Chapter 15

Construction and Analysis of Randomized Protein-Encoding Libraries Using Error-Prone PCR

Paulina Hanson-Manful and Wayne M. Patrick

Abstract

In contrast to site-directed mutagenesis and rational design, directed evolution harnesses Darwinian principles to identify proteins with new or improved properties. The critical first steps in a directed evolution experiment are as follows: (a) to introduce random diversity into the gene of interest and (b) to capture that diversity by cloning the resulting population of molecules into a suitable expression vector, en bloc. Error-prone PCR (epPCR) is a common method for introducing random mutations into a gene. In this chapter, we describe detailed protocols for epPCR and for the construction of large, maximally diverse libraries of cloned variants. We also describe the utility of an online program, PEDEL-AA, for analyzing the compositions of epPCR libraries. The methods described here were used to construct several libraries in our laboratory. A side-by-side comparison of the results is used to show that, ultimately, epPCR is a highly stochastic process.

Key words Directed evolution, Random mutagenesis, Error-prone PCR, GeneMorph II, Mutazyme II DNA polymerase, Library, Mutation spectrum, Mutational bias, PEDEL-AA

1 Introduction

In the past two decades, directed evolution has emerged as a powerful method for altering the properties of proteins. It involves mimicking the process of Darwinian evolution on a single gene, on a laboratory timescale. In the first step of a directed evolution experiment, mutations are introduced at random into copies of the target gene, resulting in a large and diverse library of variants (typically 10^3–10^9 clones). The members of this library are subjected to a suitably high-throughput screen or genetic selection, in order to identify rare variants with improvements in the desired property. Multiple rounds of mutagenesis and screening/selection enable the accumulation of beneficial mutations that may have been impossible to predict, a priori. Directed evolution has been adopted widely for tailoring industrially relevant biocatalysts with improvements in properties such as substrate specificity, enantioselectivity, and

Juliet A. Gerrard (ed.), *Protein Nanotechnology: Protocols, Instrumentation, and Applications*, Methods in Molecular Biology, vol. 996, DOI 10.1007/978-1-62703-354-1_15,

thermostability (1, 2). It has also been used to address fundamental questions about protein structure, function, and evolution (3).

Many methods have been developed for introducing molecular diversity into parent sequences (4, 5). One of these methods, error-prone polymerase chain reaction (epPCR), remains a particularly common means of generating random mutations at any position in the target gene. Conceptually, epPCR is simple: the target gene is amplified exponentially, under conditions in which the fidelity of the polymerase is reduced. Cloning the randomly mutagenized PCR product into an appropriate expression vector yields a library of variants that can be used in downstream screening or selection. While outside the scope of this chapter, it is worth noting that these downstream steps are also critically important for the success of any directed evolution experiment (6).

The original—and still the cheapest—way to carry out an epPCR is to reduce the fidelity of *Taq* DNA polymerase, by adding Mn^{2+} ions and unbalanced ratios of dNTPs (7, 8). Detailed protocols for using *Taq* polymerase to construct epPCR libraries have been described previously (9, 10). However, *Taq*-generated epPCR libraries suffer from biases in the types of mutations that are observed; in particular, mutations at A:T base pairs are massively overrepresented (11).

We have argued that an unbiased and maximally diverse library has the highest probability of containing variants with the desired function (12). The GeneMorph II Random Mutagenesis Kit from Agilent Technologies (http://tinyurl.com/3mb6x66) is designed specifically for the construction of unbiased epPCR libraries. In the protocols below, we describe the use of the GeneMorph II kit, as well as the subsequent steps that are required to construct a library. We illustrate the protocols with data from a library we have recently constructed. We also describe the analyses that we routinely perform to assess the compositions of our libraries. Our target for randomization was the *Escherichia coli cynT* gene, which was identified in a previous experiment because of its contribution to antibiotic resistance when it was over-expressed (13). We conclude the chapter by comparing the *cynT* epPCR library with two other libraries and with the results that are predicted by the manufacturer of the GeneMorph kit (Agilent). This analysis highlights the stochastic nature of epPCR.

2 Materials

2.1 Error-Prone PCR

1. Plasmid containing the gene that is to be amplified by epPCR.
2. Spectrophotometer and cuvettes for measuring DNA concentration, e.g., an Eppendorf Biophotometer and UVettes.
3. Oligonucleotide primers for the epPCR amplification (see Note 1).

4. GeneMorph II Random Mutagenesis Kit (Agilent). The kit contains Mutazyme II DNA polymerase (2.5 U/μL), 10× Mutazyme II reaction buffer, and a dNTP mix (10 mM each dNTP).
5. Thermocycler with a heated lid.
6. Agarose gels, stained with ethidium bromide at 0.5 μg/mL.
7. DNA ladder with bands that contain known amounts of DNA.
8. Apparatus for agarose gel electrophoresis.
9. QiaQuick PCR Purification Kit (Qiagen). Equivalent kits from other manufacturers are also suitable.

2.2 Vector and Insert Preparation

1. Protein expression vector, into which the epPCR product will be cloned (see Note 2).
2. QiaPrep Spin Miniprep Kit (Qiagen). Equivalent kits from other manufacturers are also suitable.
3. Restriction enzyme(s) that facilitate directional, sticky-ended cloning (e.g., SfiI from New England Biolabs).
4. Restriction enzyme that cuts within the expression vector's stuffer fragment (see Note 3).
5. Restriction enzyme DpnI (New England Biolabs).
6. Agarose gels stained with 1× SYBR Safe DNA gel stain (Invitrogen).
7. Safe Imager 2.0 Blue-Light Transilluminator (Invitrogen).
8. Clean razor blades for excising bands from gels.
9. MinElute Gel Extraction Kit (Qiagen).

2.3 Preparation of a Test Library

1. T4 DNA ligase and ligation buffer. We obtain comparable results with the T4 DNA ligases from New England Biolabs, Fermentas, and Enzymatics Inc.
2. Aliquots (50 μL) of electrocompetent *E. coli* cells (see Note 4).
3. Gene Pulser electroporation cuvettes (BioRad, 0.2 cm electrode gap).
4. Gene Pulser electroporation unit with Pulse Controller (BioRad).
5. Sterile SOC medium: 20 g/L tryptone; 5 g/L yeast extract; 10 mM NaCl; 2.5 mM KCl; 20 mM glucose.
6. 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 BioRad).
2. Primers for amplifying cloned inserts from the epPCR library (see Note 1).

3. Reagents for a standard PCR screen. While there are many alternate (and equally good) suppliers, we routinely use 5× Green GoTaq Reaction Buffer (Promega), i-Taq DNA polymerase (iNtRON Biotechnology), and the dNTP mix that is supplied with the polymerase (which contains 2.5 mM of each dNTP).
4. Ethidium bromide-stained agarose gels and electrophoresis apparatus.
5. QiaQuick PCR Purification Kit (Qiagen), or equivalent.

2.5 Construction and Storage of the Full-Sized Library

1. T4 DNA ligase and ligation buffer, as listed in Subheading 2.3.
2. QiaQuick PCR Purification Kit (Qiagen), or equivalent.
3. Standard-sized LB-agar plates (circular, 85–90 mm diameter) containing the correct antibiotic for selecting plasmid-containing cells.
4. Two square bioassay dishes (245 mm × 245 mm) from Corning or Nunc. Each bioassay dish holds 200 mL of LB agar, supplemented with the appropriate antibiotic for maintaining the library vector.
5. A fresh batch of electrocompetent *E. coli* cells (see Note 4).
6. Electroporation cuvettes and apparatus, as described in Subheading 2.3.
7. SOC medium, as described in Subheading 2.3.
8. Supercoiled pUC19 control plasmid (10 pg/μL; Invitrogen).
9. Sterile 50 mL tube (Falcon or similar).
10. LB medium supplemented with the appropriate antibiotic (~20 mL, total).
11. Refrigerated centrifuge with a rotor that takes the 50 mL tube listed above (item 9).
12. Spectrophotometer and cuvettes for measuring cell density (OD_{600}), e.g., an Eppendorf Biophotometer and UVettes.
13. Sterile glycerol (50% v/v).
14. Cryogenic vials, suitable for storage at −80°C.

3 Methods

3.1 Error-Prone PCR

1. Prepare purified plasmid DNA, containing the gene that is to be mutagenized. Measure its concentration spectrophotometrically.
2. The amount of template used in the epPCR affects the mutation rate (see Note 5). Calculate the amount of plasmid DNA that is required for the desired mutation rate (see Notes 6 and 7).

3. Prepare the epPCR reagents in a thin-walled, 0.2 mL tube:

x μL	Plasmid DNA template (see Note 6)
y μL	Water to a total volume of 50 μL
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)

4. Mix the sample and place the tube in the thermocycler.
5. Run the epPCR program:

Step 1:	1 min	95°C
Step 2:	20 s	94°C
Step 3:	20 s	Annealing temperature for primers (see Note 1)
Step 4:	1 min	72°C (for a ~1 kb gene; see Note 8)
Step 5:	Repeat steps 2–4 for an additional 29 cycles	
Step 6:	2 min	72°C
Step 7:	Hold	4°C (for product storage, if necessary)

6. Run 2 μL of the product on an ethidium bromide-stained agarose gel, alongside a DNA ladder. Determine the total yield of the epPCR product by comparing the intensity of the epPCR sample with the intensity of the bands in the ladder. The total yield of epPCR product is required to calculate the PCR efficiency (see Note 9).
7. Purify the remainder of the epPCR product using the PCR Purification Kit. Elute the purified DNA from the spin column in 30 μL elution buffer (EB).
8. Estimate the concentration of the purified sample by running 1 μL on an agarose gel, alongside a suitable DNA ladder.

3.2 Vector and Insert Preparation

1. Prepare 3–4 μg of the plasmid that will be used for cloning and expression of the epPCR library (see Note 2). In the example discussed below, we used vector pCA24N (14), which was purified from a saturated overnight culture using the QiaPrep Spin Miniprep Kit (Qiagen).

2. Digest the vector (~3 μg) and epPCR product (~1 μg) to completion, with restriction enzyme(s) that introduce sticky ends and facilitate directional cloning. We typically digest pCA24N and the epPCR product with 20 U of SfiI, in total reaction volumes of 30 μL. Under these conditions, incubating for 5 h at the enzyme's optimal temperature (50°C) is generally sufficient for complete digestion.
3. For additional improvements in the quality of the final library, add fresh restriction enzymes, as follows:
   (a) Vector preparation—add 10 U of an enzyme that cuts within the stuffer fragment (see Note 3).
   (b) Insert preparation—add 10 U of DpnI, to eliminate any of the methylated, unmutated template that may have carried over from the epPCR.

   Incubate the reactions for a further 2 h at 37°C, then heat inactivate the enzymes (where possible), according to the manufacturer's guidelines.
4. Run the two reactions on separate agarose gels, with agarose concentrations that are appropriate for the fragments being resolved (e.g., 0.8% agarose for the vector and 1.2% agarose for the insert).
5. Excise the bands that correspond to the digested vector and insert. We strongly recommend the use of a blue-light transilluminator and a compatible stain (SYBR Safe), rather than ethidium bromide and a UV transilluminator, for this step. See Note 10.
6. Purify the vector and insert DNA from the excised gel bands. We use the MinElute Gel Extraction Kit (Qiagen) and elute the DNA from each spin column in 12 μL EB.
7. Determine the concentrations of the purified vector and insert DNA by running 2 μL aliquots of each on an ethidium bromide-stained agarose gel, as described above (Subheading 3.1, step 6).
8. Store the purified DNA at −20°C, as necessary.

3.3 Preparation of a Test Library (See Note 11)

1. Prepare two ligation reactions: one with vector DNA only and one with the vector and a threefold molar excess of the insert DNA. Each reaction should contain the following: 1× ligation buffer; 50 ng of vector DNA (see Subheading 3.2, step 6); T4 DNA ligase (1 U); plus or minus the insert DNA; and water to a final volume of 10 μL. Add the T4 DNA ligase last and mix gently.
2. Incubate the ligation reactions at 16°C for 16 h.
3. Use a 1 μL aliquot of each ligation reaction to transform 50 μL aliquots of *E. coli*, by electroporation. See Note 12.
4. Immediately after electroporation, add 500 μL of SOC medium to the cuvette and transfer the cells to a sterilized, capped test

tube or a 15 mL tube (Falcon or similar). Allow the cells to recover by incubating them at 37°C, with shaking, for 1 h.

5. Store the remaining 9 μL of each ligation reaction (leftover from step 3, above) at −20°C.
6. Spread aliquots (10 and 50 μL) of the two recovery cultures on LB-agar plates. Incubate the plates at 37°C for 12–16 h.
7. Count the number of colonies on each plate. Use the results from the "vector only" plates to calculate the fraction of the library (as represented on the "vector + insert" plates) that contains recircularized vector. This background must be minimized, to avoid wasting time on futile library screens. If the "vector only" background is >1% of the total library, we recommend preparing a fresh batch of the vector (Subheading 3.2, above), and lengthening the incubation time with each restriction enzyme.
8. 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 13).

3.4 Analysis of Library Composition (See Note 14)

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 7).
2. Transfer each colony into a thin-walled, 0.2 mL tube containing 5 μL of sterile water.
3. Lyse the cells by incubating the tubes at 95°C for 5 min, in a thermocycler.
4. Amplify the randomly mutagenized gene inserts from each colony by PCR. We have listed our routine protocol, for guidance. However, many variations are possible; the goal here is merely to generate enough of the amplified product for DNA sequencing. We typically set up 25 μL PCRs in thin-walled 0.2 mL tubes, as follows:

14.75 μL	Water
5 μL	5× Green GoTaq buffer
2 μL	dNTP mix (gives 200 μM of each dNTP, final concentration)
1 μl	Forward primer (from 10 μM stock solution)
1 μL	Reverse primer (from 10 μM stock solution)
0.25 μL	*Taq* DNA polymerase (1.25 U)
1 μL	Cell lysate (from step 2, above)

Table 1
Matrix of point mutations identified in 16 *cynT* variants

		Mutation To			
		T	**C**	**A**	**G**
Mutation From	**T**	–	15	16	6
	C	10	–	8	0
	A	18	3	–	13
	G	10	4	18	–

4. Mix each sample and place the tubes in the thermocycler.
5. Run an appropriate PCR program, such as the one listed in Subheading 3.1, step 5.
6. Run a 2 μL aliquot of each PCR product on an agarose gel, to confirm successful amplification.
7. Purify the remainder of each PCR product using the QiaQuick PCR Purification Kit (or equivalent). Elute the purified DNA from each spin column in 30 μL EB.
8. Sequence each PCR product. Use the forward and/or reverse primers from the PCR as the sequencing primer(s), as necessary.
9. Align the sequence of each PCR product with the known sequence of the unmutated parental gene. Computer programs such as MacVector are useful for this analysis.
10. Tabulate all of the point mutations in the sequenced samples. Also note any insertions or deletions that may have arisen during the epPCR. The point mutations should be grouped by type. For example, we randomized the *E. coli cynT* gene and sequenced 16 clones from the resulting test library. The 121 point mutations that we identified in the 16 variants are summarized in Table 1. There were also two deletions and one insertion in the data set; in total, the sequencing revealed 124 mutations.
11. 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 15): (a) the ratio of transition (Ts) to transversion (Tv) mutations; (b) the ratio of AT→GC transitions to GC→AT transitions; and (c) the frequency of mutations at A:T base pairs, to mutations at G:C base pairs. The mutation rate and bias measures for our *cynT* epPCR library are shown in Table 2.

Table 2
Mutational spectrum of the *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[a]
AT→GC/GC→AT	1	NA[a]
A→N, T→N	71	57.3%
G→N, C→N	50	40.3%
Mutation rate		
Mutations per kb	11.8	NA[a]
Mutations per *cynT* gene[b]	7.8	NA[a]

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

12. The library analysis program PEDEL-AA (15), available online at http://guinevere.otago.ac.nz/stats.html, should now be used to predict the utility of the final epPCR library. 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 (Subheading 3.3, step 8)
    (c) The nucleotide mutation matrix (Subheading 3.4, step 10; see Table 1 for an example)
    (d) The mean number of mutations per gene in the library (Subheading 3.4, step 11; see Table 2 for an example)

Table 3
PEDEL-AA outputs for the *cynT* epPCR library

Property	Estimate
Total library size	1.4×10^7
Number of variants with no insertions, deletions, or stop codons	9.0×10^6
Mean number of amino acid substitutions per variant	5.5
Unmutated (wild-type) sequences (% of total library)	3.0%
Number of distinct, full-length proteins in the library[a]	7.4×10^6

[a]Calculated using the PCR efficiency parameter. PEDEL-AA also calculates a less accurate estimate the number of distinct, full-length proteins in the library by using the simplifying assumption of Poisson statistics

(e) The number of cycles in the epPCR (Subheading 3.1, step 3)

(f) The PCR efficiency parameter for the epPCR (see Note 9)

(g) The mean number of insertions per gene in the library (Subheading 3.4, step 10)

(h) The mean number of deletions per gene in the library (Subheading 3.4, step 10)

The program outputs a variety of statistics about the protein variants that are encoded by the epPCR library. A selection of these statistics, calculated for our *cynT* library, is shown in Table 3. Together, the data in Tables 2 and 3 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).

3.5 Construction and Storage of the Full-Sized Library

1. Prepare "vector only" and "vector + insert" ligation reactions that are tenfold larger than those described in Subheading 3.3, step 1 (see Note 16). Each reaction should contain the following: 1× ligation buffer; 500 ng of vector DNA; T4 DNA ligase (10 U); plus or minus insert DNA (3-fold molar excess over vector); and water to a final volume of 100 μL.
2. Incubate the ligation reactions at 16°C for 16 h.
3. Add the remaining 9 μL of each test ligation (Subheading 3.3, step 5) to the scaled-up "vector only" and "vector + insert" ligation reactions.
4. Purify the products from each ligation reaction using the QiaQuick PCR Purification Kit (or equivalent). Elute the purified DNA from each spin column in 42 μL EB.
5. Prepare the LB-agar plates on which the transformed cells of the library will be spread. For each library, we typically use two 245 mm × 245 mm square bioassay dishes (see Note 17).

6. Prepare a fresh batch of electrocompetent *E. coli* cells (see Note 4). In our hands, the transformation efficiencies of freshly prepared cells are four to fivefold higher than cells that have undergone a freeze/thaw cycle.
7. Add 3 μL aliquots of the "vector + insert" ligation to 14 × 50 μL aliquots of electrocompetent *E. coli* cells. Transform each aliquot, and recover the transformed cells, as described previously (Subheading 3.3, steps 3 and 4).
8. Transform a single 50 μL aliquot of cells with 3 μL of the "vector only" ligation.
9. Transform one more 50 μL aliquot of the electrocompetent *E. coli* with an appropriate plasmid for determining the transformation efficiency of the cells. We routinely use 10 pg of supercoiled pUC19.
10. Pool all of the cells that were transformed with the "vector + insert" ligation, in a sterile 15 mL tube. The total volume should be 7.7 mL (14 electroporations; 550 μL per recovery culture).
11. Mix the cells briefly, by inverting the tube 2–3 times.
12. Spread 1, 5, and 25 μL aliquots on regular LB-agar plates (diluting as necessary to obtain a spreadable volume).
13. Spread the remainder of the library on the two large plates (see step 5, above); ~3.85 mL per plate.
14. Spread 1, 5, and 25 μL aliquots of the "vector only" control on regular LB-agar plates.
15. Spread suitable aliquots (typically 2 and 10 μL) of the cells transformed with the pUC19 control on LB-agar plates that contain ampicillin (100 μg/mL).
16. Incubate all of the dilution and control plates at 37°C for 16 h. Incubate the two large library plates at 30°C, to avoid the formation of a confluent lawn.
17. Count the number of colonies on each plate, except for the large library plates (which should be covered in dense lawns of small colonies).
18. Use the pUC19 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. A good batch of *E. coli* cells should yield $>10^9$ colonies per microgram of pUC19 used in the transformation.
19. The regular LB-agar plates with aliquots of the library and the "vector only" control should be used to estimate the final size of the library, and to verify that the "vector only" background is <1% of the total library (see Subheading 3.3, step 7).

20. The large library plates should each be covered in thousands (or millions) of small colonies. To recover the library from one of the plates, pipette 4 mL of LB medium (supplemented with the appropriate antibiotic) into the center of the plate.
21. Use a glass spreader to scrape cells off the surface of the plate; pool them in one corner of the plate.
22. Remove the resuspended cells with a P1000 pipettor, and transfer them to a sterile 50 mL tube.
23. Pipette a second 4 mL aliquot of LB onto the same plate, and repeat the scraping step.
24. Use another 2 × 4 mL aliquots of LB to recover the cells from the second library plate.
25. All of the recovered cells should be pooled in the same 50 mL tube.
26. Break up cell clumps and mix well, by pipetting up and down repeatedly with a P1000 pipettor.
27. Pellet the cells by centrifugation at 3,000 × *g*, 4°C, for 15 min. Use a P1000 pipettor to remove the supernatant.
28. 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.
29. 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 per milliliter (see Note 18).
30. Split the library into 100 μL aliquots and transfer each aliquot to a cryogenic vial.
31. Add 50 μL of sterile glycerol (50% v/v) to each aliquot and mix well by pipetting.
32. Store the aliquots at −80°C until you are ready to proceed with screening/selection to identify improved variants in the library.

3.6 Summarizing Stochasticity in epPCR Library Construction

1. Tables 1 and 2 show that the *cynT* epPCR library, constructed with the GeneMorph II kit, is not free of mutational bias. For example, G→C and C→G mutations occur much less frequently than A→T and T→A mutations (Table 2)—indicating that the Mutazyme II polymerase retains some of the bias of *Taq* polymerase. For reference, we have included a side-by-side comparison of the *cynT* data with two other epPCR libraries that were constructed using the same protocol, together with the guidelines from the manufacturer (Table 4). All three randomized genes (*cynT*, *ydfW*, and *yeaD*) are latent contributors to antibiotic resistance (13).

Table 4
Comparison of three epPCR libraries with Agilent's product guidelines

	cynT	*ydfW*	*yeaD*	[Agilent][a]
epPCR details				
Gene length	657 bp	147 bp	882 bp	NA[b]
Plasmid in the epPCR	90 ng	90 ng	90 ng	NA[b]
Amount of target DNA[c]	15 ng	7 ng	18 ng	0.1–1,000 ng
Library bias indicators				
Transitions/transversions	0.86	0.56	0.56	0.90
AT→GC/GC→AT	1	1.50	1.17	0.60
A→N, T→N	57.3%	71.4%	61.1%	50.7%
G→N, C→N	40.3%	28.6%	38.9%	43.8%
Mutation rates				
Mutations per kb	11.8	9.5	4.1	0–16
Mutations per gene	7.8	1.4	3.6	NA[b]

[a]Source: Tables 1 and 2 of the GeneMorph II Random Mutagenesis Kit manual, available for download from http://tinyurl.com/3mb6x66
[b]NA: not applicable
[c]Calculated as described in Note 6

2. The data in Table 15.4 highlight the stochasticity that is inherent in epPCR. No two libraries are identical. While the GeneMorph II kit introduces less mutational bias than other epPCR methods, variation in mutational spectra is still to be expected. Mutation rates are also variable, and the recommendations given here (see Note 6) should be considered a rough guide only.
3. Bearing these facts in mind, an epPCR practitioner should be prepared to construct several libraries, combining the analyses that we have described here with considerable trial and error!

4 Notes

1. Primers should be noncomplementary and should 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/. The optimal annealing temperature to use in a PCR is typically 3–5°C cooler than the lowest primer melting

temperature. We routinely resuspend lyophilized primers in TE buffer (10 mM Tris, 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.

2. A high-quality vector preparation is critical for constructing a large library. We find it useful to use a plasmid with a stuffer fragment in the cloning cassette. Excision of this stuffer fragment allows the progress of the restriction digestion to be monitored. It also ensures that the doubly digested vector (with stuffer fragment removed) can be resolved from undigested and singly digested material on an agarose gel.
3. Using a restriction enzyme that cuts within the stuffer fragment minimizes the number of "vector only" clones in the final library. We routinely use BglII (New England Biolabs) for this purpose.
4. 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 is 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 prepare electrocompetent cells according to the method of Hanahan (16).
5. 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 few duplications 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. This is discussed further in the GeneMorph II Random Mutagenesis Kit manual, available for download from http://tinyurl.com/3mb6x66.
6. 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). Note that the amount of template is not the amount of purified plasmid DNA. Instead, it is the amount of target DNA to be amplified. For example, we used 90 ng of plasmid pCA24N-*cynT* as the starting point for one of our epPCR libraries. In total, this plasmid was 5,180 bp in size. However, the amplified product (i.e., the *cynT* gene, plus flanking sequences) was 857 bp. Therefore, the amount of template DNA in the reaction was given by

$$\left(\frac{857\text{ bp}}{5{,}180\text{ bp}}\right)\times 90\text{ ng} = 15\text{ng}$$

7. We typically aim for a medium-to-high mutation rate, because this generates libraries that contain minimal numbers of "wasted" variants (i.e., unmutated copies of the template, or multiple copies of variants with any one point mutation). This strategy is discussed in more detail elsewhere (12, 15).
8. 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.
9. Calculating the PCR efficiency parameter allows robust statistical analysis of library composition (*see* Subheading 3.4, step 12). When the total product yield and the amount of starting template are known, the number of doublings in the PCR, d, can be calculated as follows:

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

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

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

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.

10. It is well known that UV transillumination of ethidium bromide-stained DNA can induce damage, resulting in lower cloning and transformation efficiencies (17). 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 use SYBR Safe stain and a blue-light transilluminator for preparation of our library vector and epPCR insert. In our hands, this results in libraries that are 5–10 times larger than equivalent libraries prepared with ethidium bromide-stained DNA.
11. 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. It also ensures that the ligation protocol is optimized for constructing a full-sized library with low "vector only" background and the maximum number of insert-containing clones.
12. Aliquots of cells should be thawed on ice. DNA is added to each 50 µl aliquot of cells and chilled on ice in a sterile Gene Pulser cuvette. Samples are electroporated at 2.5 kV, 200 Ω, and 25 µF in a Gene Pulser unit with Pulse Controller.

13. This is a rough estimate, based on the following: (a) Tenfold scale-up of the ligation reaction, (b) transforming 10–20 aliquots of electrocompetent cells, and (c) spreading all 550 μL of each recovery culture (instead of 10–50 μL aliquots).
14. Clones from the test library (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 in 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 3.6). 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.
15. 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 (12). 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 15.2). 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 also 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 (Subheading 3.4, step 12).
16. 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 7) should be used in a 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.
17. The large volume and high surface area of the square bioassay dishes make them prone to "sweating" when they are incubated at 30–37°C (particularly if the LB agar is too hot when the plates are poured). They may need to be pre-warmed at 37°C for 4–6 h and/or dried in a laminar flow hood or a class II biosafety cabinet (10–15 min), before they are dry enough to use.
18. For *E. coli* strain DH5α-E, we find that $OD_{600} = 1$ corresponds to ~2.5×10^8 cells/mL.

Acknowledgment

The authors gratefully acknowledge financial support for this work from the New Zealand Marsden Fund.

References

1. Turner NJ (2009) Directed evolution drives the next generation of biocatalysts. Nat Chem Biol 5:567–573
2. Jäckel C, Hilvert D (2010) Biocatalysts by evolution. Curr Opin Biotechnol 21:753–759
3. Peisajovich SG, Tawfik DS (2007) Protein engineers turned evolutionists. Nat Methods 4:991–994
4. Lutz S, Patrick WM (2004) Novel methods for directed evolution of enzymes: quality, not quantity. Curr Opin Biotechnol 15:291–297
5. Otten LG, Quax WJ (2005) Directed evolution: selecting today's biocatalysts. Biomol Eng 22:1–9
6. Arnold FH, Georgiou G (eds) (2003) Directed enzyme evolution: screening and selection methods. *Methods in Molecular Biology*, vol 230, Humana Press, Totowa, New Jersey.
7. 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
8. Cadwell RC, Joyce GF (1992) Randomization of genes by PCR mutagenesis. PCR Methods Appl 2:28–33
9. Cirino PC, Mayer KM, Umeno D (2003) Generating mutant libraries using error-prone PCR. Methods Mol Biol 231:3–9
10. McCullum EO, Williams BA, Zhang J, Chaput JC (2010) Random mutagenesis by error-prone PCR. Methods Mol Biol 634:103–109
11. 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 23:304–310
12. 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
13. Soo VWC, Hanson-Manful P, Patrick WM (2011) Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli*. Proc Natl Acad Sci USA 108:1484–1489
14. Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, Toyonaga H, Mori H (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. DNA Res 12:291–299
15. 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
16. Hanahan D, Jessee J, Bloom FR (1991) Plasmid transformation of *Escherichia coli* and other bacteria. Methods Enzymol 204:63–113
17. Hartman PS (1991) Transillumination can profoundly reduce transformation frequencies. Biotechniques 11:747–748