

Program Availability

Software for running SCHEMA calculations is available at the Arnold group web site at the California Institute of Technology (<http://www.che.caltech.edu/groups/fha/>).

[5] Staggered Extension Process *In Vitro* DNA Recombination

By HUIMIN ZHAO

Introduction

In vitro DNA recombination is an extremely powerful approach for the directed evolution of proteins and nucleic acids. Unlike random mutagenesis methods in which point mutations are introduced randomly into a single parent sequence to produce a library of progeny sequences, DNA recombination methods entail the block-wise exchange of genetic variations among multiple parent sequences created in the laboratory or existing in nature to produce a library of chimeric progeny sequences. The key advantage of DNA recombination is its ability to accumulate beneficial mutations while simultaneously removing deleterious mutations, which may greatly accelerate the evolution of a protein or nucleic acid molecule of interest toward a specific function. It was demonstrated in computational simulation studies that DNA recombination plays a critical role in the evolution of biological systems.¹ In the past decade, *in vitro* DNA recombination has been used successfully to alter and engineer many types of protein function, such as stability, activity, affinity, selectivity, substrate specificity, and protein folding/solubility.^{2,3}

The first described *in vitro* DNA recombination method, or “DNA shuffling,” was developed by Stemmer in 1994,^{4,5} in which DNA fragments generated by the random digestion of parent genes with DNase I are combined and reassembled into full-length chimeric progeny genes in a polymerase chain reaction (PCR)-like process. Since then, a number of *in vitro* DNA recombination methods have been described,⁶ such as

¹ S. Forrest, *Science* **261**, 872 (1993).

² O. Kuchner and F. H. Arnold, *Trends Biotechnol.* **15**, 523 (1997).

³ C. Schmidt-Dannert, *Biochemistry* **40**, 13125 (2001).

⁴ W. P. Stemmer, *Proc. Natl. Acad. Sci. USA* **91**, 10747 (1994).

⁵ W. P. Stemmer, *Nature* **370**, 389 (1994).

staggered extension process (StEP) recombination,⁷ random-priming recombination (RPR),⁸ random chimeragenesis on transient templates (RACHITT),⁹ degenerate homoduplex recombination (DHR),¹⁰ and synthetic shuffling.¹¹ This article describes the method, protocol, and applications of StEP recombination. For additional technical discussions on the same topic, interested readers are referred elsewhere.^{12,13}

Principle of the StEP Method

StEP recombination is based on template switching during polymerase-catalyzed primer extension.⁷ As illustrated in Fig. 1, the StEP method uses full-length genes as templates for the synthesis of chimeric progeny genes. It consists of priming denatured templates, followed by repeated cycles of denaturation and extremely short annealing/extension steps. Recombinogenic events occur when the partially extended primers anneal randomly to different templates (template-switching events) based on sequence complementarity and extend further. StEP is continued until full-length genes are formed. If the product yield is low, the full-length chimeric genes can be amplified in a standard PCR. Compared to DNA shuffling and other polymerase-based recombination methods that require fragmentation or chemical synthesis of fragments, the StEP method is much simpler and less labor intensive and can be performed using a pair of flanking primers in a single PCR tube. It is noteworthy that the StEP method is somewhat similar to the process that retroviruses, including HIV, use to evolve their genomes.¹⁴

The recombination efficiency of the StEP method was compared to the most widely used DNA shuffling method using a green fluorescent protein (GFP)-based recombination test system.¹⁵ A series of truncated GFP

⁶ H. Zhao and W. Zha, in "Enzyme Functionality: Design, Engineering and Screening" (A. Svendsen, ed.), p. 353. Dekker, New York, 2003.

⁷ H. Zhao, L. Giver, Z. Shao, J. A. Affholter, and F. H. Arnold, *Nature Biotechnol.* **16**, 258 (1998).

⁸ Z. Shao, H. Zhao, L. Giver, and F. H. Arnold, *Nucleic Acids Res.* **26**, 681 (1998).

⁹ W. M. Coco, W. E. Levinson, M. J. Crist, H. J. Hektor, A. Darzins, P. T. Pienkos, C. H. Squires, and D. J. Monticello, *Nature Biotechnol.* **19**, 354 (2001).

¹⁰ W. M. Coco, L. P. Encell, W. E. Levinson, M. J. Crist, A. K. Loomis, L. L. Licato, J. J. Arensdorf, N. Sica, P. T. Pienkos, and D. J. Monticello, *Nature Biotechnol.* **20**, 1246 (2002).

¹¹ J. E. Ness, S. Kim, A. Gottman, R. Pak, A. Krebber, T. V. Borchert, S. Govindarajan, E. C. Mundorff, and J. Minshull, *Nature Biotechnol.* **20**, 1251 (2002).

¹² A. M. Aguinaldo and F. H. Arnold, *Methods Mol. Biol.* **192**, 235 (2002).

¹³ A. A. Volkov and F. H. Arnold, *Methods Enzymol.* **328**, 447 (2000).

¹⁴ W. S. Hu, E. H. Bowman, K. A. Delviks, and V. K. Pathak, *J. Virol.* **71**, 6028 (1997).

¹⁵ A. A. Volkov, Z. Shao, and F. H. Arnold, *Nucleic Acids Res.* **27**, e18 (1999).

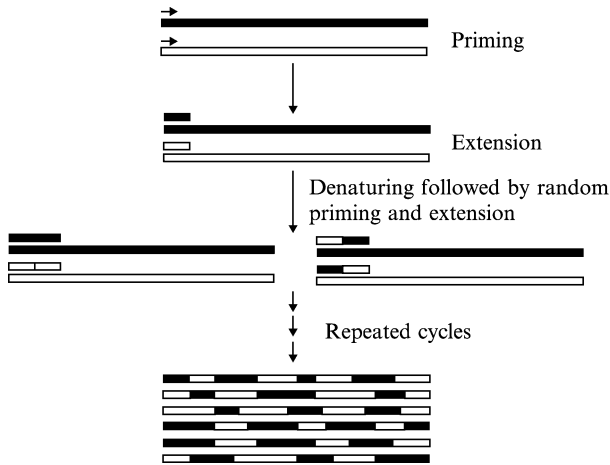


FIG. 1. Principle of the StEP recombination method. For simplicity, only one primer and single strands from two DNA templates (open and solid blocks) are shown. During priming, oligonucleotide primers anneal to the denatured templates. Short fragments are produced by brief polymerase-catalyzed primer extension. During subsequent random annealing-abbreviated extension cycles, fragments randomly prime the template (template switching) and extend further. The process is repeated for many cycles until the full-length genes are produced. The full-length chimeric genes can be amplified in a standard PCR (optional).

variants containing stop codon mutations that are nonfluorescent were created by site-directed mutagenesis at selected positions along the GFP gene. Recombination between truncated GFP variants generates the full-length wild-type gene and restores fluorescence. The percentage of fluorescent host *Escherichia coli* colonies indicates the recombination frequency or efficiency between two stop codon mutations of a given distance. As summarized in [Table I](#), the StEP method and the DNA shuffling method are equally efficient. With DNase I fragmentation, using small fragments (<100 bp) yields a slightly higher efficiency than large fragments (100–200 bp).

Materials

DNA templates containing the target sequences to be recombined

Oligonucleotide primers

Taq DNA polymerase and its 10× reaction buffer: 500 mM KCl,
100 mM Tris-HCl, pH 8.3 (Promega, Madison, WI)

25 mM MgCl₂

TABLE I
COMPARISON OF StEP RECOMBINATION AND DNA SHUFFLING¹⁵

Distance between mutations (bp)	Fraction of fluorescent colonies (%) ^a		
	StEP	DNA shuffling (<100-bp fragments)	DNA shuffling (100- to 200-bp fragments)
423	18.5	20.5	19.2
315	13.1	14.5	9.7
207	9.8	11.5	8.3
99	8.2	9.6	8.4
24	4.8	5.8	5.1

^aPercentage of fluorescent *E. coli* colonies obtained by recombining two green fluorescent protein templates containing stop codon mutations.

10× dNTP mix: 2 mM of each dNTP (Roche Diagnostics, Indianapolis, IN)

Agarose gel electrophoresis supplies and equipment

MJ PTC-200 thermocycler (MJ Research Inc., Watertown, MA)

QIAEX II gel extraction kit (QIAGEN, Valencia, CA)

DpnI restriction endonuclease (20 U/μl) and 10× supplied reaction buffer (New England Biolabs, Beverly, MA)

Experimental Approach

1. Prepare DNA template. Appropriate templates include plasmids carrying target sequences, cDNA or genomic DNA carrying the target sequences, sequences excised by restriction endonucleases, and PCR-amplified sequences.

2. Combine 5 μl of 10× *Taq* buffer, 5 μl of 10× dNTP mix (2 mM of each dNTP), 1.5 mM MgCl₂, 1–20 ng total template DNA, 30–50 pmol of each primer, sterile dH₂O, and 2.5 U *Taq* DNA polymerase in a total volume of 50 μl.

3. Run 80–100 extension cycles using the following program: 94° for 30 s (denaturation) and 55° for 5–15 s (annealing/extension).

4. Run a small aliquot (5–10 μl) of the reaction on an agarose gel. Possible reaction products are full-length amplified sequences, a smear, or a combination of both. If a discrete band with sufficient yield for subsequent cloning is obtained after the StEP reaction, no additional amplification step is needed. Proceed to step 8.

5. (Optional) If parent templates were isolated from a *dam* methylation-positive *E. coli* strain (e.g., DH5 α , XL1-Blue), the products from extension reactions can be incubated with *DpnI* endonuclease to remove parent DNA so as to reduce the background of nonchimeric genes. Combine 2 μ l of the StEP reaction, 1 \times *DpnI* reaction buffer, 6 μ l of sterile dH₂O, and 1 μ l of *DpnI* restriction endonuclease. Incubate at 37 $^{\circ}$ for 1 h.

6. Amplify the target sequence in a standard PCR. Combine 1 μ l of the StEP reaction, 0.3–1.0 μ M of each primer, 10 μ l of 10 \times *Taq* buffer, 1.5 mM MgCl₂, 10 μ l of 10 \times dNTP mix (2 mM of each dNTP), and 2.5 U *Taq* DNA polymerase in a total volume of 100 μ l. Run the PCR reaction using the following program: 96 $^{\circ}$ for 2 min, 25 cycles of 30 s at 94 $^{\circ}$, 30 s at 55 $^{\circ}$, and 60 s at 72 $^{\circ}$ for each 1 kb in length. The final step of elongation is at 72 $^{\circ}$ for 7 min.

7. Run a small aliquot of the reaction mixture (5–10 μ l) on an agarose gel. In most cases, a clear, discrete band of the correct size among a smear should be obtained.

8. Purify the product of correct size using the QIAEX II gel purification kit. Digest the product with the appropriate restriction endonucleases and ligate into the desired cloning vector.

Notes

1. *Primer design.* Primer design should follow standard criteria, including similar melting temperatures and elimination of self-complementarity or complementarity of primers to each other. Free computer programs such as Primer3 at Biology Workbench (<http://workbench.sdsc.edu>) can be used to design primers. Typically, primers should also include unique restriction sites for subsequent directional subcloning.

2. *Choice of a DNA polymerase.* The key to successful recombination by StEP is to tightly control the polymerase-catalyzed DNA extension. Too much extension during each StEP cycle will severely limit recombination events. Thermostable DNA polymerases currently used in DNA amplification are often very fast. Even very brief cycles of denaturation and annealing provide time for these enzymes to extend primers for hundreds of nucleotides. For example, extension rates of *Taq* DNA polymerase at various temperatures are: 70 $^{\circ}$, >60 nucleotides/s; 55 $^{\circ}$, ~24 nucleotides/s; 37 $^{\circ}$, ~1.5 nucleotides/s; 22 $^{\circ}$, ~0.25 nucleotides/s.¹⁶ Thus, it is

¹⁶ M. A. Innis, K. B. Myambo, D. H. Gelfand, and M. A. Brow, *Proc. Natl. Acad. Sci. USA* **85**, 9436 (1988).

not unusual for the full-length gene product to appear after only 10–15 cycles. Unfortunately, the faster the full-length gene product appears in the extension reaction, the lower the recombination frequency due to the fewer number of the template switching events. To increase the recombination frequency, various measures should be taken to minimize the time spent in each StEP cycle, including selecting a faster thermocycler, reducing the reaction volume, and using smaller PCR tubes with thin walls.

Alternatively, thermostable DNA polymerases with proofreading activity can be used. It was reported that the proofreading activity of high-fidelity DNA polymerases can significantly slow down their extension rates.¹⁷ For example, *Vent* DNA polymerase has an extension rate of 1000 nucleotides/min and a processivity of 7 nucleotides/(initiation event) as compared to >4000 nucleotides/min and 40 nucleotides/(initiation event) for *Taq* DNA polymerase at a certain extension temperature. In addition, use of these alternative polymerases is highly recommended during DNA amplification to minimize the mutagenic rate of point mutations. Commercially available thermostable DNA polymerases with proofreading activity include *Pfu* DNA polymerase (Stratagene, La Jolla, CA), *Vent* DNA polymerase (New England Biolabs, Beverly, MA), and *Pfx* DNA polymerase (Invitrogen Life Technologies, Carlsbad, CA). When setting up reactions with these polymerases, it is very important to add the polymerase last, as in the absence of dNTPs, the 3' to 5' exonuclease activity of the polymerase can degrade DNAs.

3. *Choice of annealing/extension temperatures and times.* As a general rule, the annealing temperature should be a few degrees lower than the melting temperature of the primers. The annealing temperature should be decreased when higher recombination frequency is required or when templates have low GC content. However, it should not be reduced too much in order to minimize nonspecific annealing events. The annealing/extension times are chosen based on the desired recombination frequency. Both shorter extension times and lower annealing temperatures will increase the recombination frequency. The number of the annealing/extension cycles is determined by the size of the full-length gene product.

4. *Extension products.* The progress of the StEP reaction can be monitored by taking aliquots of the reaction mixture at various time points and separating the DNA fragments by agarose gel electrophoresis. The appearance of the extension products may depend on the specific

¹⁷ M. S. Judo, A. B. Wedel, and C. Wilson, *Nucleic Acids Res.* **26**, 1819 (1998).

sequences recombined or the type of templates used. Small templates will likely show gradual accumulation of the full-length gene products with an increasing number of cycles. For example, during StEP recombination of two subtilisin E genes (~1 kb), the average size of the extension products increases gradually with increasing cycle number: 100 bp after 20 cycles, 400 bp after 40 cycles, 800 bp after 60 cycles, and a clear discrete band around 1 kb (the desired size) after 80 cycles.⁷ However, using large templates such as whole plasmids and long genes may result in nonspecific annealing of primers and their extension products throughout the templates. Although it may appear as a smear on the agarose gel, the increase of the size of their extension products may not be so obvious.

5. *PCR amplification.* If the PCR amplification reaction is not successful, i.e., no discrete band with sufficient yield is produced, repeat amplification using serial dilutions of the StEP reaction mixture: 1:10 dilution, 1:20 dilution, and 1:50 dilution. Run small aliquots of the amplified products on an agarose gel to determine the yield and quality of amplification. Select the reaction with a higher yield and lower amount of nonspecific products for subsequent cloning. An alternative solution is to use nested internal primers separated by 50–100 bp from the original primers to amplify the target sequences.

Applications of the StEP Method

The StEP method has been used successfully to recombine gene variants created from random mutagenesis and naturally occurring homologous genes that are approximately 80% identical. For example, the StEP method was used to increase the temperature optimum of subtilisin E by 18° over that of the wild-type enzyme, essentially converting a mesophilic enzyme into its thermophilic counterpart.¹⁸ The substrate specificity of biphenyl dioxygenase was broadened by StEP recombination of two homologous *bphA* genes encoding *Burkholderia cepacia* LB400 biphenyl dioxygenase and *Pseudomonas pseudoalcaligenes* KF707 biphenyl dioxygenase, respectively.¹⁹ Unlike the two parental dioxygenases, which preferentially recognize either *ortho*-(LB400) or *para*-(KF707) substituted polychlorinated biphenyls, the evolved variants can degrade both congeners to the same extent.

¹⁸ H. Zhao and F. H. Arnold, *Protein Eng.* **12**, 47 (1999).

¹⁹ F. Bruhlmann and W. Chen, *Biotechnol. Bioeng.* **63**, 544 (1999).

By combining StEP recombination and DNA shuffling based directed evolution and structure-based rational design, Altamirano and co-workers²⁰ engineered a novel function in an α/β barrel enzyme by completely converting the activity of indole-3-glycerol-phosphate synthase (IGPS) to that of phosphoribosylanthranilate isomerase (PRAI). A structure-based design was used to modify the IGPS α/β barrel by incorporation of the basic design of the loop system of PRAI, yielding a chimeric variant with very low PRAI activity. DNA shuffling, StEP recombination, and genetic selection were then used to increase the PRAI activity, which led to the creation of a variant exhibiting sixfold higher activity than wild-type PRAI and no IGPS activity. Other applications of StEP recombination include the alteration of the regioselectivity of a *Bacillus* α -galactosidase,²¹ the thermostabilization of cellulosomal endoglucanase EngB by recombining its gene with a homologous gene encoding the noncellulosomal endoglucanase EngD,²² and the improvement of the protein expression level as well as the enzyme activity of a fungal laccase in *Saccharomyces cerevisiae*.²³

Concluding Remarks

The recombination efficiency of the StEP method is similar to that of the most widely used *in vitro* gene recombination method, DNA shuffling. However, the StEP method does not require DNA fragmentation and can be carried out in a single tube. Thus, the simple and efficient StEP recombination method represents a powerful tool that can be applied to the directed evolution of genes, operons, and metabolic pathways for specific applications.

²⁰ M. M. Altamirano, J. M. Blackburn, C. Aguayo, and A. R. Fersht, *Nature* **403**, 617 (2000).

²¹ M. Dion, A. Nisole, P. Spangenberg, C. Andre, A. Glottin-Fleury, R. Mattes, C. Tellier, and C. Rabiller, *Glycoconj J.* **18**, 215 (2001).

²² K. Murashima, A. Kosugi, and R. H. Doi, *Mol. Microbiol.* **45**, 617 (2002).

²³ T. Bulter, M. Alcalde, V. Sieber, P. Meinhold, C. Schlachtbauer, and F. H. Arnold, *Appl. Environ. Microbiol.* **69**, 987 (2003).