Chapter 7

Random Mutagenesis by Error-Prone PCR

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

Abstract

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

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

1. Introduction

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

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

Jeff Braman (ed.), *In Vitro Mutagenesis Protocols: Third Edition*, Methods in Molecular Biology, vol. 634,
DOI 10.1007/978-1-60761-652-8_7, © Springer Science+Business Media, LLC 2010

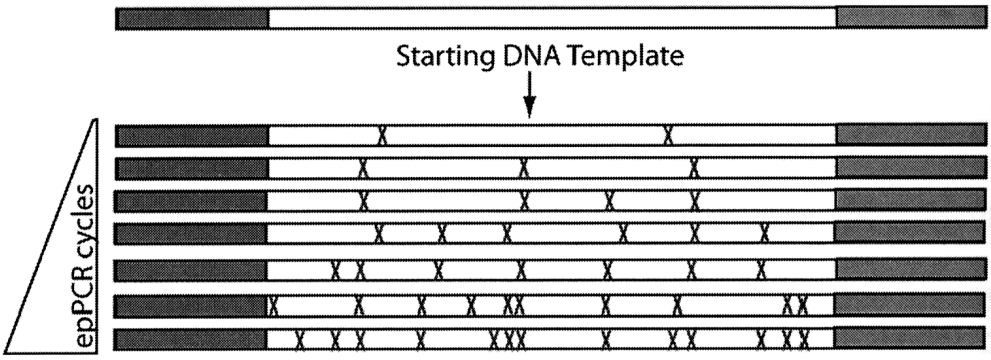

Fig. 1. Random mutagenesis by error-prone PCR. Mutations are randomly inserted into the DNA sequence under conditions that reduce the fidelity of *Taq* DNA polymerase. The number of mutations increase with the number of gene duplication events. PCR primer binding sites are denoted at the 5′ and 3′ of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

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

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

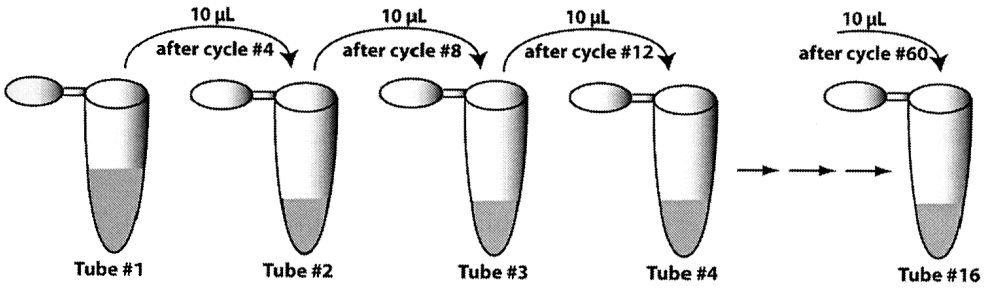

Fig. 2. Dilution and pooling technique used in epPCR. The current protocol uses a series of dilution and amplification steps to generate a mutagenic library that contains a range of single-nucleotide point mutations. After every four cycles of PCR amplification, a small portion of the PCR reaction mixture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

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

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

2. Materials

2.1. Biological and Chemical Materials

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

2.2. Equipment

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

3. Method (See Note 6)

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

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

PCR reagent mixture

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

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

PCR thermocycler program[a,b]

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

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

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

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

4. Notes

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

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

References

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