REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 18, 2003	Final Report 12 June 2000 – 11 June 2003

4. TITLE AND SUBTITLE
The Temperature Limits of Enzyme Function: Laboratory Evolution of Enzymes More Stable Than Any Natural Enzyme (**Revised title**: Evolution by Structure-Based Protein Recombination)

5. FUNDING NUMBERS
Grant No. DAAD19-00-1-0391

6. AUTHOR(S)
Frances H. Arnold

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
California Institute of Technology
1200 E. California Blvd.
Pasadena, CA 91125-4100

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
41082.2-LS

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The computational algorithm SCHEMA was developed to estimate the disruption caused when amino acid residues that interact in the three-dimensional structure of a protein are inherited from different parents upon recombination. If the folded structure is retained, the new chimeric sequences can display functional properties characteristic of the parents; they can also acquire entirely new functions. We are using SCHEMA to design chimeras of cytochrome P450 heme enzymes. In a preliminary study of 17 chimeras, we observed significant functional diversity: chimeras displayed altered substrate specificity profiles, a wide range in thermostabilities, up to a 40-fold increase in peroxidase activity, and ability to hydroxylate a substrate towards which neither parent shows detectable activity. SCHEMA-guided recombination can generate diverse sequences for exploring the evolution of function within a given protein structural framework.

14. SUBJECT TERMS
directed evolution, thermostable, recombination, protein design, oxygenase

15. NUMBER OF PAGES
Four

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	**UNCLASSIFIED**	**UNCLASSIFIED**	**UL**

NSN 7540-01-280-5500

20030925004

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. 239-18
298-102

Final Progress Report

Evolution by Structure-Based Protein Recombination (revised title)

ARO Grant No. DAAD19-00-1-0391

12 June 2000 – 11 June 2003

Frances H. Arnold, Principal Investigator

California Institute of Technology

1) Statement of problem studied

As part of this project, we studied how the evolution of thermostability influences protein dynamics. Molecular dynamics simulations were employed to study how protein solution structure and dynamics are affected by adaptation to high temperature.

In 2001 we shifted the focus of this project to developing structure-based computational approaches for improving molecular evolution methods of protein design. We have validated the SCHEMA algorithm for designing chimeric sequences and libraries, from which we can identify useful, new proteins.

2) Summary of accomplishments

Protein dynamics in a family of laboratory evolved thermophilic esterases (Wintrode):

As part of this project, we studied how the evolution of thermostability influences protein dynamics, in collaboration with the Goddard group in theoretical chemistry. Molecular dynamics simulations were employed to study how protein solution structure and dynamics are affected by adaptation to high temperature. Simulations were carried out on a para-nitrobenzyl esterase (484 residues) and two thermostable variants that were generated by laboratory evolution. Although these variants display much higher melting temperatures than wild-type (up to 18 °C higher) they are both >97% identical in sequence to the wild-type. In simulations at 300 K the thermostable variants remain closer to their crystal structures than wild-type. However, they also display increased fluctuations about their time-averaged structures. Additionally, both variants show a small but significant increase in radius of gyration relative to wild-type. The vibrational density of states was calculated for each of the esterases. While the density of states profiles is similar overall, both thermostable mutants show increased populations of the very lowest frequency modes (<10 cm-1), with the more stable mutant showing the larger increase. This indicates that the thermally stable variants experience increased concerted motions relative to wild-type. Taken together, these data suggest that adaptation for high temperature stability has resulted in a restriction of large deviations from the native state and a corresponding increase in

smaller scale fluctuations about the native state. These fluctuations contribute to entropy and hence to the stability of the native state. The largest changes in localized dynamics occur in surface loops, while other regions, particularly the active site residues, remain essentially unchanged. Several mutations, most notably L313F and H322Y in variant 8G8, are in the region showing the largest increase in fluctuations, suggesting that these mutations confer more flexibility to the loops. As a validation of our simulations, the fluctuations of Trp102 were examined in detail, and compared with Trp102 phosphorescence lifetimes that were previously measured. Consistent with expectations from the theory of phosphorescence, an inverse correlation between out-of-plane fluctuations on the picosecond time scale and phosphorescence lifetime was observed.

Computational approaches to improve molecular evolution searches:

The computational algorithm SCHEMA was developed to estimate the disruption caused when amino acid residues that interact in the three-dimensional structure of a protein are inherited from different parents upon recombination (Voigt et al., *Nature Structural Biology*, **9**, 553-558 (2002). We then showed that the simple distance metric used by SCHEMA to identify interactions and compute crossover disruption allows one to predict which chimera sequences are most likely to retain their function (Meyer et al., *Protein Science*, **12** (8) (2003). This approach can be used to evaluate crossover sites for recombination and to create highly mosaic, folded chimeras.

With ARO support we have explored recombination in a key enzyme family, the cytochrome P450s (CYP). Recombination generates chimeric proteins whose ability to fold and function depends on minimizing structural perturbations that result when portions of the sequence are inherited from different parents. If the folded structure is retained, the new chimeric sequences can display functional properties characteristic of the parents; they can also acquire entirely new functions. SCHEMA was used to design chimeras of two members of the CYP102 subfamily of the functionally diverse cytochrome P450 heme enzymes. Seventeen double-crossover chimeras were constructed from the heme domains of CYP102A1 and CYP102A2 (63% amino acid identity). Chimeras predicted to have limited structural disruption displayed CO-binding spectra characteristic of folded P450s, and many were catalytically active. Even this small population of chimeras exhibited significant functional diversity: chimeras displayed altered substrate specificity profiles, a wide range in thermostabilities, up to a 40fold increase in peroxidase activity, and ability to hydroxylate a substrate towards which neither parent shows detectable activity. These results suggest that SCHEMA-guided recombination can generate diverse P450s for exploring the evolution of function within the P450 structural framework.

We are now creating large libraries of diverse P450s by structure-guided recombination. We expect that a significant fraction will fold properly and that many will exhibit novel, and potentially useful, function. Our goal is to assess the degree of functional evolution that is possible using this evolutionary search strategy. We believe that structure-guided recombination will allow us to access useful catalysts rapidly and create libraries that encompass many of the known natural activities, as well as useful new activities.

3) Publications submitted or published under ARO sponsorship during this reporting period

P. L.Wintrode, K. Miyazaki and F. H. Arnold, "Patterns of Adaptation in a Laboratory Evolved Thermophilic Enzyme," *Biochim. Biophys. Acta* **1549**, 1-8 (2001).

F. H. Arnold, P. C. Wintrode, K. Miyazaki and A. Gershenson, "How Enzymes Adapt: Lessons from Laboratory Evolution," *Trends Biochemical Sciences* **26**, 100-106 (2001).

F. H. Arnold, "Combinatorial and Computational Challenges for Biocatalyst Design," *Nature* **409**, 253-257 (2001).

P. L. Wintrode, D. Zhang, N. Vaidehi, F. H. Arnold and W. A. Goddard III, "Protein Dynamics in a Family of Laboratory-Evolved Thermophilic Enzymes," *J. Mol. Biol.* **327**, 745-757 (2003).

C. Otey, J. J. Silberg, C. A. Voigt, J. B. Endelman, G. Bandara and F. H. Arnold "Structure-Based SCHEMA Recombination of Cytochrome P450," *Chemistry & Biology*, submitted.

4) Scientific personnel supported by project and honors/awards/degrees received

Holger Berk (held DFG Fellowship)
Cecilia Orencia (January-April, 2001)
Patrick Wintrode
Radu Georgescu (received M.S. in Biochemistry, 2001)
Geethani Bandara

5) Report of inventions

Creation of Diverse Cytochrome P450 Enzymes by Structure-Guided Recombination, C. Otey and F. H. Arnold, inventors. Invention disclosure CIT-3910, Filed: 6/17/2003.