In vitro 'sexual' evolution through the PCR-based staggered extension process (StEP)

Huimin Zhao & Wenjuan Zha

Departments of Chemical and Biomolecular Engineering, Chemistry, and Bioengineering, Institute for Genomic Biology, Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA. Correspondence should be addressed to H.Z. (zhao5@uiuc.edu).

Published online 22 November 2006; doi:10.1038/nprot.2006.309

This protocol describes a directed evolution method for *in vitro* mutagenesis and recombination of polynucleotide sequences. The staggered extension process (StEP) is essentially a modified PCR that uses highly abbreviated annealing and extension steps to generate staggered DNA fragments and promote crossover events along the full length of the template sequence(s). The resulting library of chimeric polynucleotide sequence(s) is subjected to subsequent high-throughput functional analysis. The recombination efficiency of the StEP method is comparable to that of the most widely used *in vitro* DNA recombination method, DNA shuffling. However, the StEP method does not require DNA fragmentation and can be carried out in a single tube. This protocol can be completed in 4–6 h.

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

INTRODUCTION

Directed evolution is an extremely powerful method for engineering of and fundamental studies of biomolecules and biological systems[1–3]. It mimics Darwinian evolution in the test tube and consists of iterative rounds of creating genetic diversity followed by high-throughput screening or selection. Both random mutagenesis and DNA recombination have been used effectively to create genetic diversity. However, random mutagenesis can introduce point mutations only into a single parent sequence to create a library of chimeric progeny sequences, whereas DNA recombination involves blockwise exchange of genetic variations among multiple parental sequences created in the laboratory or existing in nature to produce a library of chimeric progeny sequences. As an 'asexual' evolutionary process, random mutagenesis relies on iterative rounds of mutagenesis and screening to slowly accumulate beneficial mutations, but the deleterious mutations gathered along the way cannot be removed. In contrast, the 'sexual' DNA recombination can combinatorially explore the mutations encoded in the parental sequences, thus greatly expanding the protein sequence space that can be searched. Moreover, DNA recombination allows for rapid accumulation of beneficial mutations and simultaneous removal of deleterious mutations.

The power of DNA recombination over random mutagenesis has been well demonstrated in the molecular evolution of TEM-1 β-lactamase[4]. Three rounds of DNA shuffling followed by two rounds of backcrossing with the wild-type gene resulted in mutants with 32,000-fold improvement in the minimum inhibitory concentration (MIC) of the antibiotic cefotaxime, whereas error-prone PCR resulted in mutants with only a 16-fold improvement in MIC after three iterations[4]. In addition, the importance of DNA recombination in the evolution of biological systems has been demonstrated by computational simulation studies[5,6].

StEP is a simple yet efficient *in vitro* homology-dependent DNA recombination method[7] developed shortly after the invention of the first *in vitro* DNA recombination method, DNA shuffling[4,8]. In the DNA shuffling method, multiple homologous parental genes are randomly fragmented, and the resulting fragments are then assembled into full-length chimeric progeny genes by repeated cycles of overlap extension reaction. Recombinogenic events occur when fragments derived from different parental genes prime each other. In contrast, the StEP method uses full-length homologous genes as templates for the synthesis of chimeric progeny genes (**Fig. 1**). It consists of priming the denatured templates followed by repeated cycles of denaturation and extremely short annealing/extension steps. Recombinogenic events occur when the partially extended primers randomly anneal 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. Note that due to the intrinsically imperfect fidelity of DNA polymerases, random point mutations will be introduced into progeny genes in both methods. These random point mutations are desirable in many directed evolution experiments, since they represent an additional source of genetic

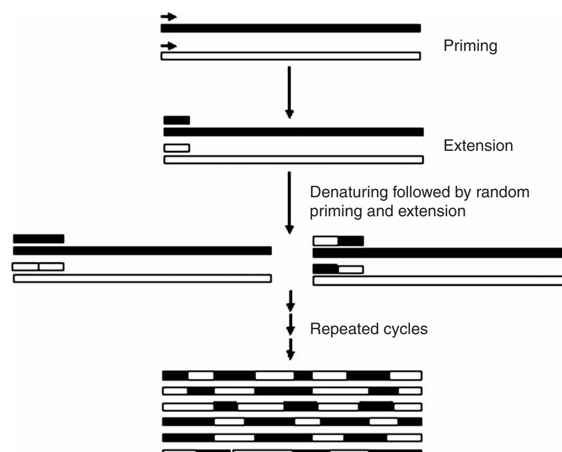

Figure 1 | Principle of the StEP method. For simplicity, only one primer and single strands from two DNA templates (open and filled rectangles) are shown. During priming, oligonucleotide primers anneal to the denatured templates. Short fragments are produced by brief polymerase-catalyzed primer extension. During subsequent random abbreviated annealing/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).

BOX 1 | COMPARISON OF THE STEP METHOD AND DNA SHUFFLING

We compared the recombination efficiency of the StEP method with the most widely used DNA shuffling method using an *Aequorea victoria* green fluorescent protein (GFP)-based recombination test system[19]. The DNA templates of this test system are a series of GFP variants created by insertion of the sequence TAAT (containing a stop codon) at selected positions along the GFP gene via site-directed mutagenesis. The stop codon results in truncated GFP products that are nonfluorescent. Recombination between truncated GFP variants generates the full-length wild-type gene and restores fluorescence (**Fig. 2**). The percentage of fluorescent host *E. coli* colonies indicates the recombination frequency or efficiency between two stop codon mutations of a given distance. Recombination by the StEP method was performed using *Taq* DNA polymerase, whereas in DNA shuffling, two sets of DNA fragments, differing in the size range after DNase I digestion, were reassembled by *Pfu* DNA polymerase. As summarized in **Table 1**, 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).

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

diversity. If needed, however, the rate of mutagenesis can be significantly reduced by using high-fidelity DNA polymerases[9,10].

Since the development of DNA shuffling and StEP, many other *in vitro* homology-dependent DNA recombination methods have been reported in literature, such as random-priming recombination (RPR)[11], random chimeragenesis on transient templates (RACHITT)[12], degenerate oligonucleotide gene shuffling (DOGS)[13], degenerate homoduplex recombination (DHR)[14] and synthetic shuffling[15]. In addition, several *in vitro* homology-independent DNA recombination methods have also been developed, such as incremental truncation for the creation of hybrid enzymes (ITCHY)[16], sequence homology–independent protein recombination (SHIPREC)[17], and nonhomologous random recombination (NPR)[18]. However, despite these new developments, DNA shuffling and StEP remain the preferred *in vitro* DNA recombination methods for directed evolution.

Compared to DNA shuffling and other polymerase-based homology-dependent DNA 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. Most importantly, a direct and systematic comparison between DNA shuffling and the StEP method has demonstrated that the recombination efficiency of these two methods is almost identical[19]. As discussed in **Box 1**, nonfluorescent truncated GFP variants containing internal stop codons at different positions were recombined by these two methods to restore the full-length GFP with fluorescence so that we were able to estimate recombination efficiency between two stop codons of a given distance by the percentage of fluorescent host *E. coli* colonies (**Fig. 2**). As summarized in **Table 1**, the StEP method gives similar recombination efficiency as the DNA shuffling method, employing 100- to 200-bp fragments for a distance ranging from 20 bp to 500 bp. Although DNA shuffling is slightly more efficient when small fragments (<100 bp) are used, it is usually difficult to reassemble fragments of such size by DNA shuffling. In another instance, the StEP method and DNA shuffling were also tested with two thermostable subtilisin E mutants and showed similar DNA recombination efficiency[7]. In both methods, a minimum of one to four crossovers were observed.

The StEP method ultimately takes advantage of the intrinsic ability of PCR to generate chimeric progeny genes from mixed homologous templates[20,21]. Notably, this approach is quite reminiscent of the homologous recombination tactic that retroviruses such as HIV naturally use to evolve their genomes[22]. For additional technical discussions on the same topic, see refs. 23–25.

The StEP method has been successfully used to engineer proteins, antibodies, and viruses with improved or novel functions. For example, the StEP method has been used to increase the temperature optimum of subtilisin E by 18 °C over that of the wild-type enzyme, essentially converting a mesophilic enzyme into its thermophilic counterpart[26]. The StEP method has also been used in combination with random mutagenesis to convert cytochrome P450 BM-3 from *Bacillus megaterium* into a soluble, self-sufficient, highly active alkane hydroxylase[27]. In addition, by coupling with a compartmentalized self-replication (CSR)-based high-throughput screening method, the StEP method has been used to significantly expand the substrate spectrum of *Taq* DNA polymerase for biotechnological applications such as generation of highly fluorescent microarray or *in situ* hybridization (FISH) probes[28]. Most recently, the StEP method was used to evolve adeno-associated virus with enhanced properties such as a better ability to evade antibody neutralization for gene delivery[29]. Other applications of the StEP method include the alteration of the regioselectivity of a *Bacillus stearothermophilus* α-galactosidase[30]; the creation of bacterial 5-enolpyruvylshikimate-3-phosphate synthase mutants with enhanced tolerance to glyphosate[31]; the thermostabilization of cellulosomal endoglucanase EngB by recombining its gene with a homologous gene encoding the noncellulosomal endoglucanase

TABLE 1 | Comparison of the recombination efficiency between the StEP method and DNA shuffling.

	Fraction of fluorescent colonies (%)		
Distance between mutations (bp)	**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

EngD[32]; the improvement of protein expression as well as enzyme activity of a fungal laccase in *Saccharomyces cerevisiae*[33]; the expansion of substrate specificity of cytochrome P450 2A6 (ref. 34); the thermostabilization of *Bacillus subtilis* family-11 xylanase[35]; and the enhancement of the binding affinity of single-domain antibodies toward plant hormone indole-3-acetic acid[36].

This protocol describes the procedure of the StEP method using the example of two thermostable subtilisin E mutants, *RC1* and *RC2*, each carrying a beneficial mutation. Note that the number of parental templates for the StEP method, however, is not limited to two. In addition, as with many other homologous recombination methods, the StEP method requires minimal sequence identity of 85% among the parental genes for efficient recombination[25].

Experimental design

Although described specifically for the recombination of two thermostable subtilisin E mutants, the StEP protocol outlined here can be used to recombine other homologous genes, with some minor modifications in the staggered extension process. In general, the annealing temperature should be a few degrees lower than the melting temperature of the primers. It should be further reduced when higher recombination efficiency is demanded or when template DNAs have low GC content, but it should not be decreased so much as to promote nonspecific annealing. The annealing/extension time should be chosen based on the desired recombination frequency, as a shorter extension time usually increases the recombination frequency. The number of extension cycles is determined by both the size of the full-length gene products and the extension time used. A typical 50-μl PCR reaction contains 1–20 ng of each template DNA in equimolar ratio and 10–50 pmol of each primer. A primer-to-template ratio of 100–500 is often preferred in the StEP method. Following the general guideline mentioned above, a few test runs should be performed to optimize the StEP conditions until the desired degree of recombination is achieved.

MATERIALS

REAGENTS

- Agarose
- Oligonucleotide primers specific to subtilisin E: P5N (5′-CCGAGCGTTG CATATGTGGAAG-3′; underlined sequence is *Nde*I restriction site) and P3B (5′-CGACTCTAGAGGATCCGATTC-3′; underlined sequence is *Bam*HI restriction site)
- *Taq* DNA polymerase and its 10× reaction buffer: 500 mM KCl, 100 mM Tris-HCl, pH 8.3 (Promega)
- 25 mM $MgCl_2$
- 2′-deoxynucleoside 5′-triphosphate (dNTPs) (Roche Diagnostics)
- QIAprep Spin Miniprep Kit, QIAquick PCR Purification Kit, QIAEX II Gel Extraction Kit or QIAquick Gel Extraction Kit (QIAgen)
- DpnI restriction endonuclease (20 U μl–1) and 10× reaction buffer (New England Biolabs)
- 10× gel-loading buffer (50% glycerol (vol/vol) and 0.1% bromophenol blue (wt/vol))
- SYBR gold nucleic acid gel stain (Invitrogen)

EQUIPMENT

- MJ PTC-200 thermocycler (Bio-Rad Laboratories)
- 0.2-ml thin-walled PCR tubes (Bio-Rad Laboratories)

REAGENT SETUP

10× dNTP mix 2 mM of each dNTP; store at –20 °C.

PROCEDURE

1| Following the manufacturer's instructions, prepare plasmid DNA containing the gene of interest (in our case, *RC1* (pBE3-RC1) and *RC2* (pBE3-RC2)) as the template sequences for StEP recombination. The subtilisin E gene has a GC content of about 46.2%. Note that in addition to plasmids carrying target sequences, appropriate templates also include cDNA or genomic DNA carrying the target sequences, sequences excised by restriction endonucleases, and PCR-amplified sequences.

■ **PAUSE POINT** The template DNA after preparation may be stored at –20 °C indefinitely.

2| Combine 5 μl of 10× *Taq* buffer, 5 μl of 10× dNTP mix, 1.5 mM $MgCl_2$, 0.15 pmol of total template DNA containing an equimolar mixture of the two plasmids (here, pBE3-RC1 and pBE3-RC2), 30 pmol of each primer (here, 5′ flanking primer P5N with a melting temperature of 58.5 °C and 3′ flanking primer P3B with a melting temperature of 53.8 °C), sterile dH_2O, and 2.5 U *Taq* DNA polymerase in a total volume of 50 μl.

▲ **CRITICAL STEP** Primer design should follow standard criteria, including similar melting temperatures for both primers and elimination of self-complementarity or complementarity of primers to each other. Free computer programs such as Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) can be used to design primers (**Box 2**). Typically, primers should also include unique restriction sites for subsequent directional subcloning.

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

▲ **CRITICAL STEP** Very short annealing/extensions will increase recombination efficiency, and the high number extension cycles (80–100) will ensure the formation of full-length progeny genes.

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. For the two subtilisin E genes recombined here, a clear, discrete electrophoretic band around 1 kb (the size of full-length subtilisin E gene) among a smear is typically observed on the agarose gel. If a discrete band with sufficient yield for subsequent cloning is obtained after the StEP reaction, no additional amplification step is needed; instead, proceed directly to Step 9.

? TROUBLESHOOTING

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

BOX 2 | PRIMER3 TUTORIAL

1. In the source sequence panel, paste any of the parental sequences (5′→3′) that will be recombined by the StEP method. At least 30 bp of flanking sequence on either side of the sequence of interest should also be included as possible sites for primer design.
2. In the product size range panel, specify the size of the sequence to be recombined.
3. Click 'Pick Primers'.
4. The best pair of primers appears with essential parameters such as their starting position, length, T_m, and GC content. The product size from this primer pair is calculated, and their positions are marked on the source sequence. Some additional choices of primer pairs may also be listed.
5. Alternatively, if the left primer or right primer sequence is to be fixed, paste it in the left or right primer panel. Primer3 will identify the optimal matching primer that meets all the predefined primer selection criteria.

5| (Optional) If parent templates, such as plasmids, were isolated from a *dam* methylation–positive *E. coli* strain (e.g., DH5α, XL1-Blue), the products from extension reactions can be incubated with *Dpn*I endonuclease to remove parent DNA so as to reduce the background of nonchimeric genes. Combine 2 μl of the StEP reaction, 1× *Dpn*I reaction buffer, 6 μl of sterile dH_2O, and 1 μl of *Dpn*I restriction endonuclease. Incubate at 37 °C for 1 h.

6| Combine the following components in a 200 μl thin-walled PCR tube to amplify the chimeric full-length DNA products.

Volume	Component	Final concentration
10.0 μl	10× *Taq* buffer	1×
6.0 μl	$MgCl_2$ (25 mM)	1.5 mM
10.0 μl	10× dNTP mix	0.2 mM
5.0 μl	Forward primer (e.g., P5N) (10 μM)	0.5 μM
5.0 μl	Reverse primer (e.g., P3B) (10 μM)	0.5 μM
1.0 μl	StEP reaction mixture	–
62.5 μl	ddH_2O	–
0.5 μl	*Taq* DNA polymerase (5 U μl^{-1})	0.025 U μl^{-1}

If the optional Step 5 is used, 5 μl of the unpurified *Dpn*I-digested StEP product should be used as template for PCR amplification. In addition, the volume of ddH_2O should be reduced to 58.5 μl.

7| Amplify the target sequence in a standard PCR using the following conditions:

Cycle	Denaturation	Annealing	Extension
1	96 °C for 2 min	–	–
2–25	94 °C for 30 s	55 °C for 30 s	72 °C for 60 s
26	–	–	72 °C for 7 min

The typical yield of PCR amplification is about 1–3 μg DNA. The extension time should be increased proportionally for longer target sequences (i.e., use 60 s for every 1 kb in length).

8| Run a small aliquot of the reaction mixture (5–10 μl) on an agarose gel. In most cases, a clear, discrete band of the correct size (about 1 kb for the subtilisin E genes) among a smear should be obtained.
? TROUBLESHOOTING

9| Run the entire reaction mixture on an agarose gel and carefully excise the bright DNA band around 1 kb, which contains the full-length recombined subtilisin E gene products. Purify the DNA product using the QIAEX II Kit or QIAquick Gel Purification Kit.
■ **PAUSE POINT** The cleaned-up DNA product from StEP recombination may be stored at –20 °C indefinitely.

10| Digest the product with *Nde*I and *Bam*HI, purify using the QIAquick Gel Purification Kit, and ligate into the desired cloning vector.

● **TIMING**
Step 1, 0–8 h
Steps 2–3, 2–3 h
Step 4, 30 min
Steps 5–7, 2–3 h
Steps 8–9, 1–2 h
Step 10, 1–2 h

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Problem	Possible reason	Solution
No visible product of correct size in Step 4	Poor primers	Design primers using Primer3
	Insufficient number of annealing/extension cycles	Increase the number of cycles
	Improper annealing temperatures	Use an annealing temperature that is a few degrees lower than the melting temperature of primers
	Templates with high GC content	Decrease the annealing temperature
No PCR product in Step 8	Improper concentration of template DNA	Use serial dilutions of the StEP reaction mixture: 1:10 dilution, 1:20 dilution, and 1:50 dilution
	Poor primers	Use nested internal primers separated by 50–100 bp from the original primers to amplify the target sequences
	Templates with high GC content	Decrease the annealing temperature
	Insufficient extension time	Increase the extension time
Low recombination efficiency in the final StEP products	Annealing temperature too high	Decrease the annealing temperature
	Annealing/extension time too long	Reduce the extension time Use a faster thermocycler Use smaller reaction volume Use proofreading DNA polymerase
	Poor digestion by *Dpn*I	Increase the digestion time and/or concentration of *Dpn*I

ANTICIPATED RESULTS

The key to successful *in vitro* DNA recombination by the StEP method is to tightly control the DNA 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 >60 nucleotides per second at 70 °C, ~24 nucleotides per second at 55 °C, ~1.5 nucleotides per second at 37 °C, and ~0.25 nucleotides per second at 22 °C (ref. 37). Thus, it is 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, owing to the lower number of template-switching events. The most direct control over the polymerization extent during each PCR cycle is the annealing/extension temperature and time. The annealing/extension step in the StEP method therefore should be carried out at sufficiently low temperature ($T_m - 25\ ^\circ C < T_{annealing} < T_m - 5\ ^\circ C$) and short time (no more than a few seconds) to limit polymerization/extension and yet allow for high-fidelity primer annealing.

As an alternative, thermostable DNA polymerases with proofreading activity can be used. It has been reported that the proofreading activity of high-fidelity DNA polymerases can significantly slow down their extension rates[38]. For example, *Vent* DNA polymerase has an extension rate of 1,000 nucleotides per minute and processivity of 7 nucleotides per initiation event, as compared to >4,000 nucleotides per minute and 40 nucleotides per 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 rate of occurrence of point mutations. Commercially available thermostable DNA polymerases with proofreading activity include *Pfu* DNA polymerase (Stratagene), *Vent* DNA polymerase (New England Biolabs), *Pfx* DNA polymerase (Invitrogen Life Technologies), and *Phusion* DNA polymerase (New England Biolabs). When setting up reactions with these polymerases, it is very important to add the polymerase last, because in the absence of dNTPs, the 3′ to 5′ exonuclease activity of the polymerase can degrade DNAs.

To further increase the recombination frequency, various measures to minimize the time spent in each StEP cycle should

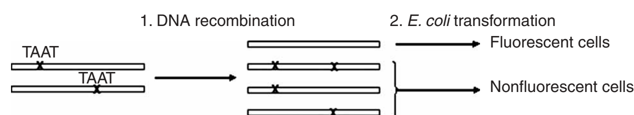

Figure 2 | Scheme of the GFP-based recombination test system. Two GFP variants, each containing a stop codon (introduced by the sequence TAAT) at a specific position, are subject to DNA recombination by either DNA shuffling or StEP. Four types of progeny are generated: two resulting from DNA recombination events between the stop codons, and two parent-like. After transformation into *E. coli*, only the full-length wild-type GFP product makes the host cell fluorescent. As a result, the percentage of fluorescent *E. coli* colonies indicates the recombination frequency or efficiency between the two introduced stop codons separated by a predefined distance.

also be taken, including selecting a faster thermocycler, reducing the reaction volume, and using smaller PCR tubes with thin walls. The MJ PTC-200 thermocycler from Bio-Rad that we have used successfully to perform StEP recombination in this protocol has a maximal rate of heating and cooling of 3 °C/s. Thermocylers with similar a ramping rate may also be used.

Preferred DNA templates for the StEP method are genes prepared by PCR amplification or endonuclease excision from the vector. Use of whole plasmids carrying the target genes might result in undesired recombination products, as the primers and/or the partially extended products anneal nonspecifically all over the vector sequence. Similar effects may also occur for large genes. The two primers used in the StEP method need to be well designed not only to meet the standard criteria for PCR primers (such as similar melting temperature and lack of complementarity to themselves or to each other) but also, more importantly, to avoid nonspecific annealing with the template, which may give rise to nonsense recombined products. Primer design using computer software such as the aforementioned Primer3 is a good practice. If restriction sites are included in the primer sequences, addition of two to ten bases at the 5′ end of the recognition site greatly increases the cleavage efficiency of most restriction enzymes. The template amount is not very crucial for a successful recombination by StEP. However, too much template should be avoided, since increased DNA concentration promotes nonspecific annealing and results in a considerable amount of unshuffled heteroduplexes of the parental strands as well as the chimeras. The StEP method generates new DNA linearly in its early cycle. Therefore, a primer-to-template ratio of 100–500 is required for StEP to obtain sufficient products, as compared to a ratio of 10^6 in a typical PCR amplification reaction.

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 sequences recombined or the type of the templates used. Small templates will likely show gradual accumulation of the full-length gene products with 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, recombining genes with unevenly distributed GC pairs or large templates, such as whole plasmids or long genes, may result in nonspecific annealing of primers and their extension products throughout the templates. The products may appear as a smear on the agarose gel, and the increase in the size of the extension products may not be so obvious.

ACKNOWLEDGMENTS We thank the Department of Defense (N000140210725), National Science Foundation (BES-0348107), National Institutes of Health, and DuPont for supporting our work on development and applications of new directed evolution tools for protein science and engineering and metabolic engineering.

AUTHOR CONTRIBUTIONS H.Z., first draft of the manuscript; W.Z., Box 2, Fig. 2, part of the Introduction and Anticipated Results; both authors participated in the editing and revision of the manuscript.

COMPETING INTEREST STATEMENT The authors declare that they have no competing financial interests.

Published online at http://www.natureprotocols.com
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Arnold, F.H. Combinatorial and computational challenges for biocatalyst design. *Nature* **409**, 253–257 (2001).
2. Arnold, F.H. (ed.) *Evolutionary Protein Design* (*Advances in Protein Chemistry*, Vol. 55) (Academic, San Diego, USA, 2001).
3. Schmidt-Dannert, C. Directed evolution of single proteins, metabolic pathways, and viruses. *Biochemistry* **40**, 13125–13136 (2001).
4. Stemmer, W.P. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389–391 (1994).
5. Holland, J.F. Genetic algorithms. *Sci. Am.* **267**, 66–72 (1992).
6. Forrest, S. Genetic algorithms: principles of natural selection applied to computation. *Science* **261**, 872–878 (1993).
7. Zhao, H., Giver, L., Shao, Z., Affholter, J.A. & Arnold, F.H. Molecular evolution by staggered extension process (StEP) *in vitro* recombination. *Nat. Biotechnol.* **16**, 258–261 (1998).
8. Stemmer, W.P. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**, 10747–10751 (1994).
9. Zhao, H. & Arnold, F.H. Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res.* **25**, 1307–1308 (1997).
10. Ninkovic, M., Dietrich, R., Aral, G. & Schwienhorst, A. High-fidelity *in vitro* recombination using a proofreading polymerase. *Biotechniques* **30**, 530–536 (2001).
11. Shao, Z., Zhao, H., Giver, L. & Arnold, F.H. Random-priming in vitro recombination: an effective tool for directed evolution. *Nucleic Acids Res.* **26**, 681–683 (1998).
12. Coco, W.M. *et al.* DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotechnol.* **19**, 354–359 (2001).
13. Gibbs, M.D., Nevalainen, K.M. & Bergquist, P.L. Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling. *Gene* **271**, 13–20 (2001).
14. Coco, W.M. *et al.* Growth factor engineering by degenerate homoduplex gene family recombination. *Nat. Biotechnol.* **20**, 1246–1250 (2002).
15. Ness, J.E. *et al.* Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol.* **20**, 1251–1255 (2002).
16. Ostermeier, M., Shim, J.H. & Benkovic, S.J. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* **17**, 1205–1209 (1999).
17. Sieber, V., Martinez, C.A. & Arnold, F.H. Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* **19**, 456–460 (2001).
18. Bittker, J.A., Le, B.V. & Liu, D.R. Nucleic acid evolution and minimization by nonhomologous random recombination. *Nat. Biotechnol.* **20**, 1024–1029 (2002).
19. Volkov, A.A., Shao, Z. & Arnold, F.H. Recombination and chimeragenesis by in vitro heteroduplex formation and in vivo repair. *Nucleic Acids Res.* **27**, e18 (1999).
20. Saiki, R.K. *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. *Science* **239**, 487–491 (1988).
21. Meyerhans, A., Vartanian, J.P. & Wainhobson, S. DNA recombination during PCR. *Nucleic Acids Res.* **18**, 1687–1691 (1990).
22. Hu, W.S., Bowman, E.H., Delviks, K.A. & Pathak, V.K. Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference. *J. Virol.* **71**, 6028–6036 (1997).
23. Zhao, H. Staggered extension process in vitro DNA recombination. *Methods Enzymol.* **388**, 42–49 (2004).
24. Aguinaldo, A.M. & Arnold, F. Staggered extension process (StEP) *in vitro* recombination. *Methods Mol. Biol.* **192**, 235–239 (2002).
25. Volkov, A.A. & Arnold, F.H. Methods for *in vitro* DNA recombination and random chimeragenesis. *Methods Enzymol.* **328**, 447–456 (2000).
26. Zhao, H. & Arnold, F.H. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng.* **12**, 47–53 (1999).

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols

27. Glieder, A., Farinas, E.T. & Arnold, F.H. Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nat. Biotechnol.* **20**, 1135–1139 (2002).
28. Ghadessy, F.J. *et al.* Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat. Biotechnol.* **22**, 755–759 (2004).
29. Maheshri, N., Koerber, J.T., Kaspar, B.K. & Schaffer, D.V. Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. *Nat. Biotechnol.* **24**, 198–204 (2006).
30. Dion, M. *et al.* Modulation of the regioselectivity of a *Bacillus* alpha-galactosidase by directed evolution. *Glycoconj. J.* **18**, 215–223 (2001).
31. He, M., Yang, Z.Y., Nie, Y.F., Wang, J. & Xu, P.L. A new type of class I bacterial 5-enopyruvylshikimate-3-phosphate synthase mutants with enhanced tolerance to glyphosate. *Biochim. Biophys. Acta Gen. Subjects* **1568**, 1–6 (2001).
32. Murashima, K., Kosugi, A. & Doi, R.H. Thermostabilization of cellulosomal endoglucanase EngB from Clostridium cellulovorans by *in vitro* DNA recombination with non-cellulosomal endoglucanase EngD. *Mol. Microbiol.* **45**, 617–626 (2002).
33. Bulter, T. *et al.* Functional expression of a fungal laccase in *Saccharomyces cerevisiae* by directed evolution. *Appl. Environ. Microbiol.* **69**, 987–995 (2003).
34. Wu, Z.L., Podust, L.M. & Guengerich, F.P. Expansion of substrate specificity of cytochrome P450 2A6 by random and site-directed mutagenesis. *J. Biol. Chem.* **280**, 41090–41100 (2005).
35. Miyazaki, K. *et al.* Thermal stabilization of *Bacillus subtilis* family-11 xylanase by directed evolution. *J. Biol. Chem.* **281**, 10236–10242 (2006).
36. Sheedy, C., Yau, K.Y.F., Hirama, T., MacKenzie, C.R. & Hall, J.C. Selection, characterization, and CDR shuffling of naive llama single-domain antibodies selected against auxin and their cross-reactivity with auxinic herbicides from four chemical families. *J. Agric. Food Chem.* **54**, 3668–3678 (2006).
37. Innis, M.A., Myambo, K.B., Gelfand, D.H. & Brow, M.A. DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl. Acad. Sci. USA* **85**, 9436–9440 (1988).
38. Judo, M.S., Wedel, A.B. & Wilson, C. Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Res.* **26**, 1819–1825 (1998).

© 2006 Nature Publishing Group http://www.nature.com/natureprotocols