

Molecular evolution by staggered extension process (StEP) in vitro recombination

Huimin Zhao, Lori Giver, Zhixin Shao, Joseph A. Affholter¹, and Frances H. Arnold*

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125. ¹Materials R&D—Biocatalysis, The Dow Chemical Company, 1707 Building, Midland, MI 48674. *Corresponding author (e-mail: frances@cheme.caltech.edu).

Received 5 November 1997; accepted 23 January 1998

We have developed a simple and efficient method for in vitro mutagenesis and recombination of polynucleotide sequences. The staggered extension process (StEP) consists of priming the template sequence(s) followed by repeated cycles of denaturation and extremely abbreviated annealing/polymerase-catalyzed extension. In each cycle the growing fragments anneal to different templates based on sequence complementarity and extend further. This is repeated until full-length sequences form. Due to template switching, most of the polynucleotides contain sequence information from different parental sequences. The method is demonstrated by the recombination of two genes encoding thermostable subtilisins carrying two phenotypic markers separated by 113 base pairs and eight other point mutation markers. To demonstrate its utility for directed evolution, we have used StEP to recombine a set of five thermostabilized subtilisin E variants identified during a single round of error-prone PCR mutagenesis and screening. Screening the StEP-recombined library yielded an enzyme whose half-life at 65°C is 50 times that of wild-type subtilisin E.

Key words: directed evolution, random mutagenesis, subtilisin

Homologous recombination is a ubiquitous process that plays an important role in species adaptation and survival. Its importance is illustrated by its duality of functions—increasing genetic diversity in populations by reshuffling genes and preserving genetic integrity by aiding in the repair of damaged genes^{1,2}. Computer simulations have shown that recombination with a low level of point mutation is efficient for the evolution of complex linear sequences^{3,4}. Natural in vivo recombination mechanisms, however, usually operate at low efficiencies, eliciting insignificant changes in gene structures or functions even after tens of generations. Furthermore, in vivo recombination in most organisms is cumbersome and difficult to adapt to the redesign of genes, operons, or pathways.

Various approaches have been developed to mimic and accelerate nature's recombination strategy to direct the evolution of protein function⁵. In vitro recombination methods generally offer higher recombination efficiencies and greater experimental flexibility than in vivo approaches. In the widely used "DNA shuffling" method developed by Stemmer^{7,8}, a set of parent genes is digested with DNase I to create a pool of short DNA fragments that are reassembled into full-length genes by repeated thermocycling in the presence of DNA polymerase.

We describe a new approach to in vitro recombination that is both technically simple and conceptually novel. Rather than reassembling recombined genes from a fragment pool, our method prepares full-length recombined genes in the presence of the template(s) by what we call the "staggered extension" process (StEP). StEP consists of priming the template sequences followed by repeated cycles of denaturation and extremely abbreviated annealing/polymerase-catalyzed extension (Fig. 1). In each cycle the growing fragments can anneal to different templates based on sequence complementarity and extend further to create "recombination cassettes." Due to the template switching, the growing polynucleotides contain sequence information from different parental genes. StEP is continued until full-length genes are formed. It can be followed by a gene amplification step, if desired. The whole process can be performed using flanking universal primers.

We have assessed the degree and efficiency of StEP recombination by recombining two genes encoding thermostable *Bacillus subtilis* subtilisin E variants, each of which carries a single thermostable mutation along with several other neutral mutations. We have also demonstrated the utility of the StEP method for directed evolution of a mesophilic subtilisin into its thermophilic counterpart.

Results and discussion

StEP recombination between two thermostable subtilisin E genes. Two thermostable subtilisin E mutants, *RC1* and *RC2*, were used to test the recombination efficiency of the StEP method. The 10 positions at which these genes differ from one another are shown in Table 1. Only those mutations leading to amino acid substitutions Asn181 → Asp (N181D) and Asn218 → Ser (N218S) confer thermostability; the remaining mutations are neutral⁹. Single variants N181D and N218S have half-lives approximately threefold and twofold longer, respectively, than wild type subtilisin E at 65°C. The variant containing both mutations has an eightfold longer half-life. Thus N218S and N181D are convenient phenotypic markers for recombination events. Recombination between these positions only 113 bp apart can be measured easily by phenotypic analysis of a small sampling from the recombined variant library¹⁰. If the point mutagenesis rate is very low, 25% of the recombined population should exhibit wild-type-like stability, 25% of the population should have double mutant (N181D+N218S)-like stability and the remaining 50% should have single mutant (N181D or N218S)-like stability. Additional information on recombination efficiency can be obtained by sequencing a small sampling of the recombined library.

An equimolar mixture of plasmid DNAs containing the *RC1* and *RC2* genes was subjected to StEP recombination with two flanking primers located 45 nucleotides before the first codon and 113 nucleotides after the stop codon of the mature sequence. The progress of staggered extension was monitored by removing 10 μ l aliquots from the reaction tube at various time points and separating the DNA fragments by agarose gel electrophoresis. The average

size of the smear increased gradually with increasing cycle number (Fig. 2). The front of the smear approached 100 bp after 20 cycles, 400 bp after 40 cycles, and 800 bp after 60 cycles. Finally, a clear, discrete approximately 1 kb band appeared within the smear after 80 cycles. This band containing a mixture of recombined products was gel purified, digested with restriction enzymes BamHI and NdeI, and ligated with vector generated by BamHI-NdeI digestion of the *Escherichia coli*/*B. subtilis* pBE3 shuttle vector⁹. This gene library was amplified in *E. coli* HB101 and transferred into *B. subtilis* DB428 competent cells for expression and screening.

The overall point mutagenesis rate associated with StEP recombination can be estimated from the catalytic activity profile of a small sampling of the recombined variant library^{11,12}. The relationship between the point mutagenesis rate and the fraction of the library that encodes enzymes with activity significantly lower than wild type (<10%) is known¹³. Catalytic activities of enzyme variants were measured in a 96-well plate format⁹. Of the 368 clones screened, about 84% retained subtilisin activity. This level of inactivation corresponds to a point mutation rate of roughly 0.1% (a rate commonly observed when using *Taq* polymerase).

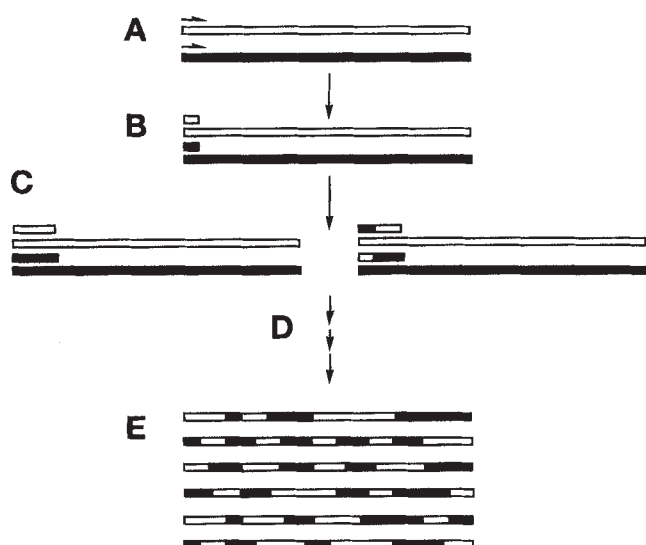


Figure 1. StEP recombination. Only one primer and single strands from two parent genes (templates) are shown. (A) Denatured template genes are primed with one defined primer. (B) Short fragments are produced by brief polymerase-catalyzed primer extension. (C) Through another cycle of StEP, fragments randomly prime the templates (template switching) and extend further. (D) This process is repeated until full-length genes are produced. (E) Full-length genes are purified and (optionally) amplified in a PCR reaction with external primers.

Table 1. DNA and amino acid substitutions in thermostable subtilisin E variants *RC1* and *RC2*.

Gene	Base	Base substitution	Position in codon	Amino acid substitution	Amino acid
<i>RC1</i>	1107	A→G	2	218	Asn→Ser
	1141	A→T	3	229	synonymous
	1153	A→G	3	233	synonymous
<i>RC2</i>	484	A→G	3	10	synonymous
	520	A→T	3	22	synonymous
	598	A→G	3	48	synonymous
	731	G→A	1	93	Val→Ile
	745	T→C	3	97	synonymous
	995	A→G	1	181	Asn→Asp
	1189	A→G	3	245	synonymous

Genes also contain base substitution A→G at position 780 relative to wild type.

The thermostabilities of the active clones are shown in Figure 3. With regards to thermostability, approximately 18% were similar to wild-type subtilisin E, 21% were comparable to the N181D+N218S double mutant, and 61% were comparable to the single mutants (N181D or N218S). This distribution was similar to that expected for random recombination of the two phenotypic markers separated by 113 bp.

The recombination efficiency was further analyzed by sequencing genes from 10 randomly selected clones. All 10 genes were novel recombinants, different from the parent genes (Fig. 4). The frequency of occurrence of any particular point mutation (marker) from parent *RC1* or *RC2* in the recombined genes ranged from 20% to 70% (fluctuating around the expected value of 50%), which indicated that the template-switching events during StEP were reasonably random. The minimum number of crossovers required to generate each chimeric gene ranged from one to four. A certain degree of linkage was apparent for mutations that are close together. The two closest, yet separable markers among this small sampling are 34 bp apart (positions 1107 and 1141 in clone 7). No significant differences in recombination efficiency were seen, compared with other in vitro recombination methods. When these two genes were recombined using the DNA shuffling method, for example, the minimum number of crossovers ranged from one to four⁹; it ranged from one to six for recombination using the random-priming technique¹⁰. None of the three methods was able to efficiently recombine the most closely spaced mutations. Six new point mutations were identified in the 10 genes. This mutation rate (0.06%) is close to that estimated from the frequency of active clones (*vide supra*).

Recombination is also observed when a pool of homologous templates is amplified by the PCR¹³⁻¹⁶. However, our PCR control experiments indicated that the efficiency of recombination by this route is very low. Ten nanograms of an equimolar mixture of plasmids containing genes *RC1* and *RC2* were used as templates in two separate conventional PCRs, one with a low annealing temperature (45°C) and the

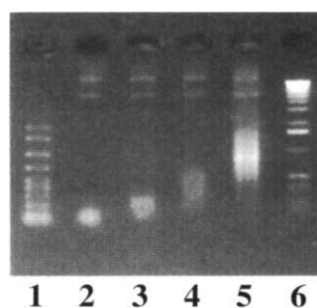


Figure 2. Agarose electrophoresis gel showing the progress of recombining two thermostable subtilisin E genes *RC1* and *RC2* by StEP. Lane 1: Amp#Size DNA Size standards (Bio-Rad, Hercules, CA), from top to bottom: 2000, 1500, 1000, 700, 500, 400, 300, 200, 100, and 50 bp; lane 2: after 20 cycles; lane 3: after 40 cycles; lane 4: after 60 cycles; lane 5: after 80 cycles; lane 6: 1 kb ladder from Boehringer Mannheim.

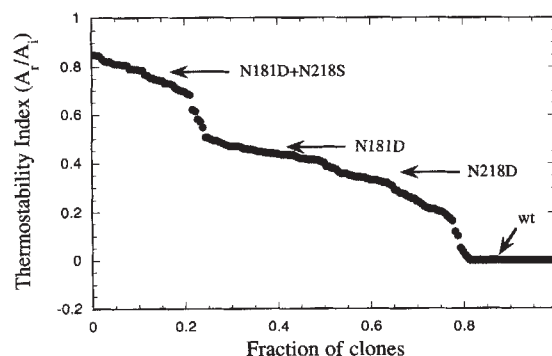


Figure 3. Results of screening 368 variants from the recombined library for activity after incubation at 65°C for 10 min (initial activity, A_i) and 40 min (residual activity, A_r). The ratio A_r/A_i (thermostability index) was used to estimate thermostability*. Data from active variants are sorted and plotted in descending order.

RESEARCH

other with an annealing temperature of 60°C. Gene amplifications were conducted with *Taq* polymerase for 25 cycles: 1 min at 94°C, 1 min at 45°C or 60°C, and 1 min at 72°C. One hundred eighty-four clones from the two libraries were screened for thermostability. The fraction of clones expressing active subtilisin E increased from 53% to 80% when the annealing temperature was raised from 45°C to 60°C. This significant change most likely reflects an improvement in polymerase specificity at the higher temperature. However, among these active variants, the frequency of recombination of the two phenotypic markers in *RC1* and *RC2* (variants with wild-type-like stability plus those with double-mutant-like stability) decreased from 11% at 45°C (4% for double-mutant-like variants plus 7% for wild-type-like variants) to 5% at 60°C (2% for double-mutant-like variants plus 3% for

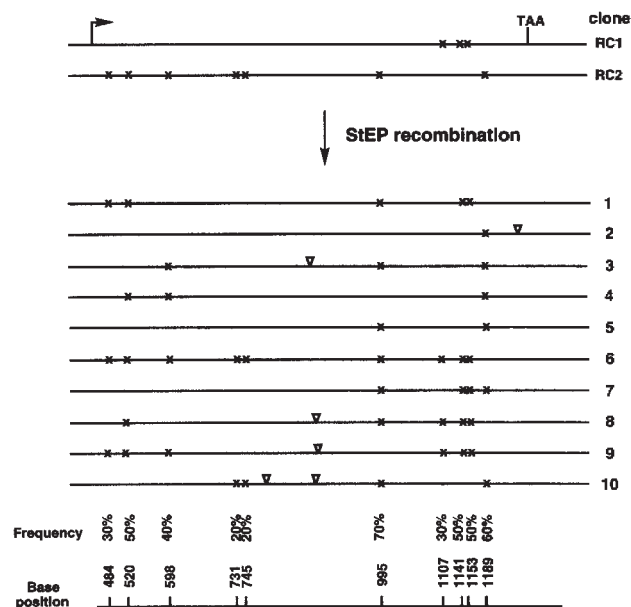


Figure 4. Sequence analysis of unscreened StEP-recombined gene libraries. Lines represent 986 bp of subtilisin E gene including 45 nucleotides of its prosequence, the entire mature sequence, and 113 nucleotides after the stop codon. X: positions of mutations from parent genes *RC1* and *RC2*; ∇: positions of new point mutations introduced during StEP. The four new point mutations in the center of genes 3, 8, 9, and 10 are located at different positions.

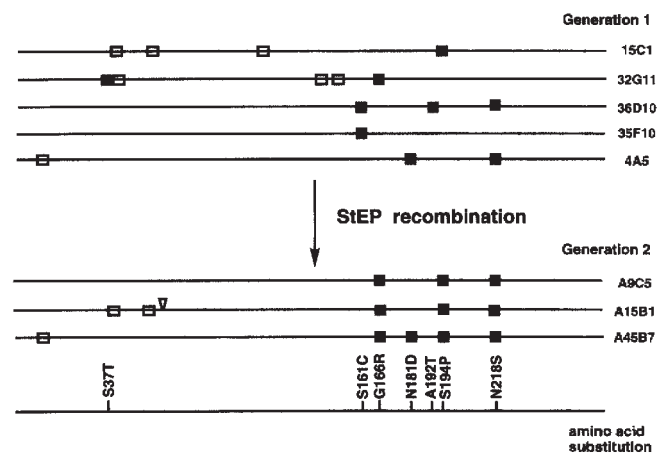


Figure 5. Sequence analysis of first- and second-generation thermostable subtilisin E variants. Filled squares indicate positions of nonsynonymous mutations, empty squares indicate positions of synonymous mutations, and triangles indicate positions of new point mutations introduced during StEP recombination.

wild-type-like variants). From the library generated at an annealing temperature of 45°C, genes from one variant exhibiting double-mutant-like stability and one variant exhibiting wild-type-like stability were sequenced. The gene with wild-type-like stability had neither the N218S nor N181D mutation, while the thermostable gene had both. No new point mutations were found (data not shown). Recombinants are postulated to arise from incomplete extension of the annealed primer during one cycle when the polymerase pauses or prematurely disassociates from the template strand for unknown reasons¹³⁻¹⁶. This phenomenon shares the same recombination mechanism as StEP. In StEP recombination, however, the partial extension of the annealed primer deliberately controlled by the very brief annealing/extension steps greatly enhances recombination efficiency.

The key to successful recombination by StEP is to tightly control the polymerase-catalyzed DNA extension. Denaturation is followed by random annealing of the extended fragments to template sequences and continued partial extension. This process is repeated multiple times, depending on the concentration of primer and template, until full-length sequences are made. Too much extension during each cycle severely limits recombination events. The annealing/extension step is therefore carried out under conditions that allow high fidelity primer annealing ($T_{\text{annealing}} > T_m - 25^\circ\text{C}$) but limit polymerization/extension (no more than a few seconds). Thermostable DNA polymerases typically exhibit maximal polymerization rates of 100–150 nucleotides/s at optimal temperatures and follow approximate Arrhenius kinetics at temperatures approaching the optimum temperature. Thus, at 55°C, a thermostable polymerase may exhibit only 20–25% of its steady-state polymerization rate at 72°C, or approximately 24 nucleotides/s. At 37°C and 22°C, *Taq* polymerase is reported to have extension activities of only 1.5 and 0.25 nucleotides/s, respectively¹⁷. The time and temperature of DNA polymerization must be optimized, depending on the template genes, polymerase, and particular reaction conditions (e.g., thermocycler used), to obtain the desired degree of recombination.

Unlike gene amplification (which generates new DNA exponentially), StEP generates new DNA linearly in its early cycles. In StEP, the ratio of primer to template is usually between 100 and 500, as compared with 10^6 in a typical gene amplification process. When significant numbers of primer-extended DNA molecules begin to reach sizes of more than one-half the length of the full-length gene, a rapid jump in molecular weight occurs, as half-extended forward and reverse strands begin to cross-hybridize to generate fragments nearly twice the size of those encountered to that point. Rapid consolidation of the smear into a discrete band of the appropriate molecular weight occurs either by continuing StEP or by altering the thermocycling program (by increasing the extension time or optimizing the extension temperature) to allow complete extension of the primed DNA and drive exponential gene amplification.

Directed evolution of a thermostable subtilisin E. Directed evolution involves the generation and selection or screening of molecular repertoires with sufficient diversity for the altered function to be represented. This “irrational” design approach has proven particularly effective for exploring and optimizing enzyme functions⁴. An effective directed-evolution strategy is to generate molecular diversity by error-prone PCR at a low error rate (two to three mutations per gene) and select or screen variants that show improvement with respect to the desired feature¹⁸. It is usually the case that several positive variants are identified after one round of selection or screening. Using in vitro recombination, beneficial mutations from these variants can be accumulated rapidly while the deleterious mutations are removed^{7,8,19}.

Subtilisin E is a protease produced by the mesophile *B. subtilis*. At 65°C, pH 8.0, and in the presence of 1 mM CaCl₂, the half-life of wild-type subtilisin E is approximately 5 min⁹. Our long-term goal was to convert this enzyme into its equivalent hyper-thermostable counterpart by directing its evolution in vitro. Mutagenesis by error-prone PCR on the mature subtilisin E gene and screening yielded five variants

with half-lives three to eight times greater than wild type at 65°C. Equal amounts of the genes from the five variants were recombined by StEP. Approximately 8000 clones expressing genes from the recombined library were screened for thermostability at 75°C. Three variants with the highest thermostabilities were identified and sequenced. Their half-lives at 65°C were 25–50 times that of wild type.

The DNA sequences of the five first-generation thermostable variants and the three thermostable variants obtained by StEP recombination are summarized in Figure 5. From inspection of the sequences of the first-generation variants, we can conclude that amino acid substitutions S194P and S161C are responsible for the enhanced thermostability of 15C1 and 35F10, respectively, as these are the only nonsynonymous mutations in their genes. Although synonymous mutations may affect expression, they are not expected to influence thermostability. Mutations leading to N181D (4A5) and N218S (36D10 and 4A5) have been confirmed previously to be thermostabilizing⁹. The remaining two variants, 32G11 and 36D10, each have more than one nonsynonymous mutation; the effects of S37T, G166R, and A192T are therefore not clear from the first generation sequences.

Among the second-generation variants, the most thermostable is A45B7, with a half-life 50 times that of wild-type subtilisin E. A45B7 contains four nonsynonymous mutations: G166R, N181D, S194P, and N218S, and the two less-thermostable variants A9C5 and A15B1 (half-life approximately 25–30 times that of wild type) contain only G166R, N181D, and N218S. The presence of G166R in all three indicates that it contributes to the observed thermostabilization. The absence of S37T and A192T in all three sequences supports the contention that these mutations are neutral, if not slightly deleterious. There are two possible explanations for our failure to identify a second-generation variant with known beneficial mutation S161C. The first possibility is that S161C does not contribute sufficiently to thermostability in the background of the other mutations. The second is that S161C is too close to G166R (only 15 bp apart) to recombine, and therefore that recombinant thermostable genes will contain one or the other, but not both. If G166R contributes more to thermostability than S161C, it will appear preferentially in the screened population. Based on these arguments, we conclude that mutations S161C, G166R, S194P, N181D, and N218S are thermostabilizing.

The efficiency of StEP recombination is similar to other in vitro recombination methods. However, the StEP recombination reaction can be carried out in a single tube; separation of parent templates from the recombined products is not necessary. StEP is in some ways reminiscent of the template-switching recombination mechanism that contributes to the evolutionary potential of retroviral populations²⁰. The simple and efficient StEP recombination method provides a powerful new tool that can be applied to directed evolution of genes, operons, pathways, and even whole bacterial or viral genomes for specific applications.

Experimental protocol

Enzymes. Restriction enzymes were purchased from Boehringer Mannheim (Indianapolis, IN). Succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide (s-AAPF-pNa) was from Sigma (St. Louis, MO).

Staggered extension process (StEP). 5' and 3' flanking primers⁹ P5N (5'-CCGAG CGTTG CATAT GTGGA AG-3', underlined sequence is NdeI restriction site) and P3B (5'-CGACT CTAGA GGATC CGATT C-3', underlined sequence is BamHI restriction site) were used for recombination. StEP conditions (100 µl final volume): 0.15 pmol (total) plasmid DNAs (pBE3 containing the subtilisin E genes⁹) were used as templates, 30 pmol of each primer, 1× *Taq* buffer, 0.2 mM of each dNTP, 1.5 mM MgCl₂, and 2.5 U *Taq* polymerase (Promega, Madison, WI). Program: 5 min of 95°C, 80 cycles of 30 sec 94°C, and 5 s 55°C. StEP was performed in an MJ Research (Watertown, MA) PTC-200 thermocycler. A clear, discrete electrophoretic band of the correct size (about 1 kb) among smears is typically obtained. All plasmids were isolated and purified from *E. coli* HB101 using QIAprep spin plasmid miniprep kit (Qiagen, Chatsworth, CA).

Cloning, expression, and thermostability screening. The product of correct size (about 1 kb) was cut from a 0.8% agarose gel after electrophoresis of

the whole reaction mixture and purified using QIAEX II gel extraction kit (Qiagen). This purified product (approximately 300 ng) was subjected to a standard restriction-digestion reaction by Nde I and BamHI (a total of 20 µl of reaction volume), followed by electrophoresis. The DNA of correct size (about 1 kb) was again cut from the 0.8% agarose gel and purified using QIAEX II gel extraction kit (Qiagen). Subsequent cloning, expression, and thermostability screening were carried out as described⁹. For recombination of thermostable subtilisins E genes RC1 and RC2, initial and residual activities were measured after incubation at 65°C for 5 and 20 min, respectively. Recombinants generated by StEP from five thermostable first-generation subtilisin E mutants were screened by measuring their activities after incubation at 75°C for 5 and 15 min.

Error-prone PCR. Random mutagenesis of subtilisin E genes was performed under conditions similar to those described^{11,21}. Primers P5N and P3B were used to amplify approximately 1 kb fragments including a partial prosequence (a length of 15 residues), mature subtilisin E gene and 113 nucleotides after the stop codon. The PCR reaction contained (100 µl final volume): 10 mM Tris (pH 8.3 at 25°C), 50 mM KCl, 7 mM MgCl₂, 0.01% (wt/vol) gelatin, 0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP, 1 mM TTP, 0.15 mM MnCl₂, 0.3 µM of both primers, 5 ng of template, and 5 U *Taq* DNA polymerase (Promega). No mineral oil was overlaid because the lid of thin PCR tube was pre-heated. PCR was performed in an MJ Research PTC-200 thermocycler for 13 cycles: 1 min at 94°C, 1 min at 50°C, and 1 min at 72°C. The PCR products were purified using Wizards PCR Preps (Promega), followed by restriction digestion by Nde I and BamHI. These digestion products were purified again using Wizards PCR Preps. Cloning, expression, thermostability screening, enzyme purification, and DNA sequencing were carried out as described⁹.

Acknowledgments

This work was supported by the US Department of Energy's program in Biological and Chemical Technologies Research within the Office of Industrial Technologies, Energy Efficiency and Renewables. The authors are grateful to Kentaro Miyazaki for helpful discussions and Yongkai Ow for her technical assistance.

1. Stahl, F.W. 1987. Genetic recombination. *Sci. Am.* **256**:91–101.
2. Kucherlapati, R. and Smith, G.R. 1988. *Genetic recombination*. American Society of Microbiology, Washington, DC.
3. Holland, J.H. 1992. Genetic algorithms. *Sci. Am.* **267**:66–72.
4. Holland, J.H. 1992. *Adaptation in natural and artificial systems*. 2nd ed. MIT Press, Cambridge, MA.
5. Forrest, S. 1993. Genetic algorithms: Principles of natural selection applied to computation. *Science* **261**:872–878.
6. Kuchner, O. and Arnold, F.H. 1997. Directed evolution of enzyme catalysts. *Trends Biotech.* **15**:523–530.
7. Stemmer, W.P.C. 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**:389–391.
8. Stemmer, W.P.C. 1994. DNA shuffling by random fragmentation and reassembly - In vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**:10747–10751.
9. Zhao, H. and Arnold, F.H. 1997. Functional and nonfunctional mutations distinguished by random recombination of homologous genes. *Proc. Natl. Acad. Sci. USA* **94**:7997–8000.
10. Shao, Z., Zhao, H., Giver, L., and Arnold, F.H. 1998. Random-priming in vitro recombination: an effective tool for directed evolution. *Nucl. Acids Res.* **26**:681–683.
11. Shafikhani, S., Siegel, R.A., Ferrari, E., and Schellenberger, V. 1997. Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. *Biotechniques* **23**:304–310.
12. Zhao, H. and Arnold, F.H. 1997. Optimization of DNA shuffling for high-fidelity recombination. *Nucl. Acids Res.* **25**:1307–1308.
13. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T. et al. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**:487–491.
14. Scharf, S.J., Friedmann, A., Brautbar, C., Szafer, F., Steinman, L., Horn, G. et al. 1988. HLA class II allelic variation and susceptibility to pemphigus vulgaris. *Proc. Natl. Acad. Sci. USA* **85**:3504–3508.
15. Scharf, S.J., Long, C.M., and Erlich H.A. 1988. Sequence analysis of the HLA-DRβ and HLA-DQβ loci from 3 pemphigus vulgaris patients. *Hum. Immunol.* **22**:61–69.
16. Bradley, R.D. and Hillis, D.M. 1997. Recombinant DNA sequences generated by PCR amplification. *Mol. Biol. Evol.* **14**:592–593.
17. Gelfand, H.A. 1989. *Taq* DNA polymerase, p. 18 in *PCR technology*. Ehrlich, H.A. (ed.). Stockton Press, New York.
18. Moore, J.C. and Arnold, F.H. 1996. Directed evolution of a para-nitrobenzyl esterase. *Nature Biotechnology* **14**:458–467.
19. Moore, J.C., Jin, H.M., Kuchner, O., and Arnold, F.H. 1997. Strategies for the in vitro evolution of protein function: enzyme evolution by random recombination of improved sequences. *J. Mol. Biol.* **272**:336–347.
20. Hu, W.-S., Bowman, E.H., Delviks, K.A., and Pathak, V.K. 1997. Homologous recombination occurs in a distinct retroviral subpopulation and exhibits high negative interference. *J. Virol.* **71**:6028–6036.
21. Cadwell, R.C. and Joyce, G.F. 1994. Mutagenic PCR. *PCR Methods and Applications* **2**:28–33.