Chapter 1

Error-Prone PCR and Effective Generation of Gene Variant Libraries for Directed Evolution

Janine N. Copp, Paulina Hanson-Manful, David F. Ackerley, and Wayne M. Patrick

Abstract

Any single-enzyme directed evolution strategy has two fundamental requirements: the need to efficiently introduce variation into a gene of interest and the need to create an effective library from those variants. Generation of a maximally diverse gene library is particularly important when employing nontargeted mutagenesis strategies such as error-prone PCR (epPCR), which seek to explore very large areas of sequence space. Here we present comprehensive protocols and tips for using epPCR to generate gene variants that exhibit a relatively balanced spectrum of mutations and for capturing as much diversity as possible through effective cloning of those variants. The detailed library preparation methods that we describe are generally applicable to any directed evolution strategy that uses restriction enzymes to clone gene variants into an expression plasmid.

Key words Directed evolution, Random mutagenesis, Error-prone PCR, GeneMorph II, Mutazyme II DNA polymerase, Library generation, Library analysis

1 Introduction

Directed evolution at a single gene level is a powerful tool for effecting rapid improvement in a desired protein function. Recent years have seen an increasing trend toward the design of targeted "smart" libraries, which exploit available structural, functional, and/or sequence-derived information to preselect promising residues for mutagenesis. These approaches seek to minimize the number of gene variants that must be screened to achieve a successful outcome [1–3] and are particularly valuable when only a low to moderate throughput screen is available (e.g., 10^2–10^4 variants per day). In contrast, approaches like error-prone PCR (epPCR) provide access to an almost unlimited number of variants, but generally have a lower hit rate and are therefore better suited to high throughput screening or selection strategies. Nevertheless, it is worth noting that epPCR can still yield rapid gains in fitness

Elizabeth M.J. Gillam et al. (eds.), *Directed Evolution Library Creation: Methods and Protocols*, Methods in Molecular Biology, vol. 1179, DOI 10.1007/978-1-4939-1053-3_1,

for some promiscuous functions; for example, one of us (DFA) was involved in a small-scale epPCR study (screening only 6,000 clones) that generated an evolved chromate reductase variant exhibiting >200-fold improvement in k_{cat}/K_M [4].

The great strengths of error-prone PCR are its simplicity and applicability; anyone who has the capability to amplify and clone a PCR product should be able to generate a high-quality epPCR library. The main limitation stems from the inherent randomness of epPCR; residues distant from the active site are just as likely to be mutated as those within, but while the former are more numerous, the latter are more likely to yield substantial improvements in activity [5]. Thus, relative to more targeted strategies, higher proportions of synonymous, neutral, or deleterious base substitutions are to be expected. Furthermore, even a balanced mutation spectrum at the genetic level yields bias at the protein level. It has been calculated that a single base substitution in a sense codon provides access to only 5.7 of the 19 alternative amino acids on average [6], owing to the degeneracy and structure of the genetic code. In practice epPCR is unlikely to generate multiple base substitutions within the same codon, and this diminishes the amount of sequence space that can be sampled.

While it is important to keep these limitations in mind, each issue also brings potential benefit. The stochastic nature of epPCR can prove advantageous as, unrestricted by preconceived notions about which residues are most likely to yield enhanced variants, entirely unpredictable mutations can be recovered that yield substantial gains in fitness (e.g., mutations of residues distant from the active site that improve substrate access, solubility, or stability [7, 8]). Likewise, although single base substitutions provide access to only a limited pool of alternative amino acids, those residues are slightly more likely to be tolerated than amino acids that are only accessible via multiple base substitutions [9].

In all these considerations, it is implicit that generation of a variant library that exhibits a balanced mutational spectrum is a desirable outcome (this argument is expanded in [10]). The original and most economical methods for performing epPCR are by enhancing the intrinsic error rate of *Taq* polymerase by the addition of Mn^{2+} to the reaction buffer to reduce base pairing specificity and/or by modifying the ratios of dNTPs present [11–13]. However, these methods generate a strong mutation bias. In particular, addition of Mn^{2+} results in an over-representation of mutations at A:T base pairs [14]. Similarly, methods that employ triphosphate derivatives of nucleoside analogues also tend to generate heavily biased libraries [15]. To minimize bias, we advocate use of the GeneMorph® II Random Mutagenesis Kit (Agilent Technologies; manual available at http://www.chem.agilent.com/library/usermanuals/Public/200550.pdf; May 2013). This kit employs a mixture of two proprietary error-prone polymerases that

collectively provide a relatively balanced mutational spectrum (albeit still slightly favoring mutations at A:T base pairs). Some fine-tuning of the mutation rate is also feasible by varying the initial concentration of template DNA and the number of amplification cycles. The relatively unbiased mutational spectrum has been independently verified by numerous independent practitioners, e.g., [15], and in our experience the mutation rate is far easier to control than alternative epPCR strategies. Nevertheless, readers who prefer to use a *Taq*/Mn^{2+}-based method for epPCR should find it straightforward to adapt the methodology of Fujii et al. (Chapter 2 of this volume) to a linear rather than rolling circle PCR format and to substitute this for the methodology that we describe under Subheading 3.2.

This chapter is based on a consensus of "best practice" approaches for generating an amplified collection of gene variants by epPCR and for restriction enzyme-based cloning of those variants into a plasmid-borne gene library, as employed in two different research laboratories. In this way, it provides a different perspective to a chapter that two of us (P.H.M. and W.M.P.) published in an earlier volume [16]. Wherever we felt that the best strategy was not clear-cut (i.e., wherever we bickered among ourselves!), we have added notes that describe alternative approaches. In this manner, we hope to provide the reader with a comprehensive series of options that can easily be tailored to existing facilities and preferences.

2 Materials

2.1 Vector Preparation and Analysis

1. Protein expression vector, into which the epPCR product will be cloned (*see* **Notes 1** and **2**).
2. Qiagen Plasmid Midi Kit (Qiagen; Valencia, CA, USA). Equivalent kits from other manufacturers are also suitable. Alternatively, plasmids can be purified using alkaline lysis, to yield DNA at high concentrations [17].
3. Restriction enzymes that enable directional, sticky-ended cloning of the epPCR product. If using two restriction enzymes, as we recommend, then ensure that they are active in a single, compatible buffer (*see* **Note 3**).
4. Restriction enzyme that cuts within the stuffer fragment (*see* **Note 1**) of the expression vector (*see* **Notes 2** and **3**).
5. Agarose gels stained with 1× SYBR Safe DNA gel stain (Invitrogen; Carlsbad, CA, USA).
6. Safe Imager 2.0 Blue-Light Transilluminator (Invitrogen; Carlsbad, CA, USA).
7. Clean razor blades or other implements for excising bands from gels.

8. QiaQuick Gel Extraction Kit (Qiagen; Valencia, CA, USA). Equivalent kits from alternate suppliers are also suitable.
9. DNA Clean & Concentrator Kit (Zymo Research; Irvine, CA, USA). Equivalent kits from other manufacturers are also suitable or DNA can be concentrated using ethanol precipitation [17].
10. Spectrophotometer to measure DNA concentration in small volumes, such as a Nanodrop 1000 (Thermo Scientific; Pittsburgh, PA, USA). If a Nanodrop spectrophotometer is not available, DNA quantification can be performed by agarose gel electrophoresis utilizing a DNA quantification ladder, e.g., the 100 bp DNA ladder from New England Biolabs (Ipswich, MA, USA).

2.2 Error-Prone PCR and Preparation of epPCR Insert for Ligation into Vector

1. Plasmid or genomic DNA containing the gene that is to be amplified by epPCR.
2. Phusion High-Fidelity polymerase (New England Biolabs; Ipswich, MA, USA). This particular brand of DNA polymerase is recommended due to its buffer compatibility with DpnI (https://www.neb.com/tools-and-resources/usage-guidelines/activity-of-restriction-enzymes-in-a-taq-or-phusion-pcr-mix).
3. Oligonucleotide primers for amplification of the target gene (*see* **Note 4**).
4. DNA Clean & Concentrator Kit (*see* Subheading 2.1, **item 9**).
5. Nanodrop spectrophotometer to measure DNA concentration (*see* Subheading 2.1, **item 10**).
6. GeneMorph II Random Mutagenesis Kit (catalog #200550, Agilent; Santa Clara, CA, USA). The kit contains Mutazyme II DNA polymerase (2.5 U/μl), 10× Mutazyme II reaction buffer, a dNTP mix (10 mM of each dNTP) and a 1.1 kb gel quantification standard (20 ng/μl).
7. Thermocycler with a heated lid (e.g., an MJ Mini from Bio-Rad; Hercules, CA, USA).
8. Agarose gels, stained with ethidium bromide at 0.5 μg/ml, and apparatus for agarose gel electrophoresis.
9. Restriction enzymes that facilitate directional, sticky-ended cloning (*see* Subheading 2.1, **item 3**).
10. Restriction enzyme DpnI (New England Biolabs; Ipswich, MA, USA).

2.3 Vector Analysis and Test Library Preparation

1. Digested vector.
2. Digested PCR insert.
3. T4 DNA ligase and ligation buffer. We obtain comparable results using each of the T4 DNA ligases from New England

Biolabs (Ipswich, MA, USA), Fermentas (Thermo Scientific; Pittsburgh, PA, USA) and Enzymatics Inc. (Beverly, MA, USA).

4. Nanodrop spectrophotometer to measure DNA concentration (*see* Subheading 2.1, **item 10**).
5. Aliquots (40 μl) of electrocompetent *E. coli* cells (*see* **Note 5**).
6. Electroporation unit and cuvettes, e.g., Gene Pulser electroporation unit with Pulse Controller (Bio-Rad; Hercules, CA, USA) and Gene Pulser electroporation cuvettes with 0.2 cm electrode gap (Bio-Rad; Hercules, CA, USA).
7. Sterile SOC medium (20 g/L tryptone; 5 g/L yeast extract; 10 mM NaCl; 2.5 mM KCl; 20 mM glucose).
8. LB-agar plates containing the correct antibiotic for selecting plasmid-containing cells.

2.4 Analysis of Library Composition

1. Thermocycler with a heated lid (e.g., an MJ Mini from Bio-Rad; Hercules, CA, USA).
2. Primers for amplifying cloned inserts from the epPCR library (*see* **Note 4**).
3. Reagents for a standard PCR screen. While there are many alternative (and equally good) suppliers, we routinely use 2× BioMix Red (Bioline; Alexandria, NSW, Australia). Equivalent kits from other manufacturers are also suitable.
4. Ethidium bromide-stained agarose gels and electrophoresis apparatus (*see* Subheading 2.2, **item 8**).
5. DNA Clean & Concentrator Kit (*see* Subheading 2.1, **item 9**).

2.5 Construction and Storage of the Full-Sized Library

1. Digested vector with minimal background level of partially digested DNA.
2. Purified epPCR-amplified insert digested with appropriate restriction enzymes.
3. T4 DNA ligase and ligation buffer (*see* Subheading 2.3, **item 3**).
4. DNA Clean & Concentrator Kit (*see* Subheading 2.1, **item 9**).
5. Nanodrop spectrophotometer to measure DNA concentration (*see* Subheading 2.1, **item 10**).
6. LB-agar plates containing the correct antibiotic for selecting plasmid-containing cells.
7. Electrocompetent *E. coli* cells (*see* **Note 5**).
8. Electroporation cuvettes and apparatus (*see* Subheading 2.3, **item 6**).
9. SOC medium (*see* Subheading 2.3, **item 7**).
10. Supercoiled control plasmid of known concentration (e.g., pUC19, 10 pg/μl; Invitrogen (Carlsbad, CA, USA)).

11. Sterile 50 ml tubes (Falcon or similar).
12. LB medium supplemented with the appropriate antibiotic (~20 ml, total).
13. Refrigerated centrifuge with a rotor that takes 50 ml tubes (*see* Subheading 2.5, **item 11**).
14. Spectrophotometer and cuvettes for measuring cell density (OD_{600}), e.g., BioPhotometer and UVettes from Eppendorf (Hamburg, Germany).
15. Sterile glycerol (50 % v/v).
16. Cryogenic vials, suitable for storage at −80 °C (*see* **Note 6**).

3 Methods

3.1 Vector Preparation and Analysis

1. Midiprep the plasmid that will be used for cloning and expression of the library using a Qiagen midiprep kit or equivalent (*see* **Notes 1** and **2**).
2. Digest 10 μg of the vector in a 500 μl reaction with 1× restriction enzyme buffer using 25 U of each restriction enzyme. Incubate the reaction at 37 °C overnight (*see* **Note 3**).
3. Add a further 10 U of an enzyme that cuts within the stuffer fragment (*see* **Note 2**). Incubate the reaction for an additional 2 h at 37 °C and then heat inactivate the enzymes according to the manufacturer's guidelines.
4. Run the entire restriction digest on a 0.8 % w/v agarose gel and excise the DNA fragment corresponding to the digested vector. We strongly recommend the use of a blue-light transilluminator and a compatible stain (SYBR Safe) for this step rather than ethidium bromide and a UV transilluminator (*see* **Note 7**). Purify the excised DNA with the QiaQuick Gel Extraction Kit, according to the manufacturer's guidelines. Use 30–50 μl Elution Buffer to elute the DNA from the spin column (*see* **Note 8**).
5. Quantify the purified vector using a Nanodrop spectrophotometer or equivalent.
6. Aliquot the vector and store at −20 °C. We typically aliquot 200–500 ng of the digested vector into multiple microcentrifuge tubes so they can be thawed and used as required. This reduces the number of freeze/thaw cycles and maintains high-quality stocks of digested vector for short-term use.
7. Ligate a test fragment (*see* **Note 9**) with 100 ng of the digested vector. Use a 3:1 molar ratio of digested insert DNA to digested vector in a 10 μl reaction with 1× T4 DNA Ligase buffer and 1 U ligase. Incubate overnight at 16 °C.

8. Prepare a control reaction as above but containing digested vector only (no insert).
9. Thaw three 40 μl aliquots of electrocompetent *E. coli* cells on ice (*see* **Note 5**). Add 1 μl of the test ligation reaction to one of the cell aliquots. Add 1 μl of the control ligation reaction to a second aliquot of cells. Add an appropriate amount of a plasmid control to the third aliquot of cells, to determine their transformation efficiency. We routinely use 10 pg of an uncut plasmid control, containing the same antibiotic resistance cassette as the vector (*see* **Note 10**). Transfer each aliquot to a prechilled electroporation cuvette, electroporate (e.g., 2.5 kV, 200 Ω, and 25 μF in a Gene Pulser unit with Pulse Controller), and immediately add 460 μl of SOC medium to the cuvette. Transfer the cell suspension to a sterilized, capped microcentrifuge test tube (or sterile 15 ml tube). Allow the cells to recover by incubating them for 1 h at 37 °C and 200 rpm. Spread 10, 20, and 50 μl aliquots of the cell suspensions onto selective agar plates. Incubate the plates overnight at 37 °C.
10. Count the number of colonies on each plate. Use the results from the "vector only" plates to calculate the fraction of the test ligation (as represented on the "vector + insert" plates) that contains recircularized or uncut vector. This background must be minimized, to avoid wasting time on futile library screens. If the "vector only" background is >1 % of the "vector + insert" colony count, then we recommend discarding all aliquots of the digested vector and preparing a fresh batch. Use the plasmid control to calculate the transformation efficiency of your electrocompetent cells. The easiest way to increase the size of an epPCR library is to improve the transformation efficiency of the cells (*see* **Notes 5** and **10**).

3.2 Template Preparation and Error-Prone PCR

1. Amplify your target gene with Phusion High-Fidelity polymerase. We typically use 20–50 μl reactions following the manufacturer's protocol (*see* **Note 11**).
2. Run 1–2 μl of the PCR reaction on an ethidium bromide-strained gel to confirm appropriate amplification of the target gene.
3. Add 10 U DpnI directly to PCR mix and incubate for 1 h at 37 °C.
4. Heat inactivate the DpnI reaction for 20 min at 65 °C and then purify using the DNA Clean & Concentrator Kit.
5. Quantify the purified DNA using a Nanodrop spectrophotometer.
6. The amount of template and the number of amplification cycles used in the epPCR affects the mutation rate (*see* **Note 12**). Calculate the amount of template DNA and the number of PCR cycles that are required for the desired mutation rate.

7. Prepare the epPCR reagents in a thin-walled, 0.2 ml tube containing 5 μl 10× Mutazyme II reaction buffer, 1 μl dNTP mix (gives 200 μM of each dNTP, final concentration), 2 μl forward primer (from 10 μM stock solution), 2 μl reverse primer (from 10 μM stock solution), 1 μl Mutazyme II DNA polymerase (2.5 U), DNA template, and double-distilled water (ddH_2O) to a total volume of 50 μl.
8. Mix the sample and place the tube in the thermocycler. If using a low template concentration, at this point the PCR mix may be split into five separate tubes, each containing 10 μl, to avoid a founder effect from mutations introduced during an early cycle. A negative control (with no template DNA added) should also be prepared to check for contamination (*see* **Note 13**).
9. Run the epPCR program beginning with an initial denaturation step of 1 min at 95 °C, followed by 30 cycles (*see* **Note 12**) of 20 s at 94 °C, 20 s at the optimal annealing temperature for primers (*see* **Note 4**) and 1 min at 72 °C (for a ~1 kb gene; *see* **Note 14**). Complete the epPCR with a 2 min elongation step at 72 °C and then hold at 4 °C (for product storage, if necessary).
10. The total yield of epPCR product is required to calculate the PCR efficiency. Run an aliquot of the epPCR product (typically 2 μl) on an ethidium bromide-stained agarose gel, alongside the 1.1 kb gel standard (*see* Subheading 2.2, **item 6**). Determine the total yield of the epPCR product by comparing the intensity of the epPCR sample with the intensity of the gel standard (*see* **Note 15**).
11. Purify the remainder of the epPCR product using the DNA Clean & Concentrator Kit. Elute the purified DNA from the spin column in 30 μl sterile water heated to 50 °C (*see* **Note 8**).
12. Quantify the purified epPCR product using a Nanodrop spectrophotometer.
13. Digest 5 μg of the epPCR product with 10 U of each of the restriction enzymes in 1× restriction enzyme buffer and a total volume of 200 μl. If your total epPCR yield is lower than 5 μg, use the entire 30 μl elution from Subheading 3.2, **step 11** above. Incubate the reaction at 37 °C for 6–8 h or overnight. Heat inactivate the restriction enzymes, as recommended by the manufacturer. Split the reaction into two 100 μl aliquots and purify using two DNA Clean & Concentrator columns. Elute each of the two aliquots into 10 μl of sterile water prewarmed to 50 °C (*see* **Note 8**).
14. Pool both elution fractions and quantify using a Nanodrop spectrophotometer.

3.3 Preparation of a Test Library (See Note 16)

1. Prepare a test ligation containing 1× ligation buffer; 50 ng of vector DNA (*see* Subheading 3.1, **step 6**); T4 DNA ligase (1 U); a 3:1 molar ratio of digested epPCR insert DNA to vector DNA; and water to a final volume of 10 μl. Add the T4 DNA ligase last and mix gently. Prepare a control ligation as above but without the insert DNA.
2. Incubate the ligation reactions at 16 °C for 12–18 h.
3. Use 1 μl of each ligation reaction to transform 40 μl aliquots of electrocompetent *E. coli* cells, as described in Subheading 3.1, **step 9** (*see* **Notes 5** and **10**).
4. Spread aliquots (10 μl and 50 μl) of the two recovery cultures on LB-agar plates. Incubate the plates at 37 °C for 12–18 h.
5. Count the number of colonies on each plate. Use the results from the "vector only" plates to confirm the quality of the library (and compare to previous results from Subheading 3.1, **step 10**).
6. The number of colonies on the "vector + insert" plates also allows the size of the final, scaled-up library to be estimated. The final library is likely to be ~10^3 times larger than the total number of colonies on the "vector + insert" test plates (*see* **Note 17**).

3.4 Analysis of Library Composition

Amplify the randomly mutagenized gene inserts from 10 to 20 colonies by PCR. We have listed our routine protocol, as a guide. However, many variations are possible; the goal here is merely to generate enough of the amplified product for DNA sequencing (*see* **Note 18**). We typically set up 20 μl PCR reactions in thin-walled 0.2 ml tubes containing 8 μl sterile water, 10 μl 2× BioMix Red (Bioline), 1 μl forward primer (from 10 μM stock solution), and 1 μl reverse primer (from 10 μM stock solution).

1. Use 2 μl pipette tips (or sterile toothpicks) to pick 10–20 colonies at random from the "vector + insert" test plates (Subheading 3.3, **step 6**). Transfer each colony into one of the 20 μl PCR reactions prepared above.
2. Run an appropriate PCR program, such as the one listed in Subheading 3.2, **step 9**. Increasing the initial denaturation step to 3 min is recommended to ensure complete cell lysis.
3. Run a 2 μl aliquot of each PCR product on an agarose gel, to confirm successful amplification.
4. Purify the remainder of each PCR product using the DNA Clean & Concentrator Kit. Elute the purified DNA from each spin column in 20 μl sterile water.
5. Sequence each PCR product. Use the forward and/or reverse primers from the PCR as the sequencing primer(s), as necessary (*see* **Notes 18** and **19**).

Table 1
Mutational spectrum of a *cynT* epPCR library

Type(s) of mutations	Frequency	Proportion of total
Transitions		
A→G, T→C	28	22.6 %
G→A, C→T	28	22.6 %
Transversions		
A→T, T→A	34	27.4 %
A→C, T→G	9	7.3 %
G→C, C→G	4	3.2 %
G→T, C→A	18	14.5 %
Insertions and deletions		
Insertions	1	0.8 %
Deletions	2	1.6 %
Summary of bias		
Transitions/transversions	0.86	NA
AT→GC/GC→AT	1	NA
A→N, T→N	71	57.3 %
G→N, C→N	50	40.3 %
Mutation rate		
Mutations/kb	11.8	NA
Mutations/*cynT* gene[a]	7.8	NA

NA not applicable
[a]The cloned *cynT* insert was 657 bp

6. Align the sequence of each PCR product with the known sequence of the unmutated parental gene. Computer programs such as MacVector (Cary, NC, USA) or DNA Baser (Heracle BioSoft; Romania) are useful for this analysis.
7. Tabulate all of the point mutations in the sequenced samples. Also note any insertions, deletions, and stop codons that may have been introduced during the epPCR. The point mutations should be grouped by type (*see* **Note 20**).
8. Use the tabulated data to calculate the overall mutation rate and to assess biases in the mutation spectrum of the epPCR library. There are three key indicators of bias (*see* **Note 20**). The mutation rate and bias measures for an example *cynT* epPCR library are shown in Table 1.
9. The library analysis program PEDEL-AA [19], available online at http://guinevere.otago.ac.nz/stats.html, can be used to predict the utility of the final epPCR library (*see* **Note 21**).

3.5 Construction and Storage of Full-Sized Library

1. Prepare a full-sized library "vector + insert" ligation reaction (*see* **Note 22**). The reaction should contain 1× ligation buffer; 500 ng of vector DNA; 10 U T4 DNA ligase; digested insert

DNA (threefold molar excess over vector); and water to a final volume of 100 µl.

2. Prepare a "vector only" control ligation containing 1× ligation buffer; 100 ng of vector DNA; 2 U T4 DNA ligase and water to a final volume of 20 µl.
3. Incubate the ligation reactions at 16 °C for 12–18 h.
4. Transform a 40 µl aliquot of electrocompetent *E. coli* cells with 1 µl of the full-size library ligation (*see* Subheading 3.1, **step 9**). Transform a second 40 µl aliquot of cells with 1 µl of the "vector only" ligation, and a third 40 µl aliquot of cells with an appropriate plasmid to confirm the transformation efficiency of the cells (*see* Subheading 3.1, **step 10**).
5. Purify the remaining full library ligation reaction using the DNA Clean & Concentrator Kit or by drop dialysis [18]. Elute the purified DNA in 50 µl sterile water that has been preheated to 50 °C (*see* **Note 8**) and quantify using a Nanodrop spectrophotometer. Store at −20 °C.
6. After overnight incubation, count the number of colonies on each plate from Subheading 3.5, **step 4** to estimate the final size of the library and to verify that the "vector only" background is <1 % of the total library as predicted from the test ligation (*see* Subheading 3.1, **step 10**).
7. Use the plasmid control to confirm the transformation efficiency of your electrocompetent cells (*see* Subheading 3.1, **step 10**).
8. Use 10–50 ng aliquots of the purified ligation to transform at least ten aliquots (each 40 µl) of electrocompetent *E. coli* cells, as described in Subheading 3.1, **step 9** (*see* **Notes 5** and **23**). If any of the purified ligation reaction products remain, store them at −20 °C. Combine the recovery cultures in a sterile 15 ml tube. The total volume should be 5 ml (ten electroporations; 500 µl per recovery culture) (*see* **Note 24**).
9. Spread the cells on two large square plates (245 mm × 245 mm; spread ~2.5 ml of cell suspension per plate) or, alternatively, spread 250 µl aliquots on 20 standard agar plates.
10. Incubate the plates at 30 °C overnight to avoid the formation of confluent lawns.
11. The plates should each be covered in thousands (or millions) of small colonies. To recover the library from one of the plates, pipette 500 µl (small plates) or 4 ml (large plates) of LB medium (supplemented with the appropriate antibiotic) into the center of the plate. Use a glass spreader to scrape cells off the surface of the plate; pool them in one corner of the plate. Remove the resuspended cells with a P1000 pipettor, and transfer them to a sterile 50 ml tube. Pipette a second aliquot

of LB onto the same plate, and repeat the scraping step. All of the recovered cells should be pooled in the same 50 ml tube.

12. Break up cell clumps and mix well, by pipetting up and down repeatedly with a P1000 pipettor.
13. Pellet the cells by centrifugation at 3,000 × *g* for 15 min at 4 °C. Use a P1000 pipettor to remove the supernatant (*see* **Note 25**).
14. Resuspend the cell pellet in 1 ml of LB medium, plus antibiotic. The pellet is likely to be large; resuspension is likely to involve vigorous pipetting and/or gentle vortexing.
15. Knowing the cell density is likely to be useful for planning downstream screening/selection experiments. Mix 1 μl of the cell suspension with 999 μl of sterile water, and measure the OD_{600} (against a water blank). The final OD_{600} of the resuspended library is usually > 100, corresponding to a cell density of $>2.5 \times 10^{10}$ cells/ml (*see* **Note 26**).
16. Split the library into 100 μl aliquots and transfer each aliquot to a cryogenic vial.
17. Add 50 μl of sterile glycerol (50 % v/v) to each aliquot and mix well by pipetting.
18. Store the aliquots at −80 °C until you are ready to proceed with screening/selection to identify improved variants in the library.

4 Notes

1. A high-quality vector preparation is critical for constructing a large library. We have found that it is helpful to make a large-scale preparation of digested vector for use in short-term library construction, as this enables several libraries to be generated and tested within a short time frame, all using a vector preparation of known quality and quantity. However, the prepared vector should ideally be used within 1–2 weeks as we have noted a considerable drop in ligation efficiency after this point, even without repeated freeze/thaw cycles. The protocol can easily be adapted to prepare smaller amounts of vector if required. We find it useful to use a plasmid with a "stuffer" fragment in the cloning cassette. This should be a DNA fragment that is ligated into the vector using the same restriction sites that will be used for cloning the epPCR library. Ideally this fragment should contain a unique internal restriction site not present in your native gene (*see* **Note 2**). Excision of this stuffer fragment from the vector enables the easy monitoring of the restriction enzyme reaction, i.e., a small aliquot of the digestion reaction can be analyzed by gel electrophoresis to

confirm the complete digestion of the vector, determined by the visualization of the stuffer fragment and cut vector as two distinct bands on the agarose gel.

2. Use of a third restriction enzyme that cuts within the stuffer fragment minimizes the number of "vector only" clones in the final library. Where possible, it is ideal if this third restriction enzyme utilizes the same buffer as the other restriction enzymes used to prepare the vector for cloning. New England Biolabs provide an easy-to-use online tool for restriction analyses: http://tools.neb.com/NEBcutter2/. A diagnostic DNA electrophoresis gel would show three distinct DNA bands in this case; the digested vector and two DNA bands showing the separate fragments of the digested stuffer fragment.
3. We recommend the use of NEB High-Fidelity restriction enzymes. This range of enzymes has the same specificity as the native enzyme with reduced star activity and improved buffer compatibility (https://www.neb.com/products/restriction-endonucleases/hf-nicking-master-mix-time-saver-other/high-fidelity-restriction-enzymes/high-fidelity-restriction-endonucleases). The reduced star activity is particularly valuable because it allows for extended incubation times without risk of off-target digestion.
4. Primers should be noncomplementary and ideally have melting temperatures that are within 5 °C of each other. Melting temperatures can be estimated accurately using the OligoAnalyzer tool from Integrated DNA Technologies (http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/) or equivalent online programs, a comprehensive range of which can be found online: http://www.molbiol-tools.ca/PCR.htm. The optimal annealing temperature to use in a PCR is typically 3–5 °C cooler than the lowest primer melting temperature. We typically incorporate a restriction site for directional cloning into each of the primers, and lyophilized primers are resuspended in TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8.0) to a concentration of 100 μM. Working stocks (10 μM) are made by tenfold dilution of these master stocks, using sterile water.
5. The choice of *E. coli* strain will depend on the downstream selection or screen that is being employed. In general, the final size of the epPCR library will be directly proportional to the transformation efficiency of the host strain. Therefore, strains with high transformation efficiencies (such as *E. coli* DH5α-E) are preferable. We adapt the method of Sambrook and Russell [17] to prepare electrocompetent cells. Two of us (J.N.C. and D.F.A.) use the modified protocol that follows. In brief, inoculate 400 ml of pre-warmed LB medium in a 2 L flask with 10 ml of an overnight culture (grown from a single *E. coli* colony from a fresh agar plate) and incubate at 37 °C,

200 rpm, until the OD_{600} reaches 0.3. Rapidly cool the culture on ice for 30 min. Transfer the culture to eight prechilled 50 ml Falcon tubes and sediment the cells by centrifugation at 4 °C for 30 min at 1,000 × *g*. Discard the supernatant. Resuspend the cell pellets with 200 ml of prechilled sterile ddH_2O and repeat the centrifugation step above. Discard the supernatant. Resuspend the cell pellets with 100 ml of prechilled 10 % w/v glycerol, combine the cell suspensions into four 50 ml tubes (e.g., Falcon), and repeat the centrifugation step above. Discard the supernatant. Resuspend the cell pellets with 50 ml of prechilled 10 % w/v glycerol, transfer into two 50 ml tubes, and repeat the centrifugation step above. Carefully discard the supernatant, taking care to remove all remaining drops of 10 % glycerol. Resuspend the cell pellet in 200 µl of ice-cold GYT medium (10 % v/v glycerol, 0.125 % w/v yeast extract, 0.25 % w/v tryptone). Measure the OD_{600} of a 1:100 dilution of this resuspension. Calculate the cell concentration (*see* **Note 26**) and dilute the cell suspension to a concentration of ~2.5×10^{10} cells/ml with ice-cold GYT. Transfer 40 µl aliquots of the suspension into sterile microcentrifuge tubes prechilled to −80 °C. Store these aliquots at −80 °C until required.

6. We have found that standard microcentrifuge tubes are also suitable for short-term storage of competent cells and library glycerol stocks.
7. Two of us (P.H.M. and W.M.P.) always purify the digested vector backbone from an agarose gel at this stage, to eliminate the possibility that the unwanted insert fragment interferes with downstream ligation steps. On the other hand, the other two of us (J.N.C. and D.F.A.) have had good results with a faster protocol purifying the digested vector using the DNA Clean & Concentrator Kit (and omitting the gel purification step). If using gel purification, be aware that UV transillumination of ethidium bromide-stained DNA can induce damage, resulting in lower cloning and transformation efficiencies [20]. Even short exposures to UV (<60 s) can have dramatic and deleterious effects. Constructing a large epPCR library (>10^6 variants) requires the highest possible quality of DNA. Therefore, we recommend use of SYBR Safe stain and a blue-light transilluminator for preparation of the library vector. In our hands, this results in libraries that are 5–10 times larger than equivalent libraries prepared with ethidium bromide-stained DNA. If a SYBR Safe stain and blue-light transilluminator are not available, a small aliquot (200–300 ng) of the digested vector should be run an agarose gel (without ethidium) next to a DNA ladder and the remainder of the vector digestion should be run on the opposite side of the gel. After electrophoresis,

the lanes corresponding to the small aliquot of vector and the ladder should be separated from the rest of the agarose gel and stained in ethidium bromide (0.5 μg/ml) for 10 min. The location of the DNA fragment corresponding to the digested vector can then be visualized by UV transillumination and should be marked with a clean scalpel. The ethidium-stained agarose gel fragment can then be lined up next to the unstained portion of the agarose gel, and the location of the DNA fragment can be estimated by the scalpel lines and excised for purification.

8. We have routinely found that preheating the Elution Buffer or sterile ddH_2O to approximately 50 °C enhances the yield when using QiaQuick and DNA Clean & Concentrator columns, respectively.
9. We typically amplify the target gene (which will be later used for epPCR) by a standard PCR protocol to serve as a test fragment. Purify the PCR product and digest it overnight with the appropriate restriction enzymes in order to clone the fragment into the prepared vector. Purify and quantify the digested insert prior to ligation.
10. This step gives the researcher an opportunity to test the electrocompetent cell stocks and ensure they are highly efficient and suitable for use in library preparation. Electrocompetent cell aliquots should be discarded, and a fresh batch of electrocompetent cells should be prepared if the efficiency is found to be too low. A good batch of *E. coli* cells should yield >10^9 colonies per microgram of the supercoiled control plasmid used in the transformation. As an alternative to homemade cells, we have also obtained excellent transformation efficiencies with premade *E. cloni* 10G Elite electrocompetent cells (Lucigen, Middleton, WI).
11. This step gives the researcher an opportunity to test and optimize PCR conditions prior to Mutazyme amplification. In our experience, the annealing temperatures that are used for Phusion amplification work well for Mutazyme amplification. We have found that use of a PCR purified, DpnI-treated template for Mutazyme amplification reduces the risk of PCR contamination, especially when aiming for a low mutagenesis rate with correspondingly high concentrations of template. If PCR conditions have previously been optimized and researchers are aiming for high mutagenesis rates (with correspondingly low concentrations of template), a plasmid-based template may be used for the epPCR and researchers may directly proceed to Subheading 3.2, **step 7**.
12. The overall mutation frequency depends on the error rate of the polymerase and also the number of times that each template is duplicated in the reaction. If the initial amount of template is

high, it will undergo little duplication in the epPCR. On the other hand, a low amount of template will result in a greater number of duplications and more mutations will be introduced. Low mutation rates may also be achieved by decreasing the number of cycles in the epPCR protocol. This is discussed further in the GeneMorph II Random Mutagenesis Kit manual, available for download from http://www.chem.agilent.com/library/usermanuals/Public/200550.pdf.

We typically aim for a medium-to-high mutation rate, because this generates libraries that contain minimal numbers of "wasted" variants (i.e., non-mutated copies of the template or multiple copies of variants with any one point mutation). This strategy is discussed in more detail elsewhere [10, 19]. Agilent recommends 500–1,000 ng of template DNA for a low mutation rate (0–4.5 mutations/kb); 100–500 ng of template for a medium mutation rate (4.5–9 mutations/kb), and 0.1–100 ng of template for a high mutation rate (9–16 mutations/kb). If you have chosen to use a plasmid template rather than a purified PCR product, it is important to note that the amount of template is not the amount of purified plasmid DNA (*see* GeneMorph II Random Mutagenesis Kit Manual). In general, we find that the resulting mutation rates tend to be slightly lower than those expected from these calculations and therefore use the lowest end of the recommended template concentrations (e.g., we would use 0.1 ng of template for a high mutation rate).

13. We typically use 1/5 of the volume of the experimental reactions for negative controls in order to conserve resources.
14. We routinely use extension times that are calculated at a rate of 1 min/kb. For example, a 30 s extension time is used for a 500 bp product, and a 90 s extension time is used for a 1,500 bp product.
15. Calculating the PCR efficiency parameter allows robust statistical analysis of library composition (*see* Subheading 3.4, **step 9**). When the total product yield and the amount of starting template are known, the number of doublings in the PCR, d, can be calculated:

$$d = \frac{\log\left(\text{Product} / \text{Template}\right)}{\log 2}.$$

The PCR efficiency (i.e., the probability that any particular sequence is duplicated in any one cycle of the PCR, eff) is then given by

$$\text{eff} = 2^{(d/n)} - 1,$$

where n is the number of PCR cycles ($n = 30$ in our protocol). An online tool for calculating eff, given d and n, can be found at http://guinevere.otago.ac.nz/cgi-bin/aef/PCReff.pl.

16. Before scaling up to a full-sized library, we find it useful and expedient to construct a test library. This allows the epPCR mutation spectrum to be determined.
17. The size of the final scaled-up library is a rough estimate based on (a) tenfold scale up of the ligation reaction; (b) transformation of ten aliquots of electrocompetent cells; and (c) spreading the entire recovery culture (instead of 10–50 μl aliquots).
18. We recommend using vector primers (e.g., for pET-based vectors, we would use the universal T7 terminator and promoter primers) to improve the quality of sequence at the beginning and end of the gene.
19. Clones from the test library (*see* Subheading 3.3) should be sequenced in order to analyze the mutation rate and the spectrum of mutations that arose in the epPCR. As discussed (*see* **Notes 5** and **6**), some control over the mutation rate is possible. However, in our experience, there is considerable experiment-to-experiment variation in the outcomes of the epPCR process (*see* Subheading 3.2). Therefore, we recommend conducting the analyses described in Subheading 3.4, to avoid wasting time and resources on a scaled-up library that contains little molecular diversity.
20. A library with an unbiased spectrum of mutations will be maximally diverse; that is, it will have the lowest probability of duplicated variants. Therefore, it is more likely to contain at least one improved variant [10]. One indicator of bias is the ratio of transitions (i.e., purine-to-purine and pyrimidine-to-pyrimidine mutations) to transversions (purine-to-pyrimidine and pyrimidine-to-purine mutations). There are four possible transitions and eight possible transversions (listed in Table 1). Therefore, a completely unbiased error-prone polymerase would generate libraries with transition/transversion (Ts/Tv) ratios of 0.5. Provided that the GC content of the gene is ~50 %, the ratio of AT → GC transitions to GC → AT transitions (i.e., AT → GC/GC → AT) in an unbiased epPCR library should be 1. Similarly, the number of mutations at A:T base pairs (A → N, T → N) should also be the same as the number of mutations at G:C base pairs (G → N, C → N). The effects of mutational bias on overall library composition can be assessed by altering the input parameters for PEDEL-AA analysis (*see* **Note 21**).
21. PEDEL-AA has an easy-to-use web interface and takes the following parameters as its inputs:

(a) The sequence of the gene that was randomized

(b) The estimated size of the scaled-up library (*see* Subheading 3.3, **step 6**)

(c) The nucleotide mutation matrix (*see* Subheading 3.4, **step 8**)

(d) The mean number of mutations per gene in the library (*see* Subheading 3.4, **step 9** and Table 1 for an example)

(e) The number of cycles in the epPCR (*see* Subheading 3.2, **step 9**)

(f) The PCR efficiency parameter for the epPCR (*see* **Note 15**)

(g) The mean number of insertions per gene in the library (*see* Subheading 3.4, **step 8**)

(h) The mean number of deletions per gene in the library (*see* Subheading 3.4, **step 8**)

The program output comprises a variety of statistics about the protein variants that are encoded by the epPCR library. Together, these data allow an informed decision to be made about whether to scale up the library (or whether to start over, with different epPCR and ligation conditions).

22. When constructing the full-sized library, the focus should be on scaling everything up by as much as possible. All of the remaining epPCR insert (Subheading 3.2, **step 14**) should be used in the scaled-up ligation, and as many aliquots of electrocompetent *E. coli* as possible should be transformed with the ligated products. The protocol that we describe is a typical example from our laboratory.

23. It is important to keep the DNA concentration low at this point to avoid transformants containing more than one plasmid [21].

24. If your selection or screening method permits, you may immediately proceed to screening/selecting variants from your library:

(a) Add 5 ml of sterile 80 % v/v glycerol to the 5 ml of recovered cells and mix the cells briefly, by inverting the tube 2–3 times. Spread 2 μl, 10 μl, and 50 μl aliquots on LB-agar plates (diluting as necessary to obtain a spreadable volume). Split the remainder of the cells into 1 ml aliquots and store them at −80 °C.

(b) Incubate all of the dilution and control plates at 37 °C for 16 h.

(c) Count the colonies on the plates corresponding to the 2 μl, 10 μl, and 50 μl aliquots and determine the most appropriate dilution to use for your screen/selection (i.e., the dilution that gives you a high number of single colonies but not a confluent lawn).

(d) Thaw separate glycerol aliquots, dilute as necessary, and plate cells on selective media. Electroporate additional 10–50 ng aliquots (*see* Subheading 3.5, **step 8**) of the purified ligation products to achieve full library coverage.

25. If you wish to screen the library in a different *E. coli* strain to the one you have utilized for cloning the library, resuspend the cell pellet from Subheading 3.5, **step 14** in 4.5 ml of LB media, split the suspension into three 1.5 ml aliquots, and purify the library using three plasmid miniprep columns (e.g., Qiagen; Valencia, CA, USA). Quantify using a Nanodrop spectrophotometer and store at −20 °C. Electroporate 10 ng aliquots of the purified library into your preferred screening or selection strain (*see* **Note 23**).
26. For the *E. coli* strain DH5α, we find that $OD_{600} = 1$ corresponds to ~2.5×10^8 cells/ml.

References

1. Lutz S (2010) Beyond directed evolution—semi-rational protein engineering and design. Curr Opin Biotechnol 21(6):734–743
2. 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(4):378–384
3. Lutz S, Patrick WM (2004) Novel methods for directed evolution of enzymes: quality, not quantity. Curr Opin Biotechnol 15(4): 291–297
4. Barak Y, Ackerley DF, Dodge CJ, Banwari L, Alex C, Francis AJ, Matin A (2006) Analysis of novel soluble chromate and uranyl reductases and generation of an improved enzyme by directed evolution. Appl Environ Microbiol 72(11):7074–7082
5. Morley KL, Kazlauskas RJ (2005) Improving enzyme properties: when are closer mutations better? Trends Biotechnol 23(5):231–237
6. 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(2): 696–700
7. Horsman GP, Liu AM, Henke E, Bornscheuer UT, Kazlauskas RJ (2003) Mutations in distant residues moderately increase the enantioselectivity of *Pseudomonas fluorescens* esterase towards methyl 3bromo-2-methylpropanoate and ethyl-3-phenylbutyrate. Chemistry 9(9): 1933–1939
8. Kumar S, Chen CS, Waxman DJ, Halpert JR (2005) Directed evolution of mammalian cytochrome P450 2B1: mutations outside of the active site enhance the metabolism of several substrates, including the anticancer prodrugs cyclophosphamide and ifosfamide. J Biol Chem 280(20):19569–19575
9. Wong JT (1980) Role of minimization of chemical distances between amino acids in the evolution of the genetic code. Proc Natl Acad Sci U S A 77(2):1083–1086
10. Patrick WM, Firth AE, Blackburn JM (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. Protein Eng 16:451–457
11. 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
12. Cadwell RC, Joyce GF (1992) Randomization of genes by PCR mutagenesis. PCR Methods Appl 2(1):28–33
13. Fromant M, Blanquet S, Plateau P (1995) Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. Anal Biochem 224(1):347–353
14. Lin-Goerke JL, Robbins DJ, Burczak JD (1997) PCR-based random mutagenesis using manganese and reduced dNTP concentration. Biotechniques 23(3):409–412
15. Rasila TS, Pajunen MI, Savilahti H (2009) Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols,

Escherichia coli mutator strain, and hydroxylamine treatment. Anal Biochem 338(1): 71–80

16. Hanson-Manful P, Patrick WM (2013) Construction and analysis of randomized protein-encoding libraries using error-prone PCR. Methods Mol Biol 996:251–267
17. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY
18. Saraswat M, Grand RS, Patrick WM (2013) Desalting DNA by drop dialysis increases library size upon transformation. Biosci Biotechnol Biochem 77(2):402–404
19. Firth AE, Patrick WM (2008) GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. Nucleic Acids Res 36:W281–W285
20. Hartman PS (1991) Transillumination can profoundly reduce transformation frequencies. Biotechniques 11:747–748
21. Goldsmith M, Kiss C, Bradbury AR, Tawfik DS (2007) Avoiding and controlling double transformation artifacts. Protein Eng Des Sel 20(7): 315–318