

# Random-priming *in vitro* recombination: an effective tool for directed evolution

Zhixin Shao, Huimin Zhao, Lori Giver and Frances H. Arnold\*

Division of Chemistry and Chemical Engineering 210-41, California Institute of Technology, Pasadena, CA 91125, USA

Received September 15, 1997; Revised and Accepted November 21, 1997

## ABSTRACT

**A simple and efficient method for *in vitro* mutagenesis and recombination of polynucleotide sequences is reported. The method involves priming template polynucleotide(s) with random-sequence primers and extending to generate a pool of short DNA fragments which contain a controllable level of point mutations. The fragments are reassembled during cycles of denaturation, annealing and further enzyme-catalyzed DNA polymerization to produce a library of full-length sequences. Screening or selecting the expressed gene products leads to new variants with improved functions, as demonstrated by the recombination of genes encoding different thermostable subtilisins in order to obtain enzymes more stable than either parent.**

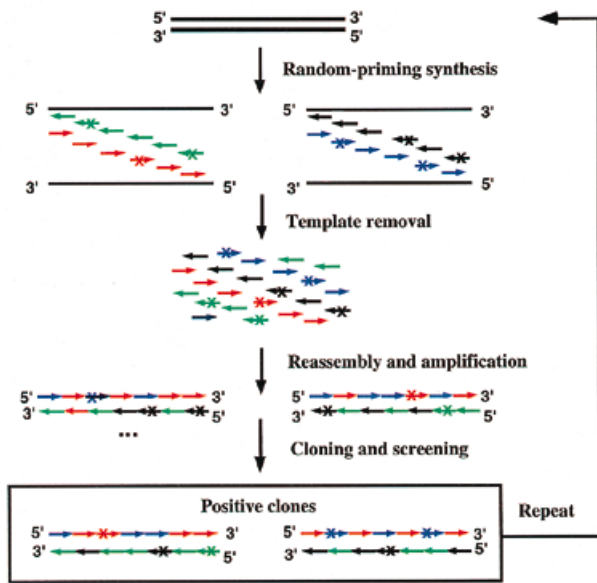
Directed evolution (1) has proven particularly effective for exploring and optimizing enzyme functions (2–5). Including recombination in the creation of gene libraries allows the rapid accumulation of beneficial mutations and removal of deleterious ones (2–4). The sequence diversity upon which recombination acts can be obtained by mutating the wild-type gene or, alternatively, by using homologous genes isolated from nature or obtained in parallel laboratory evolution experiments (e.g. genes optimized for different properties). Stemmer (2,3) recently introduced the ‘DNA shuffling’ technique for random *in vitro* mutagenesis and recombination. Here we present a new method, random-priming recombination (RPR), as an effective alternative to DNA shuffling. As shown in Figure 1, random sequence primers are used to generate a large number of short DNA fragments complementary to different sections of the template sequence(s). Due to base misincorporation and mispriming, these short DNA fragments also contain a low level of point mutations. The short DNA fragments can prime one another based on homology, and be recombined and reassembled into full-length genes by repeated thermocycling in the presence of thermostable DNA polymerase. These sequences can be further amplified by conventional PCR and cloned into a vector for expression, followed by screening or selection. RPR and screening or selection can be repeated over multiple cycles in order to evolve the desired properties.

Compared to DNA shuffling (2,3), the RPR technique has several advantages: (i) RPR can use single-stranded polynucleotide

templates without an intermediate step of synthesizing the whole second strand. Potential mutations and/or crossovers can be introduced at the DNA level from single- or double-stranded DNA template by using DNA polymerases, or directly from mRNA by using RNA-dependent DNA polymerases. (ii) DNA shuffling requires fragmentation of the double-stranded DNA template (generally done with DNase I). The DNase I must be removed completely before the fragments can be reassembled into full length sequences. Gene reassembly is generally easier with the RPR technique, which employs random priming synthesis to obtain the short DNA fragments. Furthermore, since DNase I hydrolyzes double-stranded DNA preferentially at sites adjacent to pyrimidine nucleotides (6), its use in template digestion may introduce a sequence bias into the recombination. (iii) The synthetic random primers are uniform in their length and lack sequence bias. The sequence heterogeneity allows them to form hybrids with the template DNA strands at many positions, so that, at least in principle, every nucleotide of the template should be copied or mutated at a similar frequency during extension. The random distribution of the short, nascent DNA fragments along the template(s) and the random distribution of point mutations within each nascent DNA fragment should guarantee the randomness of crossovers and mutations in the full length progeny genes. (iv) The random-priming DNA synthesis is independent of the length of the DNA template. DNA fragments as small as 200 bases can be primed equally well as large DNA molecules such as linearized plasmids or  $\lambda$  DNA (7). This offers unique potential for engineering small peptides. Hodgson and Fisk (8) found that the average size of synthesized single-stranded DNA is an inverse function of primer concentration. Based on this guideline, proper conditions for random-priming synthesis can be easily set for a given gene. (v) Since the parent polynucleotide serves solely as the template for the synthesis of nascent, single-stranded DNA, 10–20 times less parent DNA is needed as compared to DNA shuffling.

We demonstrated the RPR method by the mutagenesis and recombination of genes RC1 and RC2 encoding thermostable *Bacillus subtilis* subtilisin E variants (Fig. 2). The mutations at base positions 1107 in RC1 and 995 in RC2, giving rise to amino acid substitutions Asn<sub>218</sub>→Ser (N218S) and Asn<sub>181</sub>→Asp (N181D), lead to improvements in subtilisin E thermostability; the remaining mutations, both synonymous and non-synonymous, have no detectable effects on thermostability. At 65°C, the single variants

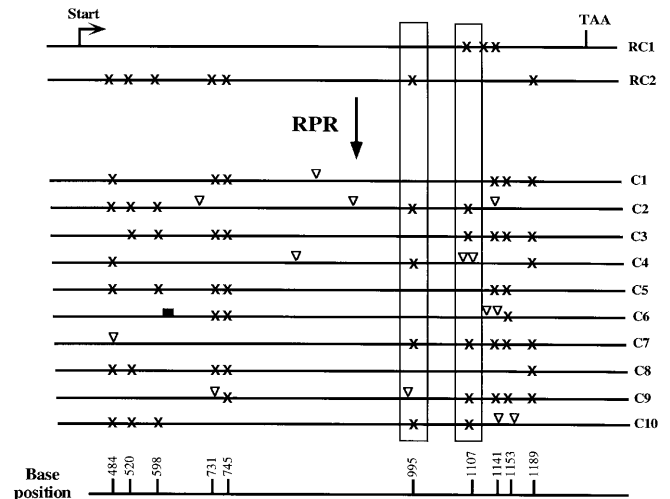
\*To whom correspondence should be addressed. Tel: +1 626 395 4162; Fax: +1 626 568 8743; Email: frances@cheme.caltech.edu



**Figure 1.** Random-priming *in vitro* recombination (RPR). (i) Synthesis of short single-stranded DNA fragments from random-sequence primers. X's indicate newly introduced mutations. (ii) Template removal. (iii) Reassembly and amplification. (iv) Cloning and screening (or selection). (v) Cycle is repeated until desired functional improvement is achieved.

N181D and N218S have ~3-fold and 2-fold longer half-lives, respectively, than wild-type subtilisin E, and variants containing both mutations have half-lives that are 8-fold longer (9). The half-lives of a population of subtilisin E variants can therefore be used to estimate the recombination efficiency. In particular, random recombination between these two mutations (in the absence of point mutations affecting thermostability) should generate a library in which 25% of the population exhibits the thermostability of the double mutant. Similarly, 25% of the population should exhibit wild-type like stability, as N181D and N218S are eliminated at equal frequency.

For random-priming synthesis using the Klenow fragment, 200 ng (0.7 pmol) of genes RC1 and RC2 (1:1) was mixed with 13.25 µg (6.7 nmol) of dp(N)<sub>6</sub> random primers (Pharmacia Biotech Inc., Piscataway, NJ). After denaturation at 100°C for 5 min, 10 µl of 10 × reaction buffer (900 mM HEPES, pH 6.6, 0.1 M MgCl<sub>2</sub>, 20 mM dithiothreitol and 5 mM each dATP, dCTP, dGTP and dTTP) was added, and the total volume of the reaction mixture was brought to 95 µl with H<sub>2</sub>O. Ten units (in 5 µl) of the Klenow fragment of *Escherichia coli* DNA polymerase I (Boehringer Mannheim, Indianapolis, IN) was added. The synthesis reaction was carried out at 22°C for 3 h and terminated by cooling the sample to 0°C. The high primer concentration facilitates production of shorter fragments, presumably by blocking extension. Under the reaction conditions described here (10 000-fold excess primer over template) the strand displacement activity of the Klenow is much reduced, and the random priming products are ~50–500 bases as determined by alkaline agarose gel electrophoresis. After the addition of 100 µl of ice-cold H<sub>2</sub>O to the reaction mixture, the random primed products were purified by passing the whole reaction mixture through a



**Figure 2.** Sequence analysis of 10 randomly-picked clones from (unscreened) *B. subtilis* subtilisin E RPR library. Each line represents 986-bp subtilisin E gene including the 45 bp prosequence, the entire mature sequence and 113 bp after the stop codon. Triangles (▽) indicate positions of new point mutations introduced during RPR. Box (■) in C6 indicates the short stretch containing four mutations between two repeats (5'-CGGTACG-3'). Positions and frequency of mutations in the 10 clones are indicated.

Microcon-100 filter (Amicon, Inc., Beverly, MA) to remove the template, proteins and large nascent DNA fragments, followed by a Microcon-10 filter to remove the primers and fragments <30 bases. The retentate fraction (~65 µl) recovered from the Microcon-10 step was buffer-exchanged against PCR reaction buffer. Of this fraction, 10 µl was used for the whole gene reassembly as described (9). Subtilisin E mutant gene amplification, cloning, expression and thermostability screening and analysis were carried out as described (9).

We measured the thermoinactivation rate at 65°C of subtilisins E from ~400 randomly-picked clones. Approximately 26% of the active clones exhibited thermostability comparable to the N181D + N218S double mutant, indicating that the RPR had efficiently recombined the N181D mutation from RC2 and the N218S mutation from RC1. Sequence analysis of the clone exhibiting the highest thermostability showed that mutations N181D and N218S were both present (data not shown). To further characterize the recombination efficiency and point mutagenesis associated with the RPR process, 10 clones from the mutant library were selected at random and sequenced. As summarized in Figure 2, all 10 clones were novel recombinants, different from the parent genes. The frequency of occurrence of any particular point mutation from parent RC1 or RC2 in the recombined genes (C1–C10) ranged from 40 to 70%, fluctuating around the expected value of 50%. The minimum number of crossovers ranged from two (C4) to six (C1). Figure 2 also shows that all 10 mutations in the parent genes can be recombined, including mutations only 12 bp apart (positions 1141 and 1153 in C6).

As shown in Figure 2, 18 new point mutations were found in the 10 random clones. This error rate of ~0.18% (1–2 new point mutations per gene) is close to ideal for directed evolution (4). The new mutations appear randomly distributed along the gene, except for a cluster of four mutations within a very short stretch of DNA

in clone C6 (see Fig. 2). The direction of mutation, however, is clearly non-random: A changes more often to G than to T or C. Transitions (in particular T→C and A→G) occur more often than transversions. One G→C, one C→G and one C→A transversion were found within the 10 sequenced clones. These mutations were generated much more rarely during the error-prone PCR mutagenesis of subtilisin BPN' (10). This result suggests that the RPR technique may allow access to a greater range of amino acid substitutions than PCR-based point mutagenesis.

RPR is a very flexible and easy to implement technique for generating mutant libraries for directed evolution. Other polymerases with different fidelities, including bacteriophage T4 DNA polymerase, T7 Sequenase® version 2.0 DNA polymerase, the Stoffel fragment of *Taq* polymerase and *Pfu* polymerase, have been used successfully for the random priming DNA fragment synthesis. The length and concentration of random primer, as well as the time, temperature and other reaction conditions can also be manipulated in order to achieve the desired mutagenic rate and recombination frequency.

### ACKNOWLEDGMENTS

This research is supported by Eli Lilly and Co., the US Office of Naval Research (N0014-96-1-340 and N00014-97-1-0433), the US Department of Energy's program in Biological and Chemical Technologies Research within the Office of Industrial Technologies, Energy Efficiency and Renewables. Z.S. is grateful to Drs Peggy

Arps and Kentaro Miyazaki for helpful discussions and to Michaeleen Callahan for excellent technical assistance.

### REFERENCES

- 1 Kuchner, O. and Arnold, F.H. (1997) *Trends Biotech.*, **15**, 523–530.
- 2 Stemmer, W.P.C. (1994) *Nature*, **370**, 389–391.
- 3 Stemmer, W.P.C. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 10747–10751.
- 4 Moore, J.C., Jin, H.-M., Kuchner, O. and Arnold, F.H. (1997) *J. Mol. Biol.*, **272**, 336–347.
- 5 Moore, J.C. and Arnold, F.H. (1996) *Nature Biotech.*, **14**, 458–467.
- 6 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 7 Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.*, **132**, 6–13.
- 8 Hodgson, C.P. and Fisk, R.Z. (1987) *Nucleic Acids Res.*, **15**, 6296.
- 9 Zhao, H. and Arnold, F.H. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 7997–8000.
- 10 Shafikhani, S., Siegel, R.A., Ferrari, E. and Schellenberger, V. (1997) *Biotechniques*, **23**, 304–310.

### RELATED PAPERS RECENTLY PUBLISHED IN NUCLEIC ACIDS RESEARCH

- He, M. and Taussig, M.J. (1997) Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for *in vitro* display and evolution of antibody combining sites. *Nucleic Acids Res.* **25**, 5132–5134.
- Zhao, H. and Arnold, F.H. (1997) Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res.* **25**, 1307–1308.
- Morl, M. and Schmelzer, C. (1990) Group II intron RNA-catalyzed recombination of RNA *in vitro*. *Nucleic Acids Res.* **18**, 6545–6551.