Methods in Molecular Biology™

VOLUME 231

Directed Evolution Library Creation

Methods and Protocols

Edited by

Frances H. Arnold
George Georgiou

HUMANA PRESS

METHODS IN MOLECULAR BIOLOGY™

Directed Evolution Library Creation

Methods and Protocols

Edited by

Frances H. Arnold

California Institute of Technology, Pasadena, CA

and

George Georgiou

University of Texas at Austin, Austin, TX

Humana Press  **Totowa, New Jersey**

999 Riverview Drive, Suite 208
Totowa, New Jersey 07512
humanapress.com

This publication is printed on acid-free paper. ∞
ANSI Z39.48-1984 (American Standards Institute) Permanence of Paper for Printed Library Materials.

Production Editor: Mark J. Breaugh.

Cover design by Patricia F. Cleary.

For additional copies, pricing for bulk purchases, and/or information about other Humana titles, contact Humana at the above address or at any of the following numbers: Tel.: 973-256-1699; Fax: 973-256-8341; E-mail: humana@humanapr.com; Website: humanapress.com

Printed in the United States of America. 10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging in Publication Data

Directed Evolution Library Creation

METHODS IN MOLECULAR BIOLOGY™

John M. Walker, SERIES EDITOR

243. **Chiral Separations:** *Methods and Protocols,* edited by *Gerald Gübitz and Martin G. Schmid, 2004*
242. **Atomic Force Microscopy:** *Methods and Applications,* edited by *Pier Carlo Braga and Davide Ricci, 2004*
241. **Cell Cycle Checkpoint Control Protocols,** edited by *Howard B. Lieberman, 2004*
240. **Mammalian Artificial Chromosomes:** *Methods and Protocols,* edited by *Vittorio Sgaramella and Sandro Eridani, 2003*
239. **Cell Migration in Inflammation and Immunity:** *Methods and Protocols,* edited by *Daniele D'Ambrosio and Francesco Sinigaglia, 2003*
238. **Biopolymer Methods in Tissue Engineering,** edited by *Anthony P. Hollander and Paul V. Hatton, 2003*
237. **G Protein Signaling:** *Methods and Protocols,* edited by *Alan V. Smrcka, 2003*
236. **Plant Functional Genomics**: *Methods and Protocols*, edited by *Erich Grotewold, 2003*
235. ***E. coli* Plasmid Vectors**: *Methods and Applications*, edited by *Nicola Casali and Andrew Preston, 2003*
234. **p53 Protocols**, edited by *Sumitra Deb and Swati Palit Deb, 2003*
233. **Protein Kinase C Protocols**, edited by *Alexandra C. Newton, 2003*
232. **Protein Misfolding and Disease:** *Principles and Protocols*, edited by *Peter Bross and Niels Gregersen, 2003*
231. **Directed Evolution Library Creation**: *Methods and Protocols*, edited by *Frances H. Arnold and George Georgiou, 2003*
230. **Directed Enzyme Evolution**: *Screening and Selection Methods*, edited by *Frances H. Arnold and George Georgiou, 2003*
229. **Lentivirus Gene Engineering Protocols**, edited by *Maurizio Federico, 2003*
228. **Membrane Protein Protocols**: *Expression, Purification, and Characterization*, edited by *Barry S. Selinsky, 2003*
227. **Membrane Transporters:** *Methods and Protocols,* edited by *Qing Yan, 2003*
226. **PCR Protocols, Second Edition,** edited by *John M. S. Bartlett and David Stirling, 2003*
225. **Inflammation Protocols,** edited by *Paul G. Winyard and Derek A. Willoughby, 2003*
224. **Functional Genomics:** *Methods and Protocols,* edited by *Michael J. Brownstein and Arkady B. Khodursky, 2003*
223. **Tumor Suppressor Genes:** *Volume 2: Regulation, Function, and Medicinal Applications,* edited by *Wafik S. El-Deiry, 2003*
222. **Tumor Suppressor Genes:** *Volume 1: Pathways and Isolation Strategies,* edited by *Wafik S. El-Deiry, 2003*
221. **Generation of cDNA Libraries:** *Methods and Protocols*, edited by *Shao-Yao Ying, 2003*
220. **Cancer Cytogenetics:** *Methods and Protocols,* edited by *John Swansbury, 2003*
219. **Cardiac Cell and Gene Transfer:** *Principles, Protocols, and Applications,* edited by *Joseph M. Metzger, 2003*
218. **Cancer Cell Signaling:** *Methods and Protocols,* edited by *David M. Terrian, 2003*
217. **Neurogenetics:** *Methods and Protocols,* edited by *Nicholas T. Potter, 2003*
216. **PCR Detection of Microbial Pathogens:** *Methods and Protocols,* edited by *Konrad Sachse and Joachim Frey, 2003*
215. **Cytokines and Colony Stimulating Factors:** *Methods and Protocols,* edited by *Dieter Körholz and Wieland Kiess, 2003*
214. **Superantigen Protocols,** edited by *Teresa Krakauer, 2003*
213. **Capillary Electrophoresis of Carbohydrates,** edited by *Pierre Thibault and Susumu Honda, 2003*
212. **Single Nucleotide Polymorphisms:** *Methods and Protocols,* edited by *Pui-Yan Kwok, 2003*
211. **Protein Sequencing Protocols, 2nd ed.,** edited by *Bryan John Smith, 2003*
210. **MHC Protocols,** edited by *Stephen H. Powis and Robert W. Vaughan, 2003*
209. **Transgenic Mouse Methods and Protocols,** edited by *Marten Hofker and Jan van Deursen, 2003*
208. **Peptide Nucleic Acids:** *Methods and Protocols,* edited by *Peter E. Nielsen, 2002*
207. **Recombinant Antibodies for Cancer Therapy:** *Methods and Protocols.* edited by *Martin Welschof and Jürgen Krauss, 2002*
206. **Endothelin Protocols,** edited by *Janet J. Maguire and Anthony P. Davenport, 2002*
205. ***E. coli* Gene Expression Protocols,** edited by *Peter E. Vaillancourt, 2002*
204. **Molecular Cytogenetics:** *Protocols and Applications,* edited by *Yao-Shan Fan, 2002*
203. ***In Situ* Detection of DNA Damage:** *Methods and Protocols,* edited by *Vladimir V. Didenko, 2002*
202. **Thyroid Hormone Receptors:** *Methods and Protocols,* edited by *Aria Baniahmad, 2002*
201. **Combinatorial Library Methods and Protocols,** edited by *Lisa B. English, 2002*
200. **DNA Methylation Protocols,** edited by *Ken I. Mills and Bernie H, Ramsahoye, 2002*
199. **Liposome Methods and Protocols,** edited by *Subhash C. Basu and Manju Basu, 2002*
198. **Neural Stem Cells:** *Methods and Protocols,* edited by *Tanja Zigova, Juan R. Sanchez-Ramos, and Paul R. Sanberg, 2002*
197. **Mitochondrial DNA:** *Methods and Protocols,* edited by *William C. Copeland, 2002*
196. **Oxidants and Antioxidants:** *Ultrastructure and Molecular Biology Protocols,* edited by *Donald Armstrong, 2002*
195. **Quantitative Trait Loci:** *Methods and Protocols*, edited by *Nicola J. Camp and Angela Cox, 2002*
194. **Posttranslational Modifications of Proteins:** *Tools for Functional Proteomics*, edited by *Christoph Kannicht, 2002*
193. **RT-PCR Protocols,** edited by *Joe O'Connell, 2002*
192. **PCR Cloning Protocols, 2nd ed.,** edited by *Bing-Yuan Chen and Harry W. Janes, 2002*
191. **Telomeres and Telomerase:** *Methods and Protocols,* edited by *John A. Double and Michael J. Thompson, 2002*
190. **High Throughput Screening:** *Methods and Protocols,* edited by *William P. Janzen, 2002*
189. **GTPase Protocols:** *The RAS Superfamily,* edited by *Edward J. Manser and Thomas Leung, 2002*
188. **Epithelial Cell Culture Protocols,** edited by *Clare Wise, 2002*
187. **PCR Mutation Detection Protocols,** edited by *Bimal D. M. Theophilus and Ralph Rapley, 2002*
186. **Oxidative Stress Biomarkers and Antioxidant Protocols,** edited by *Donald Armstrong, 2002*

Preface

Biological systems are very special substrates for engineering—uniquely the products of evolution, they are easily redesigned by similar approaches. A simple algorithm of iterative cycles of diversification and selection, evolution works at all scales, from single molecules to whole ecosystems. In the little more than a decade since the first reported applications of evolutionary design to enzyme engineering, directed evolution has matured to the point where it now represents the centerpiece of industrial biocatalyst development and is being practiced by thousands of academic and industrial scientists in companies and universities around the world. The appeal of directed evolution is easy to understand: it is conceptually straightforward, it can be practiced without any special instrumentation and, most important, it frequently yields useful solutions, many of which are totally unanticipated. Directed evolution has rendered protein engineering readily accessible to a broad audience of scientists and engineers who wish to tailor a myriad of protein properties, including thermal and solvent stability, enzyme selectivity, specific activity, protease susceptibility, allosteric control of protein function, ligand binding, transcriptional activation, and solubility. Furthermore, the range of applications has expanded to the engineering of more complex functions such as those performed by multiple proteins acting in concert (in biosynthetic pathways) or as part of macromolecular complexes and biological networks.

Not surprisingly, the growth in the ranks of practitioners of directed evolution, and also in the range of new applications, has led to a proliferation of experimental methods aimed at simplifying the process and increasing its efficiency. The purpose of this and the accompanying volume in this series is to provide a compendium of experimental protocols accessible to scientists and engineers with minimal background in molecular biology.

Directed Evolution Library Creation focuses on methods for the generation of molecular diversity. Protocols for random mutagenesis of entire genes or segments of genes, for homologous and nonhomologous recombination, and for constructing libraries in vivo in bacteria and yeast are presented. Every one of these methods has been applied for directed evolution purposes. The optimal choice depends on the many factors that characterize each evolution problem, and we have often found that any of several different methods will work. Though there may be multiple molecular solutions to any given functional problem, the library made for directed evolution must nonetheless contain at least one of those solutions. And, the higher the frequency of potential solu-

tions, the easier it is to find them. Thus, the choice of method for creating molecular diversity and its particular implementation are important. In addition to the various protocols for creating libraries, this volume also includes three chapters that describe ways to analyze libraries, particularly those made by recombination.

No directed evolution experiment is successful without a good screen or selection. *Directed Enzyme Evolution: Screening and Selection Methods* is devoted entirely to selection and screening methods that can be applied to directed evolution of enzymes. Directed evolution is not difficult, and these protocols, prepared by practitioners from many leading laboratories, should make this robust protein engineering approach accessible to anyone with a good problem.

Frances H. Arnold

George Georgiou

Contributors

VALÉRIE ABÉCASSIS • *Centre de Génétique Moléculaire, CNRS, Gif-sur-Yvette Cedex, France*

ANNA MARIE AGUINALDO • *Xencor Inc., Monrovia, CA*

MIGUEL ALCALDE • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA*

FRANCES H. ARNOLD • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA*

GEETHANI BANDARA • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA*

PAUL H. BESSETTE • *Department of Chemical Engineering, University of California at Santa Barbara, Santa Barbara, CA*

THOMAS BULTER • *Department of Chemical Engineering, University of California at Los Angeles, Los Angeles, CA*

PATRICK C. CIRINO • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA*

WAYNE M. COCO • *Tanox Inc., Houston, TX*

PATRICK S. DAUGHERTY • *Department of Chemical Engineering, University of California, Santa Barbara, CA*

OLGA ESTEBAN • *Department of Chemical and Biomolecular Engineering, University of Illinois, Urbana, IL*

RADU GEORGESCU • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA*

TAKAHIRO HOHSAKA • *Department of Bioscience and Biotechnology, Okayama University. Okayama, Japan*

LOÏC JAFFRELO • *Centre de Génétique Moléculaire, CNRS, Gif-sur-Yvette Cedex, France*

JOHN M. JOERN • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA*

LAWRENCE A. LOEB • *Department of Biochemistry, University of Washington, Seattle, WA*

STEFAN LUTZ • *Department of Chemistry, Emory University, Atlanta, GA*

KIMBERLY M. MAYER • *Biology Department, Brookhaven National Laboratory, Upton, NY*

PETER MEINHOLD • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA*

MARCO A. MENA • *Department of Chemical Engineering, University of California, Santa Barbara, CA*

KENTARO MIYAZAKI • *National Institute of Advanced Industrial Science and Technology (AIST), Ibaraki, Japan*
HIROSHI MURAKAMI • *Department of Bioscience and Biotechnology, Okayama University, Okayama, Japan*
ANNALEE W. NGUYEN • *Department of Chemical Engineering, University of California, Santa Barbara, CA*
MARC OSTERMEIER • *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD*
DENIS POMPON • *Centre de Génétique Moléculaire, CNRS, Gif-sur-Yvette Cedex, France*
VOLKER SCHELLENBERGER • *Genencor International Inc., Palo Alto, CA*
OLGA SELIFONOVA • *Cargill, Inc., Minneapolis, MN*
VOLKER SIEBER • *Degussa Texturant Systems GmbH, Freising, Germany*
JONATHAN J. SILBERG • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA*
MASAHIKO SISIDO • *Department of Bioscience and Biotechnology, Okayama University, Okayama, Japan*
JESSICA L. SNEEDEN • *Department of Biochemistry, University of Washington, Seattle, WA*
LIANHONG SUN • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA*
ALEXANDER V. TOBIAS • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA*
GILLES TRUAN • *Centre de Génétique Moléculaire, CNRS, Gif-sur-Yvette Cedex, France*
ANDREW K. UDIT • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA*
DAISUKE UMENO • *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA*
RYAN D. WOODYER • *Department of Chemistry, University of Illinois, Urbana, IL*
WENJUAN ZHA • *Department of Chemical and Biomolecular Engineering, University of Illinois, Urbana, IL*
HUIMIN ZHAO • *Department of Chemical and Biomolecular Engineering, University of Illinois, Urbana, IL*
TONGBO ZHU • *Department of Chemical and Biomolecular Engineering, University of Illinois, Urbana, IL*

Contents

I

Mutagenesis and Recombination Methods

1

Generating Mutant Libraries Using Error-Prone PCR

Patrick C. Cirino, Kimberly M. Mayer, and Daisuke Umeno

1. Introduction

Directed evolution has become a powerful tool not only for improving the utility of enzymes in industrial processes, but also to generate variants that illuminate the relationship between enzyme sequence, structure, and function. The method most often used to generate variants with random mutations is error-prone PCR. Error-prone PCR protocols are modifications of standard PCR methods, designed to alter and enhance the natural error rate of the polymerase ***(1,2)***. *Taq* polymerase ***(3)*** is commonly used because of its naturally high error rate, with errors biased toward AT to GC changes. However, recent protocols include the use of a newly-developed polymerase whose biases allow for increased variation in mutation type (i.e., more GC to AT changes) (*see* **Note 1**).

Error-prone PCR reactions typically contain higher concentrations of $MgCl_2$ (7 m*M*) compared to basic PCR reactions (1.5 m*M*), in order to stabilize non-complementary pairs ***(4,5)***. $MnCl_2$ can also be added to increase the error-rate ***(6)***. Other ways of modifying mutation rate include varying the ratios of nucleotides in the reaction ***(7–9)***, or including a nucleotide analog such as 8-oxo-dGTP or dITP ***(10)***. Fenton et al. ***(11)*** describe a mutagenic PCR protocol that uses dITP as well as provide an analysis of the effects of dITP and Mn^{2+} on PCR products. Mutation frequencies from 0.11 to 2% (1 to 20 nucleotides per 1 kb) have been achieved simply by varying the nucleotide ratio and the amount of $MnCl_2$ in the PCR reaction ***(12)***. The number of genes that contain a mutation can also be modified by changing the number of effective doublings by increasing/decreasing the number of cycles or by changing the initial template concentration.

From: *Methods in Molecular Biology, vol. 231: Directed Evolution Library Creation: Methods and Protocols*
Edited by: F. H. Arnold and G. Georgiou

Given the same error-prone PCR conditions, two different genes will likely exhibit different mutation frequencies, primarily depending on the length and base composition of the template DNA. Thus, the best way to check the mutation frequency in an experiment is to estimate it from the fraction of inactive clones by sampling small numbers (one 96-well plate) of the error-prone PCR library. This is also a good way to test various conditions to obtain an appropriate level of mutation that allows variants with improvements to be isolated. *See* **Chapter 8** in the companion volume, *Directed Enzyme Evolution: Screening and Selection Methods*, for more detailed information on library analysis.

An expression system and high-throughput assay should be developed before a library of enzyme variants is generated. To take full advantage of the power of error-prone PCR, the assay must be accurate enough to detect small improvements and sensitive enough to detect the low levels of activity typically encountered in the beginning rounds of an evolution experiment.

2. Materials

2.1. Biological and Chemical Materials

1. Appropriate PCR amplification primers, designed to have similar melting temperatures, stored at –20°C (*see* **Note 2**).
2. Plasmid containing gene of interest to be amplified by mutagenic PCR.
3. 50X dNTP mixture: 10 m*M* each of dATP, dTTP, dCTP, and dGTP (Roche, Indianapolis, IN). Prepare 20 µL aliquots of this mixture (to avoid excessive freeze/thaw cycles) and store at –20°C.
4. Individual solutions of dNTPs (10 m*M*), stored as aliquots at –20°C (*see* **Note 3**).
5. *Taq* polymerase (Roche, Indianapolis, IN), stored at –20°C (*see* **Notes 3** and **4**).
6. 10X Normal PCR Buffer (comes with Roche *Taq* polymerase): 15 m*M* $MgCl_2$, 500 m*M* KCl, 100 m*M* Tris-HCl, pH 8.3, stored at –20°C.
7. 10X $MgCl_2$ solution: 55 m*M* prepared in water. (Sterilize before use.)
8. 1 m*M* solution of $MnCl_2$ prepared in water. (Sterilized before use.)
9. Agarose gels: 1% LE agarose in 1X TAE (40 m*M* Tris-acetate, 1 m*M* ethylene diamine tetraacetic acid [EDTA]), 0.5 µg/mL ethidium bromide.
10. PCR purification kit (Zymoclean Kit; Zymo Research, Orange, CA).
11. Appropriate restriction enzyme(s) (New England Biolabs (NEB), Beverly, MA), stored at –20°C.
12. T4 DNA ligase (Roche, Indianapolis, IN), stored at –20°C.
13. Suitable vector for expressing the mutant library: digested, gel-purified, and ready for ligation of PCR insert.
14. Competent microbial strain(s).
15. Appropriate antibiotic(s).
16. LB (Luria Broth) and LB-agar plates containing antibiotics. (Sterilize before use.)
17. Water. (Sterilize before use.)
18. High-throughput screening materials (e.g., 96-well plates for cell culture and library expression, 96-well microplates for screening, plate reader, and the like.)

2.2. Equipment

1. Microcentrifuge (Eppendorf 5417R, Brinkmann Instruments, Westbury, NY).
2. Thermocycler (Model PTC200, MJ Research, Waltham, MA).
3. Agarose gel running system.

3. Methods

3.1. Error-Prone PCR Using Taq Polymerase

1. Prepare purified plasmid DNA and determine its concentration (*see* **Note 5**).
2. For each PCR sample, add to tube:

   10 µL 10X normal error-prone PCR buffer,
   2 µL 50X dNTP mix,
   Additional dNTPs (optional) (*see* **Note 6**),
   10 µL 55 mM $MgCl_2$ $MnCl_2$ (optional) (*see* **Note 7**),
   30 pmol each primer,
   2 fmol template DNA (~10 ng of an 8-kb plasmid) (*see* **Note 8**),
   1 µL *Taq* polymerase (5U),
   H_2O to a final volume of 100 µL.
3. Mix sample.
4. Place tubes in thermocycler.
5. Run Error-Prone PCR Program (*see* **Note 9**):

   30 s at 94°C, 30 s at annealing temperature for primers (*see* **Note 10**),
   1 min at 72°C (for a ~1 kb gene) (*see* **Note 11**),
   14–20 cycles (*see* **Note 12**),
   5–10 min at 72°C final extension,
   4°C (to protect samples overnight if necessary).
6. Run a sample of the product on a gel to estimate the yield of full-length gene.
7. Purify PCR products either by gel electrophoresis (removes plasmid DNA) or by Zymoclean Kit (*see* **Note 13**).
8. Digest with appropriate restriction enzymes (*see* **Note 2**). Clean the digested insert, ligate into expression vector, and transform the mutant library into appropriate host strain (*see* **Note 14**).
9. Grow cultures expressing the mutant library (e.g., in 96-well format) and perform the corresponding enzyme activity assay.
10. Determine from the activity profiles of the expressed mutant libraries the most suitable error conditions for screening, and continue screening that library (*see* **Note 15**). *See* **Fig. 1** for example activity profiles from mutant libraries prepared under various mutagenic conditions. In general, it is desirable to obtain mutants that contain only a single amino-acid substitution compared to the parent sequence. Higher mutation rates make it difficult to distinguish beneficial point mutations from those that are neutral or even slightly deleterious. Additionally, the fraction of mutants with improved function decreases as the mutation rate is increased. Thus, an appropriate PCR error rate for directed evolution corresponds to a mutation frequency of ~2 to 5 base substitutions per gene.

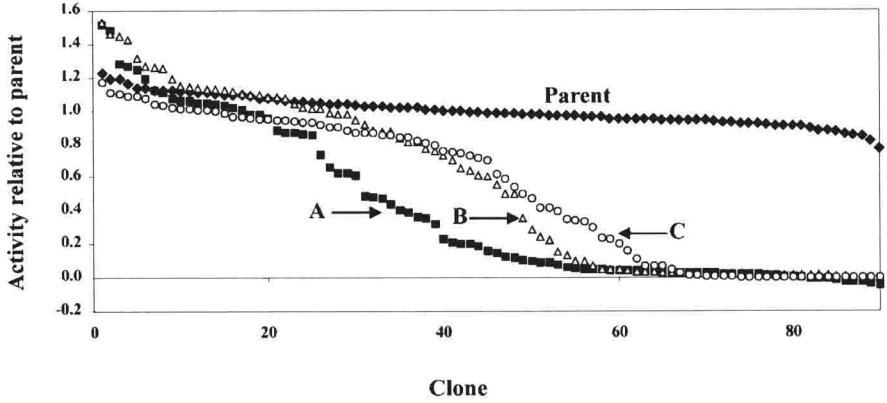

Fig. 1. Activity profiles for libraries made under different mutagenic PCR conditions. Activities are reported relative to the average activity of the parent enzyme used to prepare that generation and are plotted in descending order. The parent gene is 1.4 kb and codes for the heme domain of cytochrome P450 BM-3. The plot labeled **Parent** (◆) represents parent enzyme activity measured across an entire 96-well plate. The standard deviation in parent activity is 9.2%. The remaining three plots depict the activity profiles from 96-well plates containing different mutant libraries. All three error-prone PCR reactions contained 20 fmole of the parent gene as template, plus 7 m*M* $MgCl_2$, 0.2 m*M* each of dGTP and dATP, and 1.0 m*M* each of dCTP and dTTP. Additionally, reaction **A** (■) contained 0.1 m*M* $MnCl_2$, reaction **B** (△) contained 0.05 m*M* $MnCl_2$, and reaction **C** (○) contained no $MnCl_2$. Libraries **A**, **B**, and **C**, respectively, consist of 45%, 40%, and 31% mutants with less than 10% of the parent enzyme's activity.

Typically an error rate resulting in a library with 30–40% of mutants having less than 10% of the parent enzyme's activity (i.e., "dead" mutants) is suitable, although this value will vary depending on the enzyme and the function assayed.

4. Notes

1. Stratagene's Genemorph kit, which includes its own error-prone PCR protocol, uses a polymerase ("Mutazyme") that exhibits a mutation bias quite different from that of *Taq* polymerase. Whereas *Taq* polymerase preferentially introduces AT to GC mutations, Mutazyme mutations are biased toward GC to AT changes. It may be desirable to combine the mutation biases of these polymerases by alternating between them in successive generations, or by creating separate mutant libraries using both polymerases in a single generation.
2. Error-Prone PCR primers can be designed to anneal outside the restriction sites that will be used for subcloning or can be designed to include the restriction sites as part of the primer sequence. In our experience, higher levels of ligation efficiency are obtained when primers are located far outside the restriction sites, presumably because of better digestion efficiency.