DOE/CH/10578--T5

ENZYME CATALYSTS FOR A BIOTECHNOLOGY-BASED CHEMICAL INDUSTRY

Final Report for the Period

September 29, 1993 - September 28, 1998

Frances H. Arnold

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, CA 91125

November 16, 1998

PREPARED FOR THE UNITED STATES
DEPARTMENT OF ENERGY
Under Cooperative Agreement
No. DE-FG36-93-CH10578

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Enzymes have enormous potential for reducing energy requirements and environmental problems in the chemicals and pharmaceutical industries. Those who try to use enzymes, however, are stymied by the fact that these catalysts have evolved over billions of years to perform very specific biological functions and to do so within the context of a living organism. Some of the features required for function in a complex chemical network are undesirable when the catalyst is lifted out of context (e.g. product inhibition). Conversely many of the properties we wish an enzyme would have clash with the needs of the organism, or at least were never required: high stability, the ability to function in nonnatural environments or catalyze nonnatural reactions.

The explosion of tools that has come out of molecular biology during the last 20 years has made it possible for us to evolve enzymes for features never required in nature. We can speed up the rate and channel the direction of evolution by controlling mutagenesis and the accompanying 'selection' pressures. Darwinian evolution carried out in the test tube offers a unique opportunity for biotechnology: the ability to tailor enzymes for optimal performance in a wide range of applications. Thus it is possible, for example, to evolve enzymes that carry out reactions on nonnatural substrates or even to carry out reactions for which there is no counterpart in nature. Due to the vast size of the potential sequence space, however, explorations by directed evolution must be guided by sound principles and workable strategies. During the course of this group, this laboratory has continued to make significant progress in the evolution of industrial enzymes as well as in developing general methods for in vitro evolution.

Directed evolution of a p-nitrobenzyl esterase

Through sequential generations of random mutagenesis and screening, we have directed the evolution of an esterase for deprotection of an antibiotic p-nitrobenz yl ester in aqueous-organic solvents. Because rapid screening directly on the desired antibiotic (loracarbef) nucleus p-nitrobenzyl ester was not feasible, the much more convenient p-nitrophenyl ester was employed. Catalytic performance on the screening substrate was shown to reasonably mimic enzyme activity toward the desired ester. One p-nitrobenzyl esterase variant performs as well in 30% dimethylformamide (DMF) as the wild type enzyme in water, reflecting a 16-fold increase in esterase activity. Random pairwise gene recombination of two positive variants led to a further two-fold improvement in activity. Considering also the increased expression level achieved during these experiments, the net result of four sequential generations of random mutagenesis and the one recombination step is a 50-60-fold increase in total activity. Although the contributions of individual effective amino acid substitutions to enhanced activity are small (< two-fold increases), the accumulation of multiple mutations by directed evolution allows significant improvement of the biocatalyst for reactions on substrates and under conditions not already optimized in nature. The positions of the effective amino acid substitutions have been identified in a pNB esterase structural model developed based on its homology to acetylcholinesterase and triacylglycerol lipase. None appear to interact directly with the antibiotic substrate, further underscoring the difficulty of predicting their effects in a 'rational' design effort.

Sets of genes improved by directed evolution can be recombined in vitro to produce further improvements in protein function. Recombination is particularly useful when improved sequences are available; costs of generating such sequences, however, must be weighed against the costs of further evolution by sequential random mutagenesis. Four genes encoding para-nitrobenzyl (pNB) esterase variants exhibiting enhanced activity (Moore & Arnold, 1996) were recombined in two cycles of high-fidelity DNA shuffling and screening. Genes encoding enzymes exhibiting further improvements in activity were analyzed in order

to elucidate evolutionary processes at the DNA level and begin to provide an experimental basis for choosing in vitro evolution strategies and setting key parameters for recombination. DNA sequencing of improved variants from the two rounds of DNA shuffling confirmed important features of the recombination process: rapid fixation and accumulation of beneficial mutations from multiple parent sequences as well as removal of silent and deleterious mutations. The five- to six-fold further enhancement of total activity towards the para-nitrophenyl (pNP) ester of loracarbef was obtained through recombination of mutations from several parent sequences as well as new point mutations. Computer simulations of recombination and screening illustrate the trade-offs between recombining fewer parent sequences (in order to reduce screening requirements) and lowering the potential for further evolution. Search strategies which may substantially reduce screening requirements in certain situations are described.

Directed evolution of subtilisin for function in organic solvents

Random mutagenesis has been used to engineer the protease subtilisin E to function in a highly nonnatural environment: high concentrations of a polar organic solvent. Sequential rounds of mutagenesis and screening have yielded a variant (PC3) that hydrolyzes a peptide substrate 256 times more efficiently than wild-type subtilisin in 60% dimethylformamide. PC3 subtilisin E and other variants containing different combinations of amino acid substitutions are effective catalysts for transesterification and peptide synthesis in dimethylformamide and other organic media. Starting with a variant containing four effective amino acid substitutions (D60N, D97G, Q103R and N218S), six new mutations (G131D, E156G, N181S, S182G S188P and T255A) were generated during three sequential rounds of mutagenesis and screening. The ten substitutions are clustered on one face of the enzyme, near the active site and substrate binding pocket, and all are located in loops which connect core secondary structure elements and exhibit considerable sequence variability in subtilisins from different sources. These variable surface loops are effective handles for "tuning" the activity of subtilisin. Seven of the ten amino acid substitutions in PC3 are found in other natural subtilisins. Great variability is exhibited among naturally-occurring sequences that code for similar three-dimensional structures--it is possible to make use of this sequence flexibility to engineer enzymes to exhibit features not previously developed (or required) for function in vivo.

Seven DNA substitutions coding for three new amino acid substitutions were identified in a mutant isolated after two additional generations of directed evolution carried out on subtilisin E PC3. A B. subtilis-E. coli shuttle vector was developed in order to increase the size of the mutant library that could be established in B. subtilis, and the stringency of the screening process was increased to reflect total as well as specific activity. This directed evolution approach has been extremely effective for improving enzyme activity in a nonnatural environment: the resulting evolved 13M subtilisin exhibits specific catalytic efficiency towards the hydrolysis of a peptide substrate succinyl-Ala-Ala-Pro-Phe-p-nitroanilide in 60% DMF solution that is three times that of the parent 10M and 471 times that of wild type subtilisin E. The total activity of the 13M culture supernatant is enhanced 16-fold over that of the parent 10M.

Directed evolution of thermostable enzymes

We have used directed evolution to convert *Bacillus subtilis* subtilisin E into an enzyme functionally equivalent to its thermophilic homolog thermitase from *Thermoactinomyces vulgaris*. Five generations of random mutagenesis, recombination and screening were sufficient to create subtilisin 5-3H5, whose half-life at 83 °C (3.5 min) and temperature optimum for activity (T_{opt}, 76 °C) are identical to those of thermitase. The T_{opt} of

the evolved enzyme is 17 °C higher and its half-life at 65 °C is >200 times that of wild type subtilisin E. n addition, 5-3H5 is more active towards the hydrolysis of succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide than wild type over the whole range of temperatures, from 10 to 90 °C. Thermitase differs from subtilisin E at 157 amino acids. However, only eight amino acid substitutions were sufficient to convert subtilisin E into an enzyme equally thermostable. The eight substitutions, which include known stabilizing mutations (N218S, N76D) as well as several not previously reported, are distributed over the surface of the enzyme. Only two (N218S, N181D) are found in thermitase. Directed evolution provides a powerful tool to unveil mechanisms of thermal adaptation and is an effective and efficient approach to increasing thermostability without compromising enzyme activity.

We have also used *in vitro* evolution to probe the relationship between stability and activity in a mesophilic esterase. Previous studies of these properties in homologous enzymes evolved for function at different temperatures have suggested that stability at high temperatures is incompatible with high catalytic activity at low temperatures through mutually exclusive demands on enzyme flexibility. Six generations of random mutagenesis, recombination and screening stabilized *Bacillus subtilis p*-nitrobenzyl esterase significantly (>14°C increase in T_m) without compromising its catalytic activity at lower temperatures. Furthermore, analysis of the stabilities and activities of large numbers of random mutants indicate that these properties are not inversely correlated. Although enhancing thermostability does not necessarily come at the cost of activity, the process by which the molecule adapts is important. Mutations that increase thermostability while maintaining low-temperature activity are very rare. Unless both properties are constrained (by natural selection or screening) the evolution of one by the accumulation of single amino acid substitutions typically comes at the cost of the other, regardless of whether the two properties are inversely correlated or not correlated at all.

Recombination methods

Simple and efficient methods for mutagenesis and recombination of polynucleotide sequences have been developed. One simple and efficient method for *in vitro* mutagenesis and recombination, the staggered extension process (StEP), consists of priming the template sequence(s) followed by repeated cycles of denaturation and extremely abbreviated annealing/extension. In each cycle the extending 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 bp and eight other point mutation markers. To demonstrate its utility for directed evolution, we have used StEP to recombine a set of five thermostablized 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.

A second simple and efficient method for *in vitro* mutagenesis and recombination of polynucleotide sequences was developed. 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. These fragments are subjected to denaturation followed by 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. The method was also demonstrated by the recombination of genes encoding different thermostable subtilisins in order to obtain enzymes more stable than either parent.

A convenient 'DNA shuffling' protocol for random recombination of homologous genes *in vitro* with a very low rate of associated point mutagenesis (0.05%) was described. In addition, the mutagenesis rate can be controlled over a wide range by the inclusion of Mn^{2+} or Mg^{2+} during DNase I digestion, by the choice of DNA polymerase used during gene reassembly as well as how the genes are prepared for shuffling (PCR amplification vs. restriction enzyme digestion of plasmid DNA). These protocols should be useful for *in vitro* protein evolution, for DNA-based computing, and for structure-function studies of evolutionarily-related genes.

A method for identifying functional mutations

We also developed a method for distinguishing functional from non-functional or deleterious mutations in homologous genes. High-fidelity *in vitro* gene recombination ('DNA shuffling') coupled with sequence analysis of a small sampling of the shuffled library exhibiting the evolved behavior allows identification of those mutations responsible for the behavior in a background of neutral and deleterious mutations. Functional mutations are expected to occur in 100% of the sequenced screened sample; neutral mutations are found in 50% on average, while deleterious mutations do not appear at all. When used to analyze ten mutations in a laboratory-evolved gene encoding a thermostable subtilisin E, this method rapidly identified the two responsible for the observed protease thermostability; the remaining eight are neutral with respect to thermostability, within the precision of the screening assay. A similar approach, coupled with selection for growth and survival of the host organism, could be used to distinguish adaptive from neutral mutations.

Patents issued and pending

Para Nitrobenzyl Esterases with Enhanced Activity in Aqueous and Nonaqueous Media filed Jan 23, 1996. Ser. No. 08/589,892/PCT No. PCT/US97/01175 (Docket # 107-260) J. C. Moore and F. H. Arnold. U. S. Patent 5,741,691 April 21,1998. EPO Patent application number 97905618.1.

Divisional application Ser. No. 09/020,991, February 9, 1998.

Recombination of Polynucleotide Sequences using Random or Defined Primers, F. H. Arnold, H. Zhao, Z. Shao, J. Affholter, L. Giver, Utility application filed 8/4/97. Ser. No. 08/905,359.

Method for Creating Novel Polynucleotide and Polypeptide Sequences. F. H. Arnold and Z. Shao. Provisional patent application filed 12/8/97.

Thermally Stable para-nitrobenzyl Esterases, L. Giver and FH Arnold, Ser. No. 09/062,890. Filed US patent application, April 20, 1998.

References and publications from this grant

1. "Engineering Proteins for Nonnatural Environments," F. H. Arnold, invited review for *FASEB Journal* **7**, 744-749 (1993).

2. "Tuning the Activity of an Enzyme for Unusual Environments: Sequential Random Mutagenesis of Subtilisin E for Catalysis in Dimethylformamide," K. Chen and F. H. Arnold, *Proc. Natl. Acad. Sci. USA*, **90**, 5618-5622 (1993).

3. "Protein Engineering for Unusual Environments," F. H. Arnold, invited review for *Current Opinion in Biotechnology*, Vol. 4, 450-455 (1993).

4. "Directed Evolution of Subtilisin E in *Bacillus Subtilis* to Enhance Total Activity in Aqueous Dimethylformamide," L. You and F. H. Arnold, *Protein Engineering* **9**, 77-83 (1996).

5. "Directed Evolution of a para-Nitrobenzyl Esterase for Aqueous-Organic Solvents," J. Moore and F. H. Arnold, *Nature Biotechnology* **14**, 458-467 (1996).

6. "Directed Evolution: Creating Biocatalysts for the Future," F. H. Arnold, *Chemical Engineering Science*, **51**, 5091-5102 (1996).

7. "Strategy for the Directed Evolution of a Peptide Ligase," H. Zhao, L. You and F. H. Arnold, Proceedings of the 13th Enzyme Engineering Conference, *Annals of the N. Y. Acad. Sci.* **799**, 1-6 (1996).

8. "Engineering New Functions and Altering Existing Functions," Z. Shao and F. H. Arnold, *Current Opinion in Structural Biology* **6**, 513-518 (1996).

9. "Optimizing Industrial Enzymes by Directed Evolution," F. H. Arnold and J. C. Moore, *Advances in Biochemical Engineering* **58**, 1-14 (1997).

10. "Methods for Optimizing Industrial Enzymes by Directed Evolution," H, Zhao, J. C. Moore, A. A. Volkov and F. H. Arnold, *ASM Handbook of Industrial Microbiology*, in press.

11. "Optimization of DNA shuffling for high fidelity recombination," H. Zhao and F. H. Arnold, *Nucleic Acids Research.*, **25**, 1307-1308 (1997).

12. "Functional and non-functional mutations distinguished by random recombination of homologous genes," H. Zhao and F. H. Arnold, *Proc. Natl. Acad. Sci. USA*, **94**, 7997-8000 (1997).

13. "Combinatorial Protein Design: Strategies for Screening Protein Libraries." H. Zhao and F. H. Arnold, *Current Opinion in Structural Biology* **7**, 480-485 (1997).

14. "Strategies for the *in vitro* Evolution of Protein Function: Enzyme Evolution by Random mbination of Improved Sequences," J. C. Moore. H.-M. Jin, O. Kuchner and F. H. Arnold, *J. Molecular Biology* **272**, 336-347 (1997).

15. "Directed Evolution of Enzyme Catalysts," O. Kuchner and F. H. Arnold, invited review for *Trends in Biotechnology*, **15**, 523-530 (1997).

16. "Random Priming *in vitro* Recombination: an Effective Tool for Directed Evolution," Z. Shao, H. Zhao, L. Giver and F. H. Arnold, *Nucleic Acids Research* **26,** 681-683(1998).

17. "Design by Directed Evolution," F. H. Arnold, *Accounts of Chemical Research* **31**, 125-131 (1998).

18. "Molecular Evolution by Staggered Extension Process (StEP) in vitro Recombination," H. Zhao, L. Giver, Z. Shao, J. A. Affholter, F. H. Arnold, *Nature Biotechnology* **16**, 258-262 (1998).

19. "Combinatorial protein design by *in vitro* DNA recombination (DNA Shuffling)" L. Giver and F. H. Arnold, *Current Opinion in Chemical Biology*, **2**, 335-338 (1998).

20. "Enzyme engineering reaches the boiling point," F. H. Arnold, *Proc. Natl. Acad. Sci. USA*, **95**, 2035-2036 (1998).

21. "Laboratory evolution of a thermostable esterase," L. Giver, A. Gershenson, P.-O. Freskgard, F. H. Arnold, *Proc. Natl. Acad. Sci. USA*, **95**, 12809-12813 (1998).

22. "Directed evolution converts subtilisin E into a functional equivalent of thermitase, H. Zhao and F. H. Arnold, *Protein Engineering*, in press.

23. "When blind is better: protein design by evolution," F. H. Arnold, commentary in *Nature Biotechnology* **16**, 617-618 (1998).

24. "Directed evolution of mesophilic enzymes into their thermophilic counterparts," F. H. Arnold, L. Giver, A. Gershenson, H. Zhao, K. Miyazaki, Proceedings of the NYAS, in press (1998).

25. "Directed enzyme evolution," F. H. Arnold and Patrick Wintrode, chapter forEncyclopedia of Bioprocess Technology, M. C. Flickinger and S. W. Drew, Eds., Wiley, in press.

26. "Directed evolution of industrial enzymes," C. Schmidt-Dannert and F. H. Arnold, *Trends in Biotechnology*, in press.

27. "Molecular Evolution: A biotechnology Revolution for the Chemicals Industry or? Molecular Evolution: Bringing life to chemistry in the 21ST Century," J. Affholter and F. H. Arnold, *Chemistry in Britain*, in press.