 ng in 50 µl reaction or a different plasmid DNA will be used in the following EZClone reaction (see Subheading 3.5).

3.4. Quantitation of the Purified PCR Product

1. Perform electrophoresis of 1–10 µl of each purified amplification reaction along with 2.5 µl (50 ng) of the 1.1 kb gel standard on a 1% agarose gel.
2. Estimate the yield of purified PCR product by comparing the intensities of the PCR product bands with the 1.1 kb gel standard by visual inspection or by using an imaging system. Amplicon concentration is used in calculating the amount of megaprimer to add to an EZClone reaction.

3.5. EZClone Reaction

1. EZClone reaction setup: combine 25 µl of 2× PfuUltra Hotstart master mix, 50 ng of the plasmid template, 250 ng megaprimer (see Note 4–5), 3 µl of EZClone solution, and deionized H_2O to a final volume of 50 µl.
2. Thermal cycling: 1 cycle of 95°C for 1 min, 25 cycles of 95°C for 50 s, 60°C for 50 s, 68°C for 2 min/kb of plasmid length.

3.6. DpnI Digestion

Add 1 µl of *Dpn*I restriction enzyme (10 U/µl) directly into each EZClone reaction. Gently and thoroughly mix and spin down in a microcentrifuge for 10 s. Incubate the reactions at 37°C for 2 h to digest the parental (i.e., the unmutated) supercoiled dsDNA.

3.7. Transformation of XL10-Gold Ultracompetent Cells

1. 1.5 µl of *Dpn*I-digested DNA is transformed into 45 µl XL10-Gold chemical ultracompetent cells.
2. After incubation on ice for 10 min, heat-pulse the tubes (14 ml BD Falcon polypropylene round-bottom tube) in a 42°C water bath for 30 s. Incubate the tubes on ice for 2 min.
3. Add 0.5 ml of preheated (42°C) NZY+ broth to each tube, and shake at 37°C for 1 h.
4. Plate out the appropriate volume of each transformation reaction on an LB plate with appropriate antibiotics. Incubate the plates at 37°C over night up to 24 h.

3.8. Library Titer Calculation and Mutant Screening

1. Use the following formula to calculate the mutant library titer:

$$\text{Mutant library titer (colony forming units (cfu) per 50 µl EZClone reaction)} = \frac{((\text{Colony \# on plate}) \times 50 \times 50)}{(\text{Plating volume (ml)} \times 1.5)}.$$

2. Mutant screening can be performed by assaying gain or loss of the function, or more generally by sequencing (see Note 9).

4. Notes

1. Targets should be amplified from high-copy-number templates such as plasmid DNA (circular or linear) or PCR product. Direct amplification from low-copy-number templates (e.g., genomic DNA) should be avoided to keep mutation rates within a useful range (see Table 1). If genomic DNA is the only source, the target should be amplified with a high fidelity DNA polymerase to generate the recommended amount of PCR amplicon for subsequent amplification with Mutazyme II.
2. The plasmid DNA template used in the EZClone reaction must contain the sequence that is targeted in the error-prone PCR and must be isolated from a dam+ *E. coli* strain to allow *Dpn*I digestion (enrichment of mutant DNA). Most commonly used *E. coli* strains that are dam+ include XL1-blue, XL10-gold, and DH5α, etc. The plasmid DNA templates for

Table 1
Mutation frequency vs. initial target quantity and cycle number

Mutation rate	Mutation frequency (mutations/kb)[a]	Initial target amount (ng)[b]	Recommended fold amplification	Recommended cycle number
Low	0–4.5	500–1,000	1.5–10	20–30
Medium	4.5–9	100–500	10–100	25–30
High	9–16	0.1–100	100–10,000	30 or 2–3 rounds of PCR

Reprinted by permission of Agilent Technologies, Stratagene Products Division

[a]These values are accurate for reactions achieving the approximate fold amplification (total yield/input DNA) indicated. The actual number of mutations in each clone may differ as these values represent the average frequency for all clones

[b]The amount of template indicated is the amount of target DNA to be amplified, not the total amount of DNA template to add to the reaction. For example, to mutagenize a 1-kb target in a total 4-kb plasmid at a low mutation frequency, 2 µg of whole plasmid (equal to 500 ng of 1 kb target) is needed in the reaction

error-prone PCR and the EZClone reaction may or may not be identical. If different, a gel purification of megaprimer is required.

3. Amplicons (megaprimer) ranging in size from 0.15 to 3.5 kb have been used successfully with the GeneMorph II EZClone kit.
4. If the yield of megaprimer is low, increase the cycle number and/or perform replicate 50 µl reactions. Pool the reactions after PCR and go through the remaining steps.
5. When using long (>1 kb) megaprimers, the titer of the mutant library can be increased by adding up to 1 µg of the megaprimer to the EZCloning reaction.
6. The mutation frequency of an amplification reaction is determined by the formula:

$$\text{Mutation frequency} = \text{Error rate} \times d$$

where mutation frequency is expressed as mutations per kb, *error rate* is the intrinsic error rate of the DNA polymerase in terms of errors per kb per duplication, and *d* is the number of target duplications during PCR.

The variable *d* can be calculated from the following equation:

$$2^d = \frac{\text{PCR yield}}{\text{Initial amount of target}}$$

$$\text{or } d = \ln\left(\frac{\text{PCR yield}}{\text{Initial amount of target}}\right)$$

As can be seen in the formulas above, mutation frequency can be adjusted by changing input target amount and/or PCR cycle number. Table 1 provides recommendations for the amount of target and number of cycles to use to achieve low, medium, or high mutation rates.

7. In the absence of bias, all four bases A, C, G, and T have equal opportunity to replace and be replaced. Unfortunately, all DNA polymerases produce a biased spectrum of mutations that reflects an intrinsic tendency to create certain mutation types over others. There are several ways to assess bias in mutational spectra. Bias can be examined by analyzing the ratio of transition (Ts) to transversion (Tv) mutations. Transition mutations are purine (A and G) to purine changes and pyrimidine (C and T) to pyrimidine changes, while transversions are purine to pyrimidine and pyrimidine to purine changes. There are eight possible transversions and four possible transitions, and a PCR enzyme completely lacking bias would exhibit a Ts/Tv ratio of 0.5. Mutational bias can also be

assessed by calculating the ratio of AT→GC to GC→AT transition mutations (AT→GC/GC→AT ratio), which would equal 1 for a completely unbiased enzyme. Finally, mutational bias can be assessed by comparing the frequency of mutating A's and T's vs. the frequency of mutating G's and C's (AT→NN/GC→NN ratio), which should be equal for an unbiased DNA polymerase. Table 2 shows the mutational spectra of three commonly used error-prone PCR enzymes. *Taq* strongly favors mutations at A and T over G and C (2), while Mutazyme I has a tendency to introduce errors at G and C rather than A and T. Mutazyme II DNA polymerase is

Table 2
Mutational spectra of error-prone enzymes

Type(s) of mutations	Mutazyme II[a]	Mutazyme I[a]	*Taq*[b]
Bias indicators			
Ts/Tv	0.9	1.2	0.8
AT→GC/GC→AT	0.6	0.2	1.9
A→N, T→N	50.7%	25.6%	75.9%
G→N, C→N	43.8%	72.5%	19.6%
Transitions			
A→G, T→C	17.5%	10.3%	27.6%
G→A, C→T	25.5%	43.7%	13.6%
Transversions			
A→T, T→A	28.5%	11.1%	40.9%
A→C, T→G	4.7%	4.2%	7.3%
G→C, C→G	4.1%	8.8%	1.4%
G→T, C→A	14.1%	20%	4.5%
Insertions and deletions			
Insertions	0.7%	0.8%	0.3%
Deletions	4.8%	1.1%	4.2%
Mutation frequency			
Mutations/kb (per PCR)[c]	3–16 (per PCR)	<1 to 7 (per PCR)	4.9 (per PCR)

Reprinted by permission of Agilent Technologies, Stratagene Products Division
[a]Mutazyme I and II DNA polymerases were used with the corresponding GeneMorph Random Mutagenesis Kits
[b]*Taq* DNA polymerase was used with Mn^{2+}-containing buffer and unbalanced dNTP concentrations, which are mutagenic conditions for *Taq* DNA polymerase
[c]Initial target amounts of 16 pg to 1 μg (Mutazyme II DNA polymerase), 1 pg to 100 ng (Mutazyme I DNA polymerase), and 0.01 nM template (*Taq* DNA polymerase) were used to generate the data

a novel error-prone PCR enzyme blend of Mutazyme I and a *Taq* mutant. The *Taq* mutant produces the same mutational spectrum as wild-type *Taq* while introducing more errors under optimal reaction conditions. As shown in Table 2, the Mutazyme II enzyme blend provides a more uniform mutational spectrum than can be achieved using either enzyme separately.

8. The 3 kb control plasmid included with the GeneMorph II EZClone domain mutagenesis kit encodes the *lac*Z gene. Colonies produce blue color when plated on LB-ampicillin agar plates containing IPTG and X-gal. When 1 ng of control plasmid DNA is used to carry out 30 cycles of error-prone PCR (high mutation frequency) and 250 ng of megaprimer is incorporated in the EZClone step, greater than 25% of the resulting colonies appear white (mutant) as compared to less than 1% white colonies for the nonmutagenized control plasmid. White colonies indicate that random mutations in the *lac*Z gene have eliminated β-galactosidase activity (loss of function). There are some mutations that do not impair enzyme activity and therefore, the frequency of phenotypic changes (blue to white color change) is lower but proportional to the rate of genotypic changes.
9. The GeneMorph II EZClone domain mutagenesis kit was successfully applied to randomize Moloney murine leukemia virus (MMLV) reverse transcriptase (RT) gene by Arezi and Hogrefe (9). A new more thermostable RT was achieved by the combination of several mutations.

References

1. Innis MA, Gelfand DH, Sninsky JJ, White TJ (1990) PCR protocols: a guide to methods and applications. Academic, New York
2. Vartanian JP, Henry M, Wain-Hobson S (1996) Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. Nucleic Acids Res 24(14):2627–2631
3. You L, Arnold FH (1996) Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. Protein Eng 9(1):77–83
4. Shafikhani S, Siegel RA, Ferrari E, Schellenberger V (1997) Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. Biotechniques 23(2):304–310
5. Wan L, Twitchett MB, Elits LD, Mauk AG, Smith M (1998) In vitro evolution of horse heart myoglobin to increase peroxidase activity. Proc Natl Acad Sci U S A 95(22): 12825–12831
6. Cherry JR, Lamsa MH, Schneider P, Vind J, Svendsen A (1999) Directed evolution of a fungal peroxidase. Nat Biotechnol 17(4): 379–384
7. Cadwell RC, Joyce GF (1992) Randomization of genes by PCR mutagenesis. PCR Methods Appl 2(1):28–33
8. Cline J, Hogrefe H (2000) Randomize gene sequences with new PCR mutagenesis kit. Strategies 13(4):157–161
9. Arezi B, Hogrefe H (2009) Novel mutations in Moloney murine leukemia virus reverse transcriptase increase thermostability through tighter binding to template-primer. Nucleic Acids Res 37(2):473–481

Part VII

Mutator Bacterial Strain Mutagenesis

Chapter 29

Random Mutagenesis Using a Mutator Strain

Ghazala Muteeb and Ranjan Sen

Abstract

Random mutagenesis is a useful technique to study the functions of different gene products. Propagation of the genes cloned in plasmids through a mutator strain, like *Escherichia coli* XL1-red, produces randomly mutagenized plasmid libraries. This method offers a very simple and economic way of introducing random point mutations throughout the gene with a fairly high mutation rate. The whole process involves transformation and propagation of a plasmid containing the desired gene into the XL-1 red strain, isolating the mutagenized plasmid library and transforming this library into a desired strain for screening the mutant phenotype.

Key words: XL1-red™, Random mutation, Mutagenized library, *E. coli*, Spontaneous mutation, DNA repair

1. Introduction

The changes in the nucleotide sequences of the genetic material of an organism are defined as mutations. The changes in the nucleotide sequence can be in the form of substitution of a single nucleotide for another or insertion/deletion of more than one nucleotide sequence. Substitution mismatch or point mutation can be a transition, that exchanges a purine for a purine (example, A for G) or the less common transversion, which exchanges a purine for a pyrimidine or vice versa (example C/T for A/G). These mismatches can arise due to the errors in DNA replication, during recombination or due to DNA damage caused by UV radiation, reactive oxygen species, and alkylating agents (1). These mismatches are rectified by the machineries of the mismatch repair pathway (1).

Although DNA replication is a very accurate process due to the high nucleotide selectivity and proofreading

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_29, © Springer Science+Business Media, LLC 2010

ability (3′→5′ exonuclease activity) of DNA polymerases, the rare postreplication DNA mismatches that occur are repaired by the mutHSL mismatch repair system (2). In *Escherichia coli*, presence of the mismatch repair system improves the fidelity of DNA replication by almost 1,000-fold ((2, 3); Fig. 1). The mismatch repair system involves three proteins:

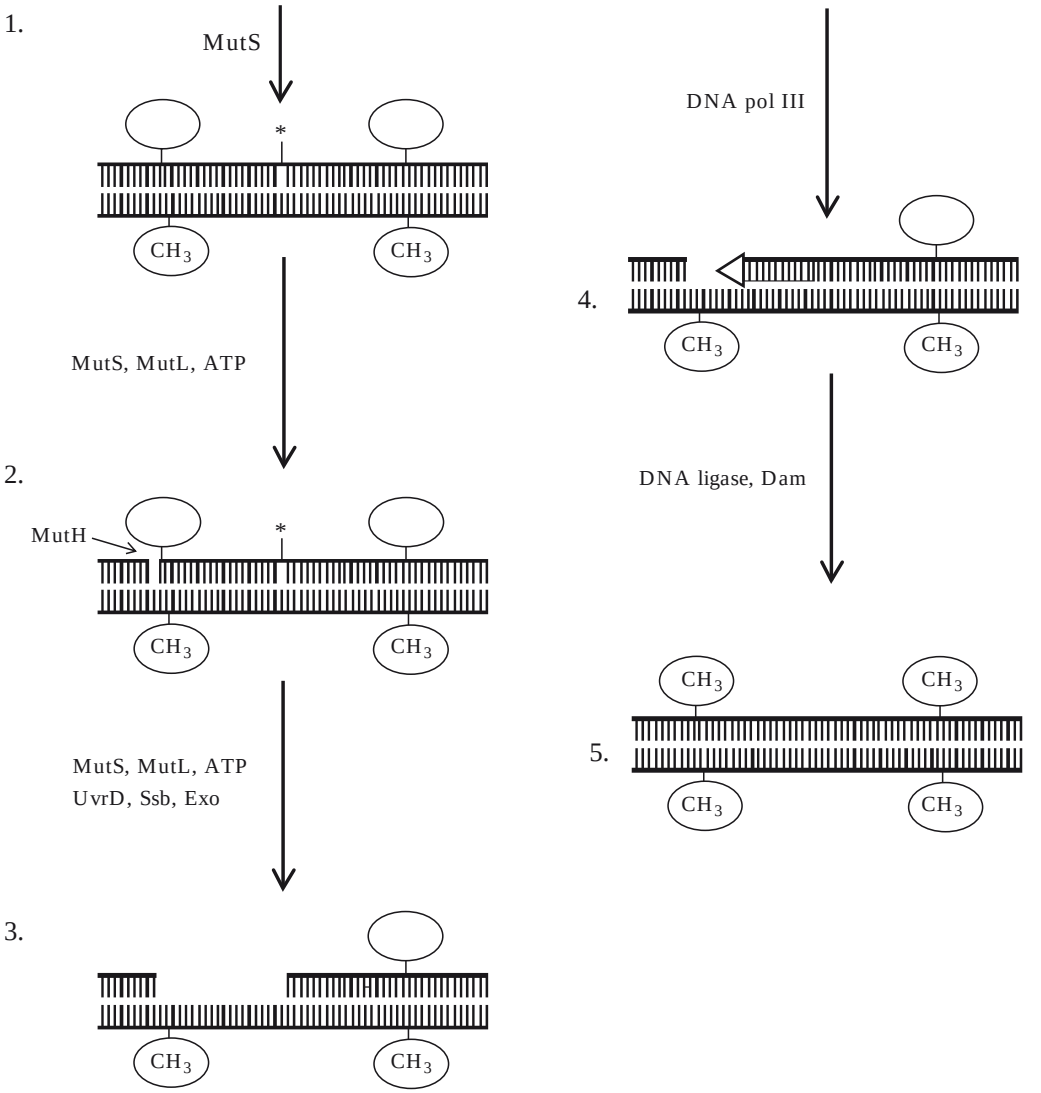


Fig. 1. *Escherichia coli* mismatch repair pathway. (1) Newly synthesized DNA is temporarily unmethylated (*empty circles*). The mismatch (*) is recognized by MutS. (2) MutS-MutL–ATP complex activates MutH to cleave the newly synthesized strand. (3) UvrD helicase unwinds DNA, exonuclease generates a gap that extends from the nick to approximately 100 nucleotides past the mispair, which is stabilized by single stranded DNA binding protein. (4) Gap is filled by DNA polymerase III. (5) DNA ligase seals the nick, and deoxyadenosine methylase methylates the newly synthesized strand.

MutS, MutL, and MutH. MutS is an ATPase that recognizes the mismatch. MutL, a member of the GHKL superfamily of ATPases (3), is recruited by MutS in the presence of ATP. These two proteins together activate MutH that cleaves only the newly synthesized, unmethylated DNA strand. The methyl-directed nature of the *E. coli* MMR (methyl-directed mismatch repair) system is an efficient way to discriminate template and daughter strands during DNA replication. Following the cleavage, UvrD (MutU) helicase unwinds the duplex DNA molecule (4). Excision of the newly synthesized strand between the nick and the mismatch is carried out by single-strand exonucleases (3). Once the mismatch has been removed, DNA polymerase III fills the gap and the DNA ligase seals the remaining nick. Mismatches in DNA can also arise due to the damage caused by reactive oxygen species. In *E. coli*, MutT (a base excision repair enzyme), along with MutM and MutY glycosylases, repairs this type of mismatch (5). *E. coli* strains defective in repair pathway genes have elevated rates of spontaneous mutations ((6); Fig. 1).

Introducing random mutations into a gene and subsequent phenotype screening is an important technique to understand gene function. Random mutations can be introduced into the gene of interest by error-prone PCR, chemical mutagenesis (using MNNG: *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine, EMS: ethyl methanesulfonate), or by introducing the gene of interest into a mutator strain, which is defective in DNA repair. *E. coli* XL1-red (*endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac mutD5 mutS mutT Tn10* (*Tet*^{r})) engineered by Stratagene is deficient in functional DNA repair pathway genes, like *mutS*, *mutD*, and *mutT*. The strain is severely defective in error-prone mismatch repair (due to mutation in *mutS*) (7), in 3′- to 5′-exonuclease activity of DNA polymerase III (due to mutation in *mutD*) (8) and in repairing oxidative DNA damages (due to mutation in *mutT*) (9). Because of these defects, this strain functions as a mutator strain and is capable of introducing random mutations with a very high rate. In the wild type *E. coli*, the rate of spontaneous mutation per genome per replication is 0.0025 (10), whereas in XL1-red, it is about 5,000-fold higher. The advantage of using XL1-red lies in the fact that it can introduce random mutations in to the gene of interest quite easily and quickly. The gene of interest is cloned in a plasmid. The XL1-red strains are transformed with the plasmid and the random mutations are introduced during each round of DNA replication. The mutation frequency can be simply increased by extending the culture period. More importantly, the technique does not involve any hazardous mutagens or carcinogens required for performing mutagenesis.

2. Materials

2.1. XL1-Red Competent Cells

Frozen XL1-red chemically competent cells are available from Stratagene (La Jolla, CA), as 200 µl vials (see Note 1). Cells stored properly at −80°C should retain their efficiency for 6 months. From each 200 µl vial, four transformations can be performed. When aliquoting, the competent cells should always remain on ice and should be transferred into the prechilled tubes.

2.2. Plasmid DNA

In this random mutagenesis procedure, the gene of interest in which mutations are desired is cloned into plasmid DNA and transformed into XL1-red competent cells. The plasmid was prepared from an overnight *E. coli* culture using a Qiagen Plasmid Miniprep Kit (Valencia, CA). Approximately 40 ng of plasmid DNA was used for transformation.

2.3. β-Mercaptoethanol

β-Mercaptoethanol (Stratagene) is used to increase the efficiency of transformation two-to threefold. The final concentration was 25 mM in each transformation mixture.

2.4. SOC Medium

1. Take 20 µl of filter sterilized 20% (w/v) glucose and add 980 µl SOB medium (autoclaved) to make the final volume 1 ml. The resulting SOC medium should be prepared immediately before use.
2. SOB medium: 20.0 g tryptone (Difco, Lawrence, KS), 5.0 g yeast extract (Difco), 0.5 g of NaCl (USB, Cleveland, OH), and add deionized water to make the final volume 1 l. Autoclave and add 10 ml of filter sterilized 1 M $MgCl_2$ (Sigma, St. Louis, MO) and 10 ml of filter sterilized 1 M $MgSO_4$ (Sigma) prior to use.

2.5. LB Broth, LB Agar Plates, and Agarose

1. LB broth: 10 g NaCl, 10 g tryptone, 5 g yeast extract, add deionized water to a final volume of 1 l. Adjust pH to 7.0 with 5 N NaOH (Sigma). Autoclave at 121°C for 15 min.
2. LB Agar: 10 g NaCl, 10 g tryptone, 5 g yeast extract, and 20 g bacto-agar (Difco). Add deionized water to a final volume of 1 l. Adjust pH to 7.0 with 5 N NaOH and autoclave. Cool to room temperature, and add filter sterilized antibiotic according to the final concentration required (for Ampicillin (USB): 100 µg/ml). Pour the media in petri dishes (approximately 25 ml/100-mm plate).
3. Agarose: Electrophoresis grade agarose from Sigma is used to cast gels in order to visualize the isolated plasmid.

2.6. TAE Buffer

TAE: 40 mM Tris–acetate (Tris base from USB, glacial acetic acid from Qualigens, India), 2 mM EDTA (USB) (pH 8.0) was used

for running agarose gel. TAE is prepared as 50× concentrated stock solution using deionized (18.2 mΩ) water and used as 1× concentration.

2.7. Electro Competent Cells Preparation

1. Sterile deionized water: Collect 1 l of fresh deionized water (18.2 mΩ), and autoclave at 121°C for 30 min.
2. Sterile 10% glycerol: Make 10% glycerol from 100% glycerol (Sigma) stock using deionized water and autoclave at 121°C for 30 min.

2.8. RO Water

Reverse osmosis water (approximately 10 mΩ).

2.9. DNA Marker

0.5 mg/ml of Lambda *Hin*dIII marker from Fermentas (Hanover, MD) is used for quantification of plasmid preparations.

2.10. 6× Loading Dye (Fermentas)

Composition is 10 mM Tris–HCl (pH 7.6), 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol, and 60 mM EDTA.

2.11. Oakridge Centrifuge Tube

Autoclavable 50 ml capacity from Tarsons (Kolkata, India).

2.12. Bio-Rad Gene Pulser

Electroporator, Gene pulsar Xcell microbial system from Bio-Rad (Hercules, CA) consists of a pulse controller module and a shock pod suitable for electroporation of bacteria and yeast.

2.13. Electroporation Cuvette

Cuvettes used for the electroporation having a gap width of 1 mm.

3. Methods

3.1. XL1-Red Transformation

1. Thaw a vial of XL1-red competent cells on ice.
2. For each transformation, aliquot 50 µl of the XL1-red competent cells into a prechilled tube. The 200 µl of competent cells, once thawed, should be used immediately for four separate transformations.
3. Add β-mercaptoethanol to each aliquot of competent cells, to the final concentration of 25 mM.
4. Incubate the tube on ice for 10 min.
5. Add approximately 40 ng of plasmid DNA to each aliquot of competent cells and mix gently with a pipette.
6. Incubate on ice for 30 min.
7. Give a heat shock to the tube by quickly transferring the tube to a 45°C water bath for 45 s. The time period of the heat pulse is very critical as the efficiency decreases sharply if the heat shock is given for less than 45 s or greater than 60 s.

8. Transfer the tube immediately to ice after the heat shock. Incubate for 2 min.
9. Add 1 ml of SOC medium (pre-warmed to 42°C) to each tube, and incubate at 37°C for 1 h with shaking at 250 rpm.

3.2. Plating the Transformation Mixture

1. Spin the tube at 1,700 × *g* for 4 min. Immediately discard approximately 800 µl of supernatant with a pipette, and resuspend the cells in 200 µl of SOC medium. This is done to concentrate the cells.
2. Put the entire 200 µl of cell suspension on an LB-agar plate containing the appropriate antibiotic (see Note 2). Spread the mix on the plate by using a sterile spreader.
3. Incubate the plates at 37°C for 24–30 h. As the mutator strain grows slowly in LB media (rich media), so the colonies are visible only after 24–30 h (see Note 3). The incubation can be extended longer for larger sized colonies. Different colony sizes are observed due to high mutation rate of the strain.

3.3. Preparation of the Mutagenized Library

1. Scrape all the colonies from the transformation plate using a sterile loop and inoculate into 100 ml LB broth containing the appropriate antibiotic. Grow overnight at 37°C.
2. Prepare mutagenized plasmid DNA library by following the protocol of the Qiagen Plasmid Miniprep Kit (see Note 4).

3.4. Running the Agarose Gel

1. Clean the agarose gel casting tray by scrubbing with ethanol and then RO water. Seal the sides of the tray with tape.
2. Take TAE buffer in a clean flask and add agarose powder to it to give 0.8% (w/v) final concentration. Heat the contents to dissolve the agarose. Cool slightly by letting the flask remain at room temperature. Then, pour the contents in the agarose gel casting tray. Place the combs of desired thickness (usually 1-mm thick combs, as very small quantity of plasmid is to be loaded). After the gel solidifies, place it in a tray containing TAE buffer.
3. Load 2 µl of plasmid midiprep DNA after mixing with the DNA loading dye (6× loading dye from Fermentas, used as 1× concentration) together with appropriate DNA markers.
4. Run the gel at 100 V (requiring approximately 400 mA) for 30 min. Stain the gel with ethidium bromide and visualize under UV light to confirm the presence of the plasmid DNA (see Note 5).

3.5. Preparing Electrocompetent Cells for Functional Screening

The mutagenized plasmid so isolated has to be transformed in the desired background strain for functional screening of the mutants. Highly efficient electrocompetent cells are required for this.

1. Inoculate single colony of the desired background strain (see Note 6) in 3 ml LB broth, with appropriate antibiotic, and grow overnight at 37°C.
2. Subculture the bacteria in 200 ml LB broth with appropriate antibiotic, by adding 2 ml of overnight culture (sub-culturing with 1% inoculum). Grow at 37°C with shaking at 200 rpm until the OD 600 reaches 0.3–0.4.
3. Stop growth by placing the culture flask on ice.
4. Dispense culture in precooled 50-ml Oakridge tubes and spin at 5,200 × *g* for 3 min at 4°C.
5. Discard the supernatant and resuspend in an equal volume of precooled sterile deionized water (see Note 7). Resuspend completely with the help of a pipette. Take care so as to avoid frothing. Maintain the Oakridge tubes throughout this process on ice (see Note 8).
6. Spin at 5,200 × *g* for 3 min at 4°C. Discard the supernatant immediately and dissolve the a pellet in half the initial volume of precooled sterile deionized water.
7. Spin at 5,200 × *g* for 3 min at 4°C. Discard supernatant immediately and dissolve pellet in one-fourth the initial volume of sterile deionized water.
8. Spin at 5,200 × *g* for 3 min at 4°C. Discard supernatant and dissolve in 5 ml of precooled sterile 10% glycerol.
9. Spin at 5,200 × *g* for 3 min at 4°C. Discard supernatant immediately and finally resuspend the cell pellet in 200 µl of 10% glycerol.
10. Aliquot 45 µl into prechilled 1.5-ml microcentrifuge tubes. Freeze in liquid nitrogen and store at –80°C.

3.6. Electrocompetent Cell Transformation

1. Keep the vial of electrocompetent cells on ice.
2. Add approximately 1 ng of the mutagenized plasmid DNA to the 45 µl of electrocompetent cells (see Note 9).
3. Mix well and keep on ice for 60 s.
4. Pipette the entire mixture into a precooled Bio-Rad 1mm electroporation cuvette. Dry the cuvette from outside by wiping.
5. Place the cuvette in the Bio-Rad shock pod.
6. Using the Bio-Rad Gene Pulser, give 1 pulse, meant for 1mm cuvette (setting is: 1,800 V, 25 µF, 200 Ω) (see Note 10).
7. Immediately add 1 ml of SOC media and mix well.
8. Transfer cells to a 15 ml test tube and incubate with gentle rotation (20 rpm) for 1 h at 37°C.
9. Spread 100–200 µl of the culture onto an appropriate indicator/selection plate for screening the mutant phenotypes (see Note 11).

4. Notes

1. The mutator phenotype of the XL1-red strain cannot be guaranteed if the strain is maintained by the researcher, as the rapid mutation rate affects the chromosome, and after prolonged growth, the subsequent colonies are probably not genetically identical to the original strain. In our experience for the best results, one should use the competent cells directly available from Stratagene. Repeated freeze-thawing of these frozen competent cells greatly reduces the transformation efficiency.
2. Although XL1-red cells contain the tetracycline-resistance gene, due to the rapid mutator phenotype, the cells will frequently give rise to tetracycline-sensitive (Tet^s) variants. Therefore, the resulting transformants may or may not be tetracycline resistant, so it is advised not to add tetracycline to the growth media or the plate.
3. If large-sized colonies are desired, longer incubation times are required. The high mutation rate of the XL1-red mutator strain causes a wide range of colony sizes.
4. The plasmid isolation should be done immediately, and the colonies should not be left on the plate.
5. Multiple DNA topoisomers are usually observed in the gel.
6. We have used *E. coli* WT strain MC4100 Δ*rho galEP3* as background strain for transforming with the mutagenized plasmid library having the transcription terminator *rho* gene (11). The background strain also contains a WT *rho* gene cloned in a shelter plasmid, replication of which is dependent on presence of IPTG. This plasmid can be removed easily in the absence of IPTG following the transformation with the mutagenized library. *galEP3* is an IS2 insertion in the galactose operon, which contains several Rho-dependent terminators. Mutation in Rho protein will cause transcription termination defects, and this strain will utilize galactose on a McKonkey-galactose plate and will appear as red colonies. In general, any strain can be used as a background strain provided that it can be screened for some phenotype of the gene of interest.
7. Alternatively, 10% autoclaved and precooled glycerol can be used instead of sterile deionized water.
8. Steps 5–9, involving washing of the competent cells, should be performed rapidly without any time lag to ensure good transformation efficiency of the cells and also care should be taken while decanting the supernatant after each centrifugation step to avoid losing the cell pellets.

9. The amount of DNA should be determined depending on the size of the plasmid and transformation efficiency of the electrocompetent cells.
10. The pulsing time for the electroporation should be between 4.8 and 5.1 ms.
11. Using this mutator strain, we have successfully screened mutants of different genes of the transcription machinery of *E. coli* and its phages such as *rho* (11), *psu* (12), *nusG*, *N*, *rpoB*, *rpoC*, etc. Mutation rates depend on the copy number of the plasmid and also on the size of genes.

References

1. Rupp WD (1996) DNA repair mechanisms. In: Neidhardt FC, Curtiss R III, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE (eds) *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd edn. ASM, Washington, DC, pp 2277–2294
2. Modrich P, Lahue R (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu Rev Biochem 65:101–133
3. Schofield MJ, Hsich P (2003) DNA mismatch repair: molecular mechanisms and biological function. Annu Rev Microbiol 57:579–608
4. Dao V, Modrich P (1998) Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex. J Biol Chem 273:9202–9207
5. Lu AL, Li X, Gu Y, Wright PM, Chang DY (2001) Repair of oxidative DNA damage: mechanisms and functions. Cell Biochem Biophys 35:141–170
6. Modrich P (1987) DNA mismatch correction. Annu Rev Biochem 56:435–466
7. Radman M, Wagner RE, Glickman W, Meselson M (1980) In: Alacevic M (ed) Progress in environmental mutagenesis. Elsevier, Amsterdam, pp 121–130
8. Scheuermann R, Tam S, Burgers PM, Lu C, Echols H (1983) Identification of the epsilon subunit of *Escherichia coli* DNA polymerase III holoenzyme as the *dnaQ* gene product: a fidelity subunit for DNA replication. Proc Natl Acad Sci U S A 80: 7085–7089
9. Cox EC (1976) Bacterial mutator genes and the control of spontaneous mutation. Annu Rev Genet 10:135–156
10. Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148:1667–1686
11. Chalissery J, Banerjee S, Bandey I, Sen R (2007) Transcription termination defective mutants of rho: role of different functions of rho in releasing RNA from the elongation complex. J Mol Biol 371:855–872
12. Pani B, Ranjan A, Sen R (2009) Interaction surface of bacteriophage P4 protein Psu required for the complex formation with the transcription terminator Rho. J Mol Biol 389(4):647–660

Chapter 30

En Passant Mutagenesis: A Two Step Markerless Red Recombination System

B. Karsten Tischer, Gregory A. Smith, and Nikolaus Osterrieder

Abstract

Bacterial artificial chromosomes are used to maintain and modify large sequences of different origins in *Escherichia coli*. In addition to RecA-based shuttle mutagenesis, Red recombination is commonly used for sequence modification. Since foreign sequences, such as antibiotic resistance genes as well as frt- or loxP-sites are often unwanted in mutant BAC clones, we developed a Red-based technique that allows for the scarless generation of point mutations, deletions, and insertion of smaller and larger sequences. The method employs a sequence duplication that is inserted into the target sequence in the first recombination step and the excision of the selection marker by in vivo I-SceI cleavage and the second Red recombination. To allow for convenient and highly efficient mutagenesis without the use of additional plasmids, the *E. coli* strain GS1783 with a chromosomal encoded inducible Red- and I-SceI-expression was created.

Key words: Red recombination, BAC, Markerless, En passant mutagenesis, I-SceI

1. Introduction

Bacterial artificial chromosomes (BACs) are widely used for cloning large DNA sequences of up to 300 kb, such as eukaryotic genome libraries or full-length genomes of large DNA and RNA viruses (1–4). Since cleavage and ligation or QuikChange™ based methods that cannot be employed with such large constructs, mutagenesis based on Red recombination is mainly used (5). The Red system originates from λ phages and allows for the insertion of linear double-stranded DNA molecules by homologous recombination (Fig. 1). The Red system consists of three components named Exo, Beta, and Gam (5, 6). The Gam protein blocks the *E. coli* RecBCD helicase–nuclease complex, which otherwise would degrade linear DNA from both free ends (7).

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_30, © Springer Science+Business Media, LLC 2010

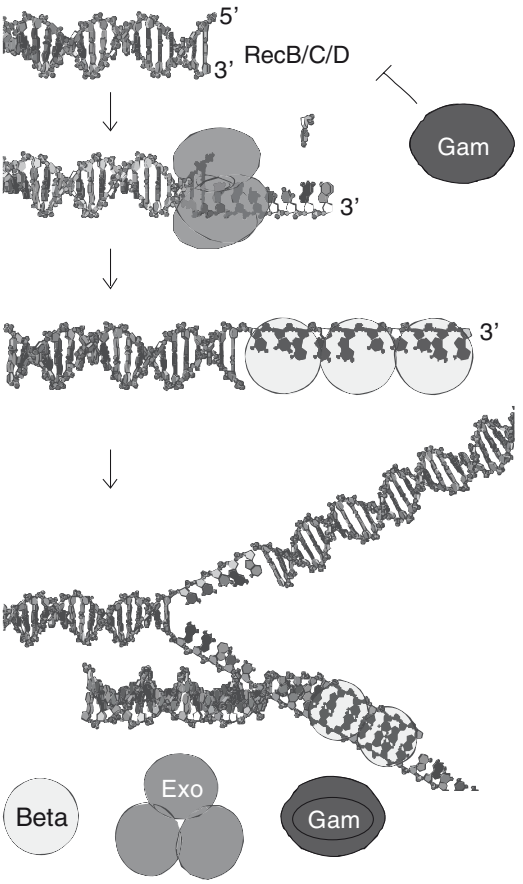


Fig. 1. *Schematic presentation of λ Red recombination.* A double-stranded DNA end is substrate for the Red recombination system. Gam inhibits degradation of linear DNA by the bacterial RecBCD helicase-nuclease complex. Exo forms a toroidal homotrimeric molecule and possesses 5′ → 3′ exonuclease activity. Beta binds to the resulting single-stranded DNA and anneals it to complementary sequences, preferably to the lagging strand.

In a second step, the 5′–3′ exonuclease Exo, a homotrimer, produces a 3′ single strand extension (8, 9), which is protected by the single strand binding activity of the λ protein Beta (10). The second function of Beta is the annealing of the single strand end produced from the linear DNA with complementary sequences (10, 11) of the target DNA, thereby conferring an integration into replicating DNA, with a preference to the lagging strand (12). Red recombination, therefore, is focused on the double-strand end of its substrate, which is introduced into replicating homologous target sequences.

Red recombination can be used to insert a linear DNA, typically a PCR product of an antibiotic resistance gene that can be selected for by *E. coli*, into a target site of the BAC. Recombination is mediated between 30 and 50 bp flanking sequences encoded in the 5′ ends of each PCR primer. Selection for the positive

selection marker certifies a high efficiency of the reaction. Insertion of selectable markers is the most basic application of Red recombination and can be useful to disrupt target genes.

To provide for the insertion of non-selectable sequences (epitope tags, promoters, reporter genes, etc.) and to produce "scarless" point mutations and deletions, the I-SceI enzyme is used. I-SceI is a homing endonuclease that originates from *Saccharomyces cerevisiae*. The enzyme has an 18 bp recognition site that is very rarely present in genome sequences (13). Several systems make it possible to induce *I-sceI* expression in bacteria or eukaryotic cells. The homing endonuclease performs with slow kinetics but is active *in vitro*, in *E. coli* as well as in eukaryotic cells (13, 14).

En passant recombination utilizes a sequence duplication within inserted sequences and an adjoining I-*Sce*I restriction site to remove the marker cassette in a second Red recombination step (15). In detail, the linear DNA used for the first Red recombination harbors a sequence duplication in addition to the selection marker and the targeting flanks (Fig. 2 **I**,

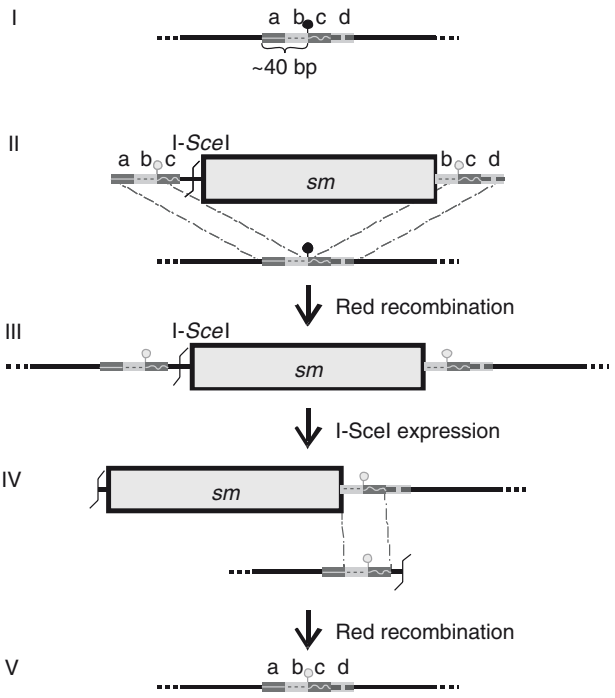


Fig. 2. *Generation of point mutations.* **I –** target sequence; **II** – PCR product of a marker cassette recombines with target sequences; **III** – co-integrate; **IV** – *in vivo* I-SceI cleavage results in new substrate for second Red recombination between duplicated sequences; **V** – final mutant sequences; **a**–**d** – identical sequences of approx. 20 bp each, also represented by bars with identical shape and shading; sm – positive selection marker (e.g., kanamycin resistance gene *aphAI*); – I-*Sce*I restriction site; – sequence to be modified; – modified sequence (may also be insertions that can be fitted into primers such as small epitope tags).

sequences **b** and **c** of the PCR product) and a single I-*Sce*I site (Fig. 2 **I**). In a first recombination step, the whole cassette is integrated into the target site (Fig. 2 **II**), and intermediates (Fig. 2 **III**) are identified via antibiotic resistance. Correct co-integrates are confirmed with RFLP, (colony) PCR, sequencing, or other methods. Subsequently, the expression of the inducible *I-sceI* leads to the production of the homing endonuclease and cleavage of the I-*Sce*I site *in vivo* (Fig. 2 **IV**). The produced double-stranded DNA end with the adjoining duplex sequence is now capable of serving as a substrate for a second Red recombination, which removes the complete marker cassette (Fig. 2 **V**) (15).

1.1. The GS1783 E. coli Strain

Successful application of *en passant* mutagenesis requires successive transient expression of the Red proteins and the I-*Sce*I enzyme. Although these proteins can be expressed from high copy plasmids, the presence of additional plasmids in *E. coli* harboring BAC plasmids adds unwanted complications and work for the investigator. A major advancement to Red recombination was achieved by the insertion of the Red genes in the chromosome of the DH10B strain of *E. coli* under the control of a temperature-inducible promoter. The resulting *E. coli* strain, DY380, greatly simplified Red recombination with BAC plasmids (16). To further adapt DY380 for use in *en passant* mutagenesis, the *I-sceI* gene was inserted in the chromosome under the control of an arabinose-inducible promoter. This was achieved by modifying a derivative of the DY380 strain, EL250, which encoded an arabinose-indubile promoter driving expression of the Flp recombinase gene (*flp*). A scarless replacement of the coding sequences for Flp recombinase with that of I-SceI was performed by the application of the *en passant* methods. An *I-sceI* transfer plasmid was made (transfer plasmids are described below in more detail) and used as a template in a PCR reaction with primers encoding homologies to target recombination with the 5′ and 3′ ends of the *flp* coding sequence. The initial Red recombinant encoded a disrupted *I-sceI::aphAI* gene precisely replacing the *flp* coding sequences, leaving the upstream arabinose-inducible promoter intact. The kanamycin cassette was then removed by introducing an intact *I-sceI* gene on a high copy plasmid (pBAD-*I-sceI*), followed by a second round of Red recombination (15). The pBAD-*I-sceI* could not subsequently be cured from the kanamycin-sensitive strain upon serial passage, demonstrating a major disadvantage of using high-copy plasmids for I-SceI expression. Therefore, the bacteria were transformed with another high-copy plasmid based on the same pUC origin of replication, but encoding kanamycin resistance and an I-*Sce*I restriction site. Following selection with kanamycin, the pBAD-*I-sceI* plasmid was lost

in favor of the new plasmid. The new plasmid was easily cured by adding l-arabinose to the bacterial culture, thereby inducing expression of I-SceI from the bacterial chromosome and cleaving the high-copy plasmid to assist in curing it from the bacteria. This procedure had the added benefit of confirming the expression of the chromosomally encoded I-SceI enzyme. The resulting GS1783 *E. coli* strain (see Subheading 2) is the preferred host for *en passant* mutagenesis and has the following genotype:

DH10B λ *cI*857 Δ(*cro-bioA*)<>*araC*-P_{BAD}*I-sceI*

1.2. Generation of Point Mutations, Deletions and Insertion of Smaller Sequences

To create point mutations, deletions or to insert smaller sequences, such as epitope tags, the primer sequences for the amplification of the marker cassette play a central role. For the design of appropriate primers, one should first create the final sequence to be mutated or introduced. The sequences surrounding the desired mutation are named **a**–**d** with **b** and **c** bracketing the new sequence. Each of those segments is approx. 20 bp in length (Fig. 2 **V**). The forward primer, starting from the 5′ end, is composed of sequences identical to segments **a**–**c** harboring the new sequence. The 3′ end of the primer represents sequences that anneal to the marker cassette. If, for example, plasmid pEPkan-S is used to amplify the selection cassette I-*Sce*I-*aphAI*, the 3′ end of the forward primer is represented by the sequence 5′-tag gga taa cag ggt aat cga ttt-3′ (EPfw) (15). The reverse primer contains 5′ sequences that are the reverse complement of segments **d** to **b** and again contains a 3′ segment annealing to the marker cassette. For the template pEPkan-S, this would be the sequence 5′-gcc agt gtt aca acc aat taa cc-3′ (EPrv) (15). The PCR product of the marker cassette (Fig. 2 **II**) is inserted into the target site (Fig. 2 **I**) via Red recombination of sequences **a**+**b** on the one, and **c**+**d** on the other side. With this reaction, the sequences **b**+**c**, with the new sequence in the middle, are now present as a duplicate in the co-integrate (Fig. 2 **III**). The induction of *I-sceI* expression and the resulting cleavage of the I-*Sce*I site results in a new substrate for a second Red recombination between those duplicated sequences (Fig. 2 **IV**). This reaction releases the complete marker cassette and only leaves behind the desired sequence modification (Fig. 2 **V**).

1.3. Insertion of Larger Sequences

For the insertion of larger sequences (*ins*), such as promoters, genes encoding autofluorescent proteins or complete expression cassettes, a universal transfer construct has to be created. Those constructs have to comprise the three essential features for *en passant* mutagenesis: a positive selection marker, a short sequence duplication, and an adjoining I-*Sce*I site. One way to generate those transfer constructs utilizes a unique restriction site

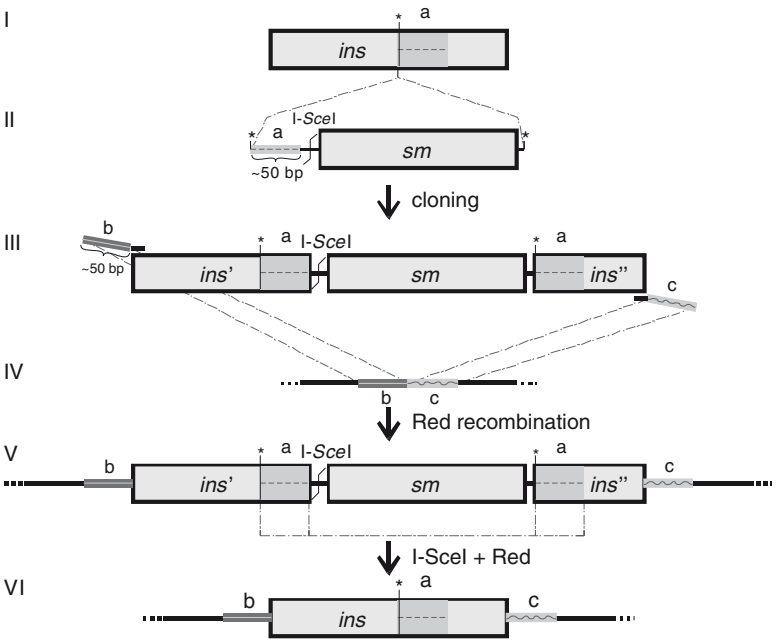


Fig. 3. *The insertion of sequences*. **I** – A unique restriction site (asterisk) within the sequence to be inserted (*ins*) is chosen; **II** – PCR product of a marker cassette with a 50 bp extension cloned into the unique restriction site; **III** – the universal transfer construct is PCR amplified, thereby adding flanks for the first Red recombination; **IV** – target sequence; **V** – co-integrate; **VI** – final insertion mutant with the insertion sequence present in the target site. **a–c** – identical sequences of approx. 50 bp each, also represented by bars with same shape and shading; *ins*, *ins'*, *ins''* – sequence or portions of the sequence that shall be inserted; *sm* – positive selection marker (like the kanamycin resistance gene *aphAI*); ⟆ – I-*Sce*I restriction site.

within the *ins* sequences (Fig. 3 **I**, asterisk). First, the *en passant* marker cassette is PCR amplified. The forward primer for that purpose contains at the 5′ end a restriction site that is identical or compatible to that in *ins*, approx. 50 bases of the *ins* sequences downstream of that site, and a sequence that anneals to the marker cassette, like EPfw. The reverse primer contains a 5′ compatible restriction site and a sequence for annealing to the marker cassette in the reverse orientation (e.g., EPrv). The resulting PCR product (Fig. 3 **II**) is then cloned into the unique restriction site chosen in *ins*, which results in a universally applicable transfer construct (Fig. 3 **III**). Ultimately, the universal transfer construct is amplified by PCR using primers that contain the extensions of 40–50 bases (Fig. 3 **III**, **b** and **c**) for Red based recombination into the target sequence. The linear DNA is introduced into the target sequence with a first Red recombination (Fig. 3 **IV**). Finally, the marker cassette is released from the co-integrate (Fig. 3 **V**) by *in vivo* cleavage of the I-*Sce*I site and a second Red recombination (Fig. 3 **VI**).

2. Materials

1. PCR template: plasmid pEPkan-S (15). Plasmid pEPkan-Scan is obtained from N. Osterrieder (no.34@fu-berlin.de).
2. Taq-polymerase: ABgene ReddyMix™ for PCR.
3. Gel extraction kit: peqGOLD Gel Extraction Kit (peqlab).
4. DpnI (New England BioLabs).
5. *E. coli* strain GS1783, derivate of EL250 (16) but with an L-arabinose inducible *I-sceI* expression cassette instead of *flpe*. The Red recombination system is heat-inducible at 42°C. GS1783 cells can be obtained from Gregory Smith (g-smith3@northwestern.edu).
6. Antibiotics: kanamycin, chloramphenicol (Sigma).
7. LB broth and LB agar: powder from BD – Diagnostic Systems.
8. L-arabinose (Sigma).
9. Electroporator: Gene Pulser (Bio-Rad).
10. 1 mm electroporation cuvettes.
11. Water bath shaker: at 42°C, 220 rpm; C76 water bath shaker (New Brunswick Scientific).
12. Bacteria shaker set at 32°C, 220 rpm.
13. Bacteria incubator at 32°C.

3. Methods

3.1. Preparation of Recombination- and Electrocompetent GS1783

(Method adapted from Lee *et al.* (16))

1. Grow overnight cultures of GS1783 with the respective BAC clones in LB broth with 30 µg/ml chloramphenicol in a bacterial shaker at 32°C (see Notes 1 and 2).
2. Inoculate 5 ml pre-warmed LB broth with 30 µg/ml chloramphenicol (see Note 3) at a 1:50 ratio with the overnight culture. Shake at 32°C, 220 rpm until OD_{600} reaches 0.5–0.7.
3. Transfer culture immediately into water bath shaker at 42°C, 220 rpm for 15 min.
4. Chill bacteria culture for 20 min in ice bath (see Note 4).
5. Spin bacteria for 5 min at ≤4°C, 4,500 × g. Discard supernatant.
6. Resuspend pellet in 1 ml 10% ice-cold glycerol and transfer into 2 ml tubes (see Note 5).
7. Spin bacteria for 1 min at ≤4°C, 12,000 × g. Discard supernatant (see Note 5).

8. Repeat washing steps 6 and 7.
9. Resuspend bacteria with 10% glycerol in a total volume of 50 µl. Keep bacteria on ice before electroporation (see Note 3).

3.2. PCR Product for En Passant Mutagenesis

1. Set up PCR with 0.1 ng template plasmid (e.g. pEPkan-S) and 10 pmol of each primer using a standard Taq polymerase protocol.
2. Cycles: 2 min 95°C, 10× (30 s 95°C, 30 s 52°C, 1 min 72°C), 25× (30 s 95°C, 30 s 68°C, 1 min 72°C), 5 min 72°C, ∞ 4°C (see Note 6).
3. Digest PCR with DpnI to remove template DNA (see Note 7).
4. Run PCR product on an agarose gel and purify amplification product with gel extraction kit (see Note 7).

3.3. Electroporation and First Red Recombination

1. Add approx. 100 ng of PCR product (or other linear DNA) to 50 µl of recombination and electrocompetent bacteria.
2. Transfer DNA/bacteria mix to chilled electroporation cuvette. Electroporate immediately with 15 kV/cm (1.5 kV with 1 mm cuvettes) using settings of 25 µF and 200 Ω.
3. Remove bacteria from cuvette with 1 ml LB broth without antibiotics.
4. Shake bacteria 1–2 h at 32°C.
5. Plate 100 µl on one and the remainder on another LB agar plate with 30 µg/ml chloramphenicol and 30 µg/ml kanamycin (see Note 8).
6. Incubate plates for approx. 24 h at 32°C.
7. Identify positive co-integrates from colonies using RFLP, colony PCR, sequencing or any other appropriate technique.

3.4. Resolution of Co-integrates

1. Inoculate bacteria harboring positive co-integrates into 1 ml of LB broth with 30 µg/ml chloramphenicol. Shake for 1–2 h at 32°C and 220 rpm until the solution becomes faintly cloudy.
2. Add 1 ml of pre-warmed LB broth with 30 µg/ml chloramphenicol and 2% L-arabinose (see Note 9).
3. Shake 1 h at 32°C, 220 rpm.
4. Transfer immediately to 42°C water bath shaker and shake for another 30 min at 220 rpm.
5. Transfer culture to 32°C and shake for 2–3 h at 220 rpm.
6. Take 1 ml of the culture to measure OD_{600}.
7. Plate 5–10 µl of a 1:100 ($OD_{600} \leq 0.5$) or a 1:1,000 ($OD_{600} > 0.5$) dilution at an LB agar plate with 30 µg/ml chloramphenicol and 1% L-arabinose (see Note 10).

8. Incubate at 32°C for 1–2 days until average size bacteria colonies are grown.
9. Optional: Pick replicas of the bacteria colonies from step 8 onto plates with 30 µg/ml chloramphenicol and plates with 30 µg/ml chloramphenicol plus 30 µg/ml kanamycin. The resolved co-integrates should be kanamycin-sensitive (see Note 8).
10. Confirm positive clones with RFLP, colony PCR, sequencing or any other appropriate technique.

4. Notes

1. All mutagenesis reactions should be performed in independent duplicates.
2. The described mutagenesis is done with GS1783 maintaining a BAC clone that harbors a chloramphenicol resistance gene. If constructs with other antibiotic resistance genes or chromosomal sequences are targeted, the used antibiotics should be adapted.
3. Quantities are for one electroporation. For the preparation of large scale batches, the method can be up scaled. The competent bacteria can be snap frozen in aliquots in a dry ice/ethanol bath and stored at −70°C. *Optional*: Test electrocompetence of frozen bacteria batches. It should be $\geq 10^7$ colonies per µg of high copy plasmid.
4. Ice bath should contain water. For better results, the ice bath can be placed on orbital shaker.
5. For larger scales, use corresponding tubes and spin 5 min at ≤4°C, 4,500 × g for washing steps.
6. PCR conditions are for primers with 3′ annealing segments EPfw and EPrv (see Subheading 1.1), both with ≥40 bp 5′ extensions, and with the marker cassette I-*Sce*I-*aphAI* (pEPkan-S) as template. For maximum yields, PAGE purifications of ≥50 bases oligonucleotides are recommended.
7. Usually, step 3 or 4 alone is sufficient. For the first set up, both steps are recommended.
8. Indicated antibiotics are for the mutagenesis of chloramphenicol resistant BAC clones using a marker cassette that confers kanamycin resistance (like I-*Sce*I-*aphAI*).
9. L-arabinose in LB broth should always be freshly prepared, sterile filtered and not autoclaved.
10. To prepare LB agar with 1% L-arabinose, first prepare a 10% solution in LB broth, sterile filter that solution, and finally mix it 1:10 with liquid LB agar at ≤65°C.

Acknowledgments

The authors thank Jens von Einem, Benedikt B. Kaufer, and Felix Wussow for their help in optimizing the presented method. We gratfully recognize Jenifer Klabis for her technical contributions to the isolation of the GS1783 *E. coli* strain.

References

1. Almazan F, Gonzalez JM, Penzes Z, Izeta A, Calvo E, Plana-Duran J, Enjuanes L (2000) Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. Proc Natl Acad Sci USA 97:5516–5521
2. Domi A, Moss B (2002) Cloning the vaccinia virus genome as a bacterial artificial chromosome in *Escherichia coli* and recovery of infectious virus in mammalian cells. Proc Natl Acad Sci USA 99:12415–12420
3. Messerle M, Crnkovic I, Hammerschmidt W, Ziegler H, Koszinowski UH (1997) Cloning and mutagenesis of a herpesvirus genome as an infectious bacterial artificial chromosome. Proc Natl Acad Sci USA 94:14759–14763
4. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci USA 89:8794–8797
5. Zagursky RJ, Hays JB (1983) Expression of the phage lambda recombination genes *exo* and *bet* under lacPO control on a multicopy plasmid. Gene 23:277–292
6. Murphy KC (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. J Bacteriol 180:2063–2071
7. Sakaki Y, Karu AE, Linn S, Echols H (1973) Purification and properties of the gamma-protein specified by bacteriophage lambda: an inhibitor of the host RecBC recombination enzyme. Proc Natl Acad Sci USA 70:2215–2219
8. Kovall R, Matthews BW (1997) Toroidal structure of lambda-exonuclease. Science 277:1824–1827
9. Weissbach A, Korn D (1962) The effect of lysogenic induction on the deoxyribonucleases of *Escherichia coli* K12 lambda. J Biol Chem 237:C3312–C3314
10. Wu Z, Xing X, Bohl CE, Wisler JW, Dalton JT, Bell CE (2006) Domain structure and DNA binding regions of beta protein from bacteriophage lambda. J Biol Chem 281:25205–25214
11. Kmiec E, Holloman WK (1981) Beta protein of bacteriophage lambda promotes renaturation of DNA. J Biol Chem 256:12636–12639
12. Ellis HM, Yu D, DiTizio T, Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci USA 98:6742–6746
13. Colleaux L, D'Auriol L, Betermier M, Cottarel G, Jacquier A, Galibert F, Dujon B (1986) Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. Cell 44:521–533
14. Perrin A, Buckle M, Dujon B (1993) Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. EMBO J 12:2939–2947
15. Tischer BK, von Einem J, Kaufer B, Osterrieder N (2006) Two-step red-mediated recombination for versatile high-efficiency markerless DNA manipulation in *Escherichia coli*. Biotechniques 40:191–197
16. Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, Court DL, Jenkins NA, Copeland NG (2001) A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics 73:56–65

INDEX

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8, © Springer Science+Business Media, LLC 2010

D

E

F

G

H

I

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z