Enhanced Enzyme Performance by DNA Shuffling

Andreas Crameri*

Abstract. In the quest for enhanced enzyme performance in non-natural applications, directed evolution emulates natural processes for the generation of diversity and selection of desirable traits. Classical methods for improving protein characteristics rely upon point mutation or cassette mutagenesis of a selected region within a desired sequence (*i.e.*, focusing on a narrow sequence space). However, computer simulations of the evolution of linear sequences have demonstrated the importance of recombination of blocks of related sequences rather than sequence mutation alone.

DNA shuffling includes *in vivo* and *in vitro* methods for recombination of nucleic acid sequences. In one format, DNA shuffling involves methods for *in vitro* homologous recombination of pools of related genes. For example, fragmentation of nucleic acid sequences that encode genes, can be followed by reassembly using the polymerase chain reaction (PCR). As the complete gene sequence can be shuffled, a wider sequence space can be accessed in searching for the desired, improved variant. Repeated cycles of recombination, optionally together with error-prone PCR to introduce point mutations, allow efficient molecular evolution of complex sequences *in vitro*, to provide biological molecules with improved properties.

Introduction

Enzymes are used in a wide variety of applications including food and feed processing, laundry detergents, chemical production, paper bleaching, and pharmaceutical manufacturing. The benefits of enzymes as catalysts are that reactions occur at moderate temperatures, toxic solvents or reactants can often be eliminated, and reactions are usually stereospecific, which is of particular benefit in the synthesis of pharmaceuticals and fine chemicals.

Nature provides an enormous number of enzymes for optimized performance in a variety of natural environments. Natural enzymes used for catalysis in an industrial reaction rarely exhibit the required activity and stability. Simply enhancing microbial production of the natural enzyme is seldom sufficient to generate a commercially viable catalyst. The directed evolution of improved or altered enzymatic functions represents a new paradigm in the molecular biology of enzyme-based catalysis. Directed evolution involves the

production of novel enzymes tailor-made for new applications. However, while the total possible number of sequences that can be generated is vast, those sequences that possess the desired properties are rare.

Rational design relies on the identification and deliberate mutation of important amino acids within an enzyme of interest. Site-directed mutagenesis is typically employed to result in the exchange of these residues. Analysis of mutant variants often allows identification of important residues for future mutagenesis, confirmation of primary functional components and sequence motifs, and identification of enzymes with improvements in properties of interest. New mutants can then be generated, re-evaluated and altered, until the process delivers a commercially relevant solution. The results of such studies account for much of what we know today about protein structure and function and have provided valuable insights for the design and expression of functionally altered proteins.

The process of rational design is slow and is dependent on initial structure/function information to indicate which residues are important. Generating such information, however, and applying it to the generation of particular mutants, is extremely labour intensive. Moreover, the benefit of focusing on promising regions

of sequence space is countered by dramatically limiting the amount of actual sequence space sampled [1]. The variant with optimal characteristics may indeed lie outside the particular sequence region being examined and consequently may never be found by this method.

The directed evolution of enzymes is a fundamentally different design algorithm to that of rational protein design. In general, directed evolution processes require no previous knowledge of the enzyme sequence or protein structure nor of the molecular changes necessary to achieve a novel or improved function [2][3]. Instead, the methodology relies on powerful screening and selection methods to identify and isolate improved variants from libraries of thousands of different mutants [4][5]. Several methods have been used successfully to isolate enzymes with novel or improved function.

Directed evolution effectively performs the complex computations required to change a particular sequence that results in an altered catalytic function. In addition to the active-site geometry, the impact of sequence changes on protein expression, stability and folding, and interactions with other host proteins and small molecules are all simultaneously considered, simply by directly measuring the activity of the mutant enzyme or an effect of a metabolic

*Correspondence: Dr. A. Crameri
Biozentrum University Basel
Klingelbergstrasse 70
CH–4056 Basel
Tel.: 076 344 06 02
E-Mail:crameri@bluewin.ch
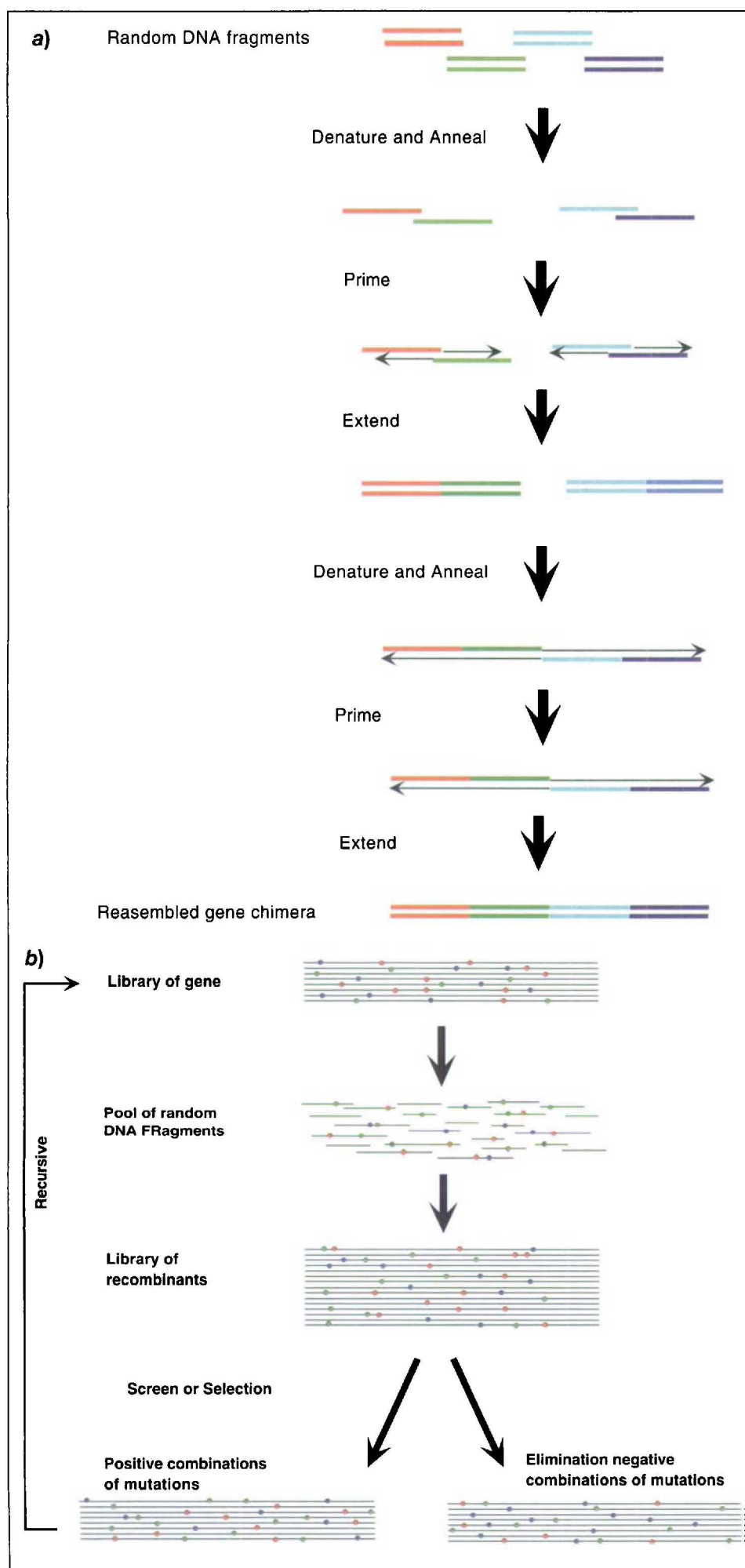
pathway. In considering the multiple effects of potential sequence changes, DNA shuffling closely mimics natural evolution, thereby improving the chances of obtaining the improved variant.

In vitro DNA Shuffling Using DNA Fragmentation

By fragmenting a pool of related DNAs and then re-assembling them using PCR methods, DNA shuffling uses recombination to take sequence diversification well beyond the limits of random mutagenesis (*Fig. 1A*) and effectively recombines the diversity of all the parental genes in the initial gene pool. By permutating the diverse sequences in the selected pool, this process creates random but functional combinations of mutations enabling one to search a much greater sequence space and to access the improved mutants that result from the additivity or synergism of the mutations (*Fig. 1B*). Additivity benefits have been demonstrated by comparing DNA shuffling and error-prone PCR by using a β-lactamase construct selected for increased resistance to moxalactame [6]. The study of *Stemmer* is of particular interest because it allows direct comparison of the products of cassette mutagenesis and DNA-shuffling techniques because DNA shuffling was carried out with the same gene as used in the cassette mutagenesis.

DNA Shuffling can also be used for backcrossing of the mutated genes with the wild-type enzyme to eliminate non-beneficial mutations. Many examples of successful directed evolution using DNA shuffling have been recently reviewed. [7–14]. Within the last year, several additional formats of DNA shuffling of genes have been published [15–17]. All of these methods rely on the general principle that

Fig. 1. In vitro *DNA shuffling. a)* Genes are randomly fragmented. This results in fragments of various lengths that, after denaturation, hybridize to form an equal mixture of 5'- and 3'-overhangs. By using PCR techniques, reassembling occurs by a reaction in which homologous fragments act as primers and template for each other. *b)* The diversity consists of the variable level of mutations that are introduced by PCR reaction (or other mutagenic methods). After the first selection cycle, the mutations are formed within a conserved framework sequence, and a controlled level of point mutations can be formed as well. When coupled with selection or screening, this process allows a rapid accumulation of useful combinations of mutants.

the most efficient way to explore sequence space is by recombination of active variants.

The initial diversity that DNA shuffling uses to access sequence space can be generated using protocols similar to those used in random mutagenesis (*e.g.*, error-prone PCR, gene assembly, cassette mutagenesis, or their combinations) but applying the techniques to the complete sequence space rather than a subsequence of interest. The selection and screening employed to evaluate the variants can be essentially the same as those used in rational design. Because of the much larger numbers of possible mutants created with DNA shuffling, high-throughput methods can be easily applied to identify the very best clone.

Family Shuffling of Naturally Occuring Diversity

Until now, we have discussed the application of DNA shuffling on a single gene and the mutations within that gene (*i.e.*, a pool of related sequences) for evolution in the laboratory (*Fig. 2A*). The genome project provides large amounts of sequence knowledge, identifying sequence homologies between sequences from different organisms which potentially encode enzymes with the same or similar function. If one could harvest this natural diversity, the jump to enzymes with altered or improved functions could be realized more efficiently than by using conventional methods [18][19].

Family shuffling achieves this, for example, by shuffling multiple related wild-type genes as starting sequences in a shuffling reaction. The ability to utilize both natural and artificial diversity allows for greater access to the functional sequence space. More importantly, the mutations present in the starting sequences have been pre-selected as mutations having functional activity. Deleterious mutations, which commonly arise in random mutagenesis, have already been removed from these populations by natural selection. As a result, the sequence space that can be accessed with these libraries is larger and likely to be comparatively rich in functional sequences.

In one example of family shuffling, four different β-lactamase genes were shuffled together to produce a chimera with 270-fold greater resistance to moxalactame than the best parental enzyme [20]. The chimeric enzyme produced in this experiment differed from each parent by at least 100 amino acids, yet was still a fully
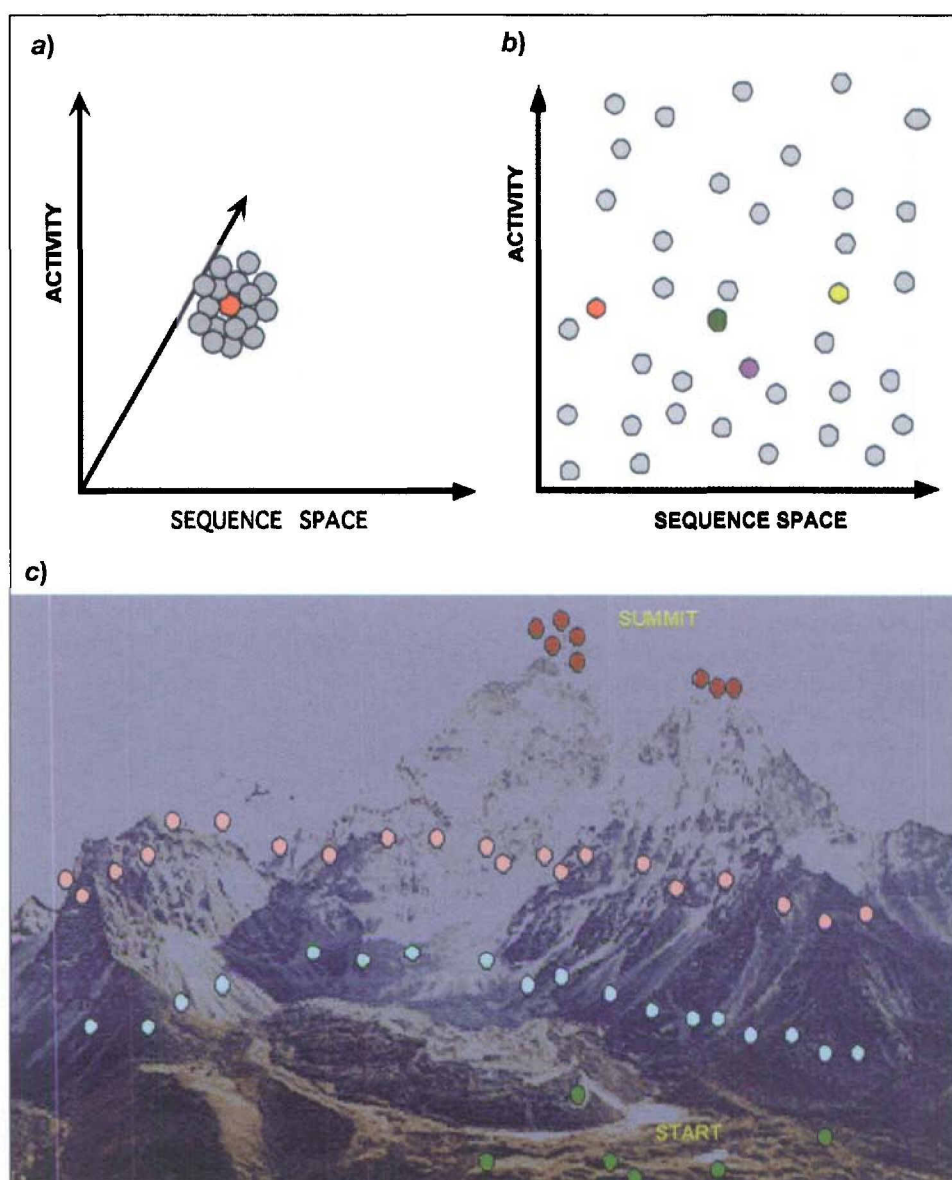

Fig. 2. *Searching sequence space.* a) Single sequence shuffling yields clones with a few point mutations, and the library members are typically 97–99% identical. Therefore, the whole library members are closely related to the parental sequence. b) Family shuffling causes sequence block exchange, which yields chimeras that have greater sequence divergence. At equal library size, the increased sequence diversity results in sparse sampling of a much greater area of sequence space, allowing more promising areas to be found and subsequently explored at increased sample density. c) This allows exploring different sequencing-space peaks and, therefore, the possibility of finding new activities increases significantly.

functional cephalosporinase. Sequences that occur in naturally existing enzymes have already been tested for their ability to function within the context of the protein's overall structure. Recombining natural blocks of sequence with each other allows a broad region of functional sequence space to be sampled sparsely (*Fig. 2B*). Christians *et al.* have performed recursive cycles of family shuffling with multiple parental sequences. By recombining two herpes simplex virus thymidine-kinase genes and robotically screening for variants that were better able to phosphorylate the therapeutic nucleotide analogue AZT, the concentration of AZT required to inhibit cell growth was re-

duced 32-fold relative to that required for the best parent. The resulting enzyme was a chimera that had undergone six crossovers between the two parental genes, and had also accumulated six point mutations, leading to a protein differing by 22 amino acids from the closest parent [21].

The process of recombination between different but functional parents to make large changes in sequence, coupled with point mutagenesis to fine-tune the activity of the protein is only achievable with family shuffling. Often, desired properties or combinations of properties do not exist among natural enzymes; this may simply be because these combinations have not been selected for in nature. For example,

activity at 23° and thermostability are unlikely to be required by the same organism. On the other hand, activity at low pH will not have been selected for in an alkalophilic species. A small library of 654 functional subtilisins created from only 26 starting sequences provided chimeras whose performance was improved over the best parents in each of the five different conditions tested, and chimeras with combinations of properties that go beyond the combinations seen in the parental enzymes [22].

By using natural sequence diversity, family shuffling mimics classical breeding, producing properties and combinations of properties not found in nature. In combination with a well-designed screen, family shuffling greatly accelerates the optimization of enzymes for reagent, clinical, or industrial uses. The efficient generation of large libraries of functional genes with mutational distances from the parents far greater than can be accessed by the successive accumulation of random point mutations should allow access to new catalytic activities or combinations of properties optimal for various desired applications.

Conclusion

By mimicking the natural mechanisms by which existing diversity can be recombined, DNA shuffling can be used to generate high-quality libraries of novel sequences. Chimeras between naturally occurring enzymes that differ by only a few amino acids often possess activities that are significantly different from their parents. By screening these libraries using innovative high-throughput assay techniques, it is possible to identify enzymes with new catalytic functions.

Because random point mutagenesis does not distinguish between functional and non-functional sequence space, the limits imposed by the screening capacity make the probability of accessing highly improved variants low. Employing a rational approach to identify and focus on particular residues, motifs, or domains, also limits the amount of actual sequence space sampled. It is relatively easy to search these limited possibilities for improved or different variants, but if the variant lies outside the particular sequence space, it will never be found by these methods.

At the other extreme, DNA shuffling utilizes non-rational recombination to access a large and diverse volume of functional sequence space. It is the inclusion of recombination, a powerful evolutionary mechanism, which allows access to improved variants not obtainable by other approaches. Inducing recombination in the test tube appears to confer the same benefits to evolution that sexual recombination offers in nature, as the resulting enzyme chimeras can possess novel combinations of the parental properties as well as improvements in desired properties. The ability to select or screen large libraries for such improved variants is the limiting factor. Screens and selections must be carefully designed and verified to ensure that positive clones can be detected and are valid.

[1] W.P.C. Stemmer, 'Searching sequence space', *Bio/Technology* **1995**, *13*, 549.
[2] C. Forst, 'Molecular evolution: a theory approaches experiments', *J. Biotechnol.* **1998**, *64*, 101.
[3] O. Kuchner, F.H. Arnold, 'Directed evolution of enzyme catalysts', *Trends Biotechnol.* **1997**, *15*, 523.
[4] J. Ogawa, S. Shimizu, 'Microbial enzymes: new industrial applications from traditional screening methods', *Tibtech* **1999**, *17*, 13.
[5] H. Zhao, F. Arnold, 'Combinatorial protein design: strategies for screening protein libraries', *Curr. Opin. Struct. Biol.* **1997**, *7*, 480.
[6] W.P.C. Stemmer, 'Rapid evolution of a protein *in vitro* by DNA shuffling', *Nature (London)* **1994**, *370*, 389.
[7] S. Harayama, 'Artificial evolution by DNA shuffling', *Trends Biotechnol.* **1998**, *16*, 76.
[8] J.C. Moore, H.-M. Jin, O. Kuchner, F.H. Arnold, 'Strategies for the *in vitro* evolution of protein function: enzyme evolution by random recombination of improved sequences', *J. Mol. Biol.* **1997**, *272*, 336.
[9] U. Bornscheuer, 'Directed evolution of enzymes', *Angew. Chem. Int. Ed.* **1998**, *37*, 3105.
[10] L. Giver, F.H. Arnold, 'Combinatorial protein design *in vitro* recombination' *Curr. Opin. Chem. Biol.* **1998**, *2*, 335.
[11] A. Crameri, E.A. Whitehorn, E. Tate, W.P.C. Stemmer, 'Improved green fluorescent protein by molecular evolution using DNA shuffling', *Nat. Biotechnol.* **1996**, *14*, 315.
[12] A. Crameri, G. Dawes, E. Rodriguez Jr., S. Silver, W.P.C. Stemmer, 'Molecular evolution of an arsenate detoxification pathway by DNA shuffling', *Nat. Biotechnol.* **1997**, *15*, 436.
[13] J. Minshull, W.P.C. Stemmer, 'Protein evolution by molecular breeding', *Curr. Opin. Chem. Biol.* **1999**, *3*, 284.
[14] L. Giver, A. Gershenson, P.-O. Freskgard, F.H. Arnold, 'Directed evolution of a thermostable esterase', *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12809.
[15] H. Zhao, L. Giver, Z. Shao, J.A. Affholter, F.H. Arnold, 'Molecular evolution by staggered extension process (StEP) *in vitro* recombination', *Nat. Biotechnol.* **1998**, *16*, 258.
[16] T. Matsuura, K. Miyai, S. Trakulnaleamsai, T. Yomo, Y. Shima, S. Miki, K. Yamamoto, I. Urabe, 'Evolutionary molecular engineering by random elongation mutagenesis', *Nat. Biotechnol.* **1999**, *17*, 58.
[17] Z. Shao, H. Zhao, L. Giver, F.H. Arnold, 'Random-priming *in vitro* recombination: an effective tool for directed evolution', *Nucleic Acids Res.* **1998**, *26*, 681.
[18] J.-H. Zhang, G. Dawes, W.P.C. Stemmer, 'Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening', *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4505.
[19] T. Kumamaru, H. Suenaga, M. Mitsuoka, T. Watanabe, K. Furukawa, 'Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase', *Nat. Biotechnol.* **1998**, *16*, 663.
[20] A. Crameri, S.-A. Raillard, E. Bermudez, W.P.C. Stemmer, 'DNA Shuffling of a family of genes from diverse species accelerates directed evolution', *Nature (London)* **1998**, *391*, 288.
[21] F.C. Christians, L. Scapozza, A. Crameri, G. Folkers, W.P.C. Stemmer, 'Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling', *Nat. Biotechnol.* **1999**, *17*, 259.
[22] J.E. Ness, M. Welch, L. Giver, M. Bueno, J.R. Cherry, T.V. Borchert, W.P.C. Stemmer, J. Minshull, 'DNA Shuffling of subgenomic sequences of subtilisin', *Nat. Biotechnol.* **1999**, *17*, 893.