Acc. Chem. Res. **1998,** *31,* 125−131

Design by Directed Evolution

FRANCES H. ARNOLD*

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

Received October 9, 1997

An Introduction to "Irrational" Design

The stunning array of features and functions exhibited by proteins in nature should convince most scientists of the power of evolutionary design processes. Natural selection acting on populations over long periods of time has generated a vast number of proteins ideally suited to their biological functions. When we try to recruit these remarkable molecular machines for new tasks—from serving as industrial catalysts to being used as additives for laundry detergents—we find that they are often not so well suited. (The chemist is not terribly impressed with a synthesis requiring a reactor the size of a football field simply because the enzyme functions in water and the substrate does not dissolve, nor is she pleased with a catalyst shut down by the products of its own reaction.) Evolution is usually the culprit: proteins are optimized, and often highly specialized, for specific biological tasks.

Most protein reengineering efforts have been by so-called rational design. The filtering effect of scientific publication (successes get published, failures mostly do not) might lead one to believe that we can, with reasonable probability of success, identify and modify the amino acids responsible for key properties such as an enzyme's substrate preference, stability, or activity in a nonnatural environment. In reality, we are far from being able to do this reliably. This is true even for the relatively small number of enzymes for which considerable structural and mechanistic data are available. Admitting ineptitude in rational design, however, frees us to consider other approaches which are hardly irrational. An alternative and highly effective design strategy can be found by looking to the processes by which all these proteins came about in the first place.

My research has been devoted to recreating in the laboratory the key processes of evolution and doing it in such a way that we can design scientifically interesting and technologically useful molecules. The challenge is to collapse the time scale for evolution from millions of years to months or even weeks.

An Evolutonary Strategy

Evolutionary mechanisms at work in nature ensure adaptability to ever-changing environments. Evolution does not work toward any particular direction, nor is there a goal; the underlying processes occur spontaneously during reproduction and survival. The laboratory evolution experiment, in contrast, often has a defined goal, and the key processes—mutation, recombination, and screening or selection—are carefully controlled by the experimenter. Obtaining an enzyme with a particular new feature is not a trivial task. It requires a good evolutionary strategy, carefully thought out to maximize the chances of success. Consider the following facts: (i) Protein sequence space is huge: with 20 letters in the protein alphabet, there are 20^{300}—essentially an infinite number of ways to string together 300 amino acids. It is also multidimensional (there are 300×19 sequences just one amino acid mutational step away from any given sequence 300 amino acids long) and highly interconnected (*any* point in this space is at most 300 mutational steps away). (ii) Protein space is mostly empty of function, especially the particular function you want. (iii) Beneficial mutations are rare. Combinations of beneficial mutations are very rare.

Features (i) and (ii) of protein space argue against a purely random sampling of even rather short sequences in the hope that something useful will surface. While researchers doing in vitro selection experiments with nucleic acids have successfully used sparse samplings of large portions of sequence space to pull out functional molecules,[1] protein evolution, at least for practical applications, is best done as nature appears to do it: by random walk from a good starting point. A good starting point is a protein "close" to what you want. We might also use several starting points simultaneously, as I will discuss further below.

Evolution is often referred to as a hill-climbing exercise in the "fitness landscape" of sequence space.[2] The fitnesses of the protein sequences in sequence space make up this landscape, only very small portions of which have been explored during natural evolution and whose most basic features are still quite unknown. Feature (iii) is another way of saying that most paths lead downhill. Furthermore, the probability that one will find oneself at a higher fitness point decreases rapidly the farther away one jumps. This points to a local landscape near existing, functional proteins that is more smooth than rugged and that may even be monotonically increasing (Fujiyama-like), at least in some of its many dimensions. The important consequence, however, of feature (iii) is that the uphill climb in this landscape is more likely to be

Frances H. Arnold's career has traversed (by seemingly random walks) engineering and science. Her B.S. is in mechanical engineering (Princeton) and her Ph.D. in chemical engineering (University of California, Berkeley). Following postdoctoral research in chemistry at Berkeley and the California Institute of Technology, she joined Caltech's Division of Chemistry and Chemical Engineering in 1987 where she is Professor of Chemical Engineering. Her research interests include molecular recognition by metal ion complexes and in vitro evolution (http://www.che.caltech.edu/groups/fha). She particularly enjoys raising her own three products of evolution by recombination.

* E-mail address: frances@cheme.caltech.edu. Web site: http://www.che.caltech.edu/groups/fha.

S0001-4842(96)00017-9 CCC: $15.00
Published on Web 02/28/1998

successful if it can take place in small steps. Although this type of random walk with "baby steps" might bother those who envision traversing real three-dimensional landscapes with their valleys and cliffs, traversing a multidimensional protein landscape is qualitatively different. The chances of getting stuck on a local optimum are greatly diminished by having so many opportunities to escape.[3] While we may never reach the "global optimum", the improvements achieved by taking this route may nonetheless yield a highly successful result.

In the end, it is the search mechanism that poses a technical, but very important determinant of the size of the step for laboratory evolution. Even a good selection may be limited to searching 10^8 protein variants; screens often can cover no more than 10^4.[4] The first law of random mutagenesis ("you get what you screen for") usually severely erodes the value of selections compared to a screen that directly measures, or at least couples strongly with, the specific function(s) of interest. In fact, most of the experimental effort of directed evolution is devoted to devising, validating, and implementing a suitable screen.

These arguments lead to an evolutionary strategy in which the steps are small (preferably single amino acid substitutions in each generation) and in which one relies on the accumulation of multiple such mutations to acquire the desired function.[5]

Some Directed Evolution Experiments: "Asexual" Evolution by Sequential Rounds of Random Mutagenesis

An early directed evolution experiment in our laboratory was the sequential random mutagenesis of the protease subtilisin E to function in high concentrations of dimethylformamide (DMF).[6,7] An enzyme that tolerates organic solvents is useful for synthetic reactions on substrates poorly soluble in water. Organic solvents can also alter reaction selectivities: in the case of proteases, addition of solvent promotes peptide synthesis over hydrolysis.[8] Most useful (polar) cosolvents, however, dramatically reduce enzyme activity. Kevin Chen and Li You in my group used multiple generations of random mutagenesis and screening to evolve subtilisin E to work almost as well in 60% DMF as the original enzyme did in aqueous solution, a nearly 500-fold increase in total enzyme activity. Ten amino acid substitutions distributed on the enzyme surface around the active site and substrate binding pocket conferred a 176-fold increase in specific activity, while two additional mutations improved expression as well as specific activity.

In 1994 Stephen Queener of Eli Lilly & Co. challenged us to evolve an efficient *p*-nitrobenzyl (pNB) esterase from an enzyme they had isolated from *Bacillus subtilis* and whose natural function was unknown. To deprotect the pNB ester intermediate in the synthesis of loracarbef (Figure 1a), the enzyme would have to function in the presence of a polar organic solvent, since the protected antibiotic substrate is poorly soluble in water. With only four generations of mutagenesis and screening in 15−20% DMF (first on a surrogate, *p*-nitrophenyl substrate and then on the desired pNB substrate), graduate student Jeffrey Moore increased the enzyme's catalytic efficiency more than 20-fold.[9] The evolutionary progression of efficiency over the four generations (Figure 1a) shows that the evolved enzymes were all more efficient in 1% DMF as well as in 15% DMF. Directed evolution of a loracarbef pNB esterase, however, came at the cost of the enzyme's activity toward a smaller substrate, *p*-nitrophenyl acetate, in low concentrations of DMF (Figure 1b). It is often seen that while one particular feature evolves, other properties will drift. If, however, two properties are coupled to one another, then evolution of one can have a dramatic effect on the other.

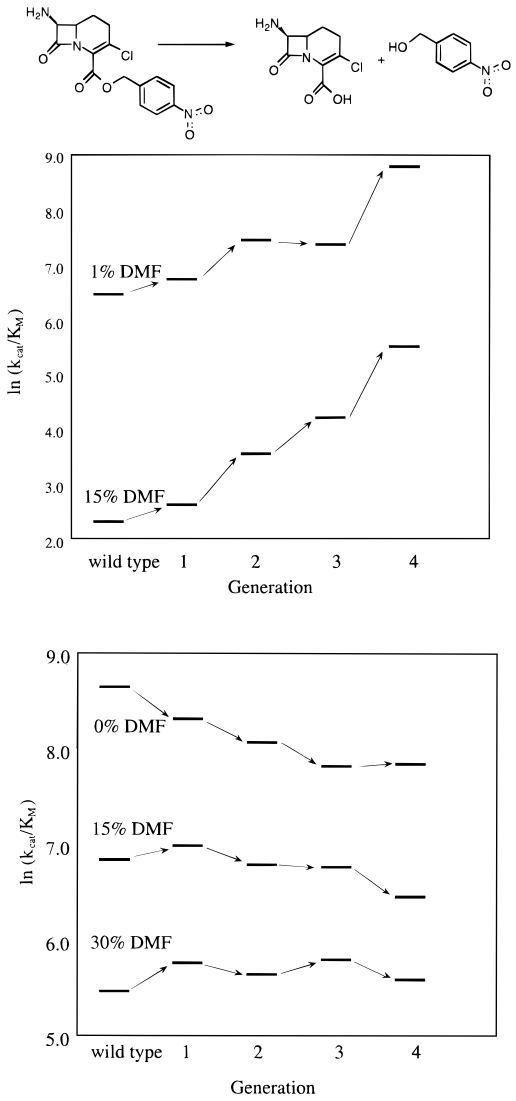

FIGURE 1. Evolutionary progressions show how mutations that enhance the catalytic efficiency of *p*-nitrobenzyl esterase toward the pNB ester of loracarbef (a, top) affect activity toward *p*-nitrophenyl acetate (pNPA) (b, bottom) in various concentrations of DMF. Directed evolution to increase activity toward the loracarbef pNB ester in aqueous DMF causes activity toward pNPA to drift or even decrease.

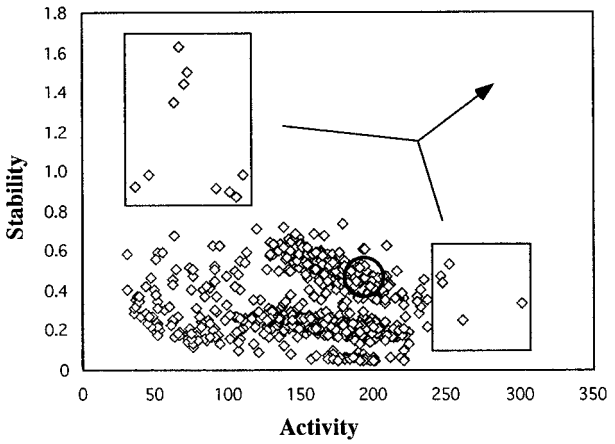

FIGURE 2. Activities and thermostabilities (as measured by the ratio of residual to initial activity, in arbitrary units) of a library of randomly mutagenized pNB esterases. Parent pNB esterases all lie within the circle, showing that the screen is sensitive to changes brought about by primarily single amino acid substitutions. Variants exhibiting improvements in single properties are observed much more frequently than variants with improvements in both properties simultaneously. Thermostabilizing mutations *tend* to deactivate, while activating mutations *tend* to destabilize. Individual populations (in the boxes) can be recombined to obtain enzymes that are both more active and more stable.

We are using directed evolution, in fact, to determine the extent to which two or more properties (e.g., activity and stability) are coupled. With a slight modification of the esterase activity screen to include measuring residual activity after incubation at high temperature, Lori Giver, Anne Gershenson, and Per Ola Freskgard evolved pNB esterases that are both more stable and more active than the natural enzyme. After six generations the melting temperature had increased from 52 to more than 66 °C. The activity at 60 °C increased by nearly a factor of 30.[10] While natural enzymes isolated from thermophilic organisms are active and stable at higher temperatures, their activities at lower temperature are often compromised. This observation has been used to support the idea that an improvement in one property (thermostability) necessarily comes at the cost of the other (activity at low temperature). It is widely believed that enhanced molecular rigidity is a prerequisite for thermostability, while maintaining flexibility is required for catalytic activity. An alternate explanation for the observation that natural proteins from thermophiles are less active than their mesophilic counterparts at lower temperatures, however, is that natural selection exerts pressure on one, but not both, properties. Because enzymes from a mesophile need not function at high temperature, their activities at high temperature are free to drift. And, in fact, a random sampling of the activities and stabilities of a random mutant library shows that of the many solutions to the stabilization problem, most come at the cost of activity (at low temperature) (Figure 2). Solutions to the problem of evolving stability to high temperatures and activity at low temperature simultaneously are considerably more rare, and therefore are not usually observed in nature.

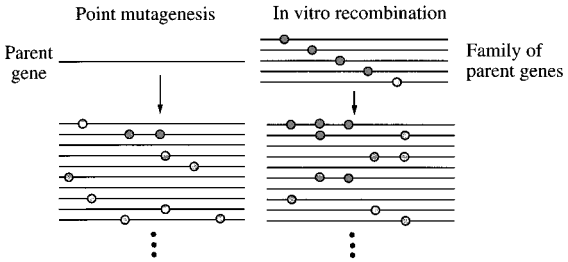

FIGURE 3. Starting from a single parent sequence, random mutagenesis methods (for example, using error-prone PCR) generate a library of genes containing point mutations. Recombination generates gene libraries with different combinations of the mutations from a pool of parent sequences. The parent sequences can be positive mutants from a round of screening, genes evolved separately for different properties, or even closely related natural sequences.

"Sexual" Evolution by Gene Recombination

When we evolve an enzyme by sequential generations of random mutagenesis and screening, we use the best variant identified in each generation to parent the next generation. It is painful to set aside other potentially useful variants; screening is time-consuming, and their beneficial mutations must be rediscovered in future generations to become incorporated. This approach can also accumulate deleterious mutations, possibly limiting the fitness that can be reached. (This cost to asexual evolution is known as Müller's ratchet.[11]) Sex provides some significant benefits in natural evolution (to make up for its obvious costs?). Some of the benefits can be captured in the test tube. By recombining parental genes to produce libraries of different mutation combinations, we can quickly accumulate the beneficial mutations, while removing any deleterious ones. Using the "DNA shuffling" method for in vitro recombination invented by Pim Stemmer of Maxygen[12,13] and other methods recently developed in this laboratory,[14,15] we have been adding a little sex to our evolutionary design strategy.

The goal is to create gene libraries containing all possible combinations of mutations present in the parent genes (Figure 3). This library can be screened to find the combination of mutations giving rise to the best enzymes. Stemmer demonstrated the power and utility of in vitro recombination in directed evolution with DNA shuffling.[12,13] His group at Maxygen has gone on to provide some fascinating examples of enzyme[16,17] and metabolic pathway[18] design by directed evolution. We found that the rate of point mutagenesis associated with DNA shuffling was higher than optimal for directed evolution of most enzymes (we generally like to have ~3 new DNA mutations per gene, to give an average of one functional amino acid substitution). Huimin Zhao made modifications to the Stemmer shuffling protocol to allow careful control over this mutagenesis rate and obtain recombined libraries with fewer point mutations.[19] Jeffrey Moore and Hua Min Jin applied high-fidelity DNA shuffling to four higher-activity pNB esterases isolated from the fourth generation of random mutagenesis. Six improved pNB esterase genes isolated from this recombined library

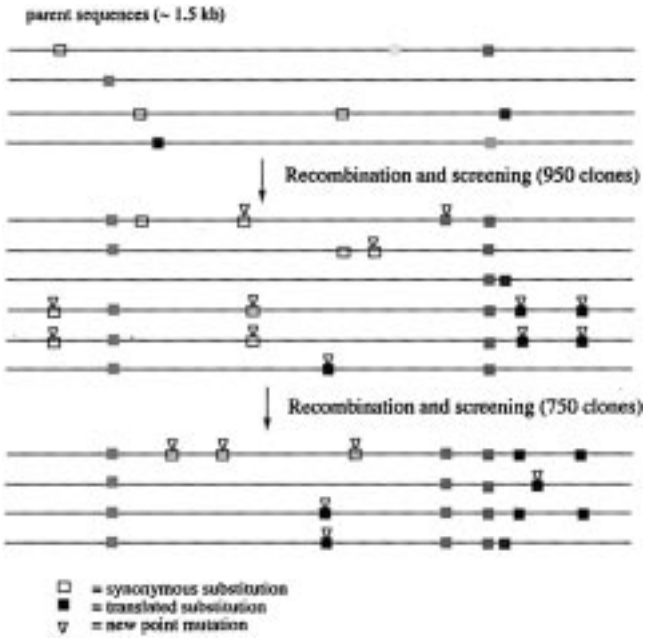

FIGURE 4. Sequencing reveals the accumulation of beneficial mutations (red, magenta) and removal of deleterious mutations (yellow) during pNB esterase evolution by two cycles of DNA shuffling and screening. A new beneficial point mutation identified in the first cycle of recombination (green) becomes fixed. Synonymous DNA mutations from the parental sequences are flushed out by recombination.

were shuffled again to create a sixth generation, which was screened for further increases in activity.

Sequencing the genes for the evolved pNB esterases from the two cycles revealed how recombination contributed to the evolutionary process. Figure 4 illustrates the accumulation and removal of mutations during these two cycles of recombination and screening. Beneficial mutations from two of the parent sequences (shown in red and magenta) were essentially already fixed among the improved variants identified by screening 950 clones from the fifth generation. The further improvement in activity observed in the sixth generation actually arose as a result of a new, beneficial point mutation (green). The remaining mutations that were beneficial in the parent sequences (blue, black, yellow) do not appear in the best recombined sequences, which indicates that the functional effects of these mutations cannot be accumulated.

New and convenient methods to accomplish in vitro recombination have been developed in this laboratory by postdoctoral researchers Zhixin Shao and Lori Giver, graduate student Huimin Zhao, and Joseph Affholter of Dow Chemical. In one approach, we use DNA polymerase to extend random-sequence primers into a pool of gene fragments that are recombined and reassembled into full-length genes.[14] If we use defined-sequence primers to prepare the gene fragments, we can bias recombination to particular sites (Zhao et al., manuscript in preparation). In vitro recombined gene libraries are also efficiently prepared from a set of template sequences by a staggered extension process (StEP), in which multiple cycles of denaturation and very short annealing/extension are used to prepare the gene library. The growing genes switch templates and thereby acquire information from multiple parent sequences as they extend.[15]

A further important use of these in vitro recombination methods is in combining several desired features, starting from libraries of enzymes evolved separately for individual properties. I have illustrated this concept in Figure 2, which shows the results of screening a randomly mutagenized enzyme (pNB esterase) library for activity and stability simultaneously. It is not surprising that few variants exhibit improvements in both properties; few single amino acid substitutions are likely to enhance both activity and stability. Several variants, however, are either more stable or more active (sometimes at the cost of the second property). These populations can be recombined to create enzymes that are improved with respect to both properties.

We have looked at the statistics of recombination to help us choose key parameters, such as the number of parent sequences to recombine or the screening required to find improved sequences.[20] For N sequences with M total mutations, the probability of generating the sequence containing all of the mutations (the rarest sequence) by random recombination is $1/N^M$. This becomes small very quickly when multiple sequences (or many mutations) are recombined. In practice, of course, oversampling is required to ensure that a particular variant has been examined. The screening required to identify the best recombinants is further increased by the point mutations that accompany any in vitro recombination method. Because recombination of relatively few mutations can lead to very large screening requirements, it is often best to use alternative search strategies involving recursive cycles of recombination and screening to find the best mutation combinations.[12]

Using DNA shuffling (or other in vitro or in vivo recombination methods), laboratory evolution can begin from multiple, closely related starting points instead of a single sequence.[21] The molecular diversity in existing functional sequences (i.e., homologous enzymes) has been "prescreened"; recombination of such sequences will create considerable diversity, and with a higher frequency of functional proteins than will equivalently high levels of random mutagenesis. Homologous enzymes evolved from a common ancestor, and it is believed that the vast majority of sequence differences have little effect on fitness (are neutral). Neutral mutations can recombine to yield novel, functional sequences. The sequence space between existing functional proteins may therefore be a particularly promising place to search for improvements in function, and perhaps even novel function.

Extracting Useful Information from the Results of Evolution

One particularly satisfying feature of design by directed evolution is that, if all goes well, improvements come quite rapidly. Improving protein function precedes, and is far easier than, understanding the molecular mechanisms

underlying those improvements. Many research sponsors are of course thrilled with this permutation of the usual order of events. However, we would like to use these laboratory evolution experiments to not only create new, functional proteins but also probe fundamental features of protein structure and function.

Directed evolution is in fact very well suited to fundamental studies. If one wishes to probe the molecular basis of a function, is it not attractive to be able to study a thousand related proteins instead of one, two, or even a few dozen natural sequences? Is it not useful to probe functions (and combinations of functions) not required or explored in nature? Furthermore, natural sequences often owe more to historical accident and random genetic drift than to functional adaptation. Proteins extracted from species adapted to diverse environments, for example, often differ at dozens of amino acids, some small but unknown fraction of which are responsible for specific functional differences. Large numbers of neutral changes add to the confusion of the scientist attempting to discern the molecular rules. In laboratory evolution experiments the lineages are clear, and the mutations are primarily adaptive. Large libraries of mutants studied at the phenotypic level can help us to understand coupling between properties, couplings which are masked in nature by neutral drift or by elusive and otherwise irrelevant evolutionary mechanisms.

We can probe more deeply into specific mechanisms by sequencing evolved genes to determine their mutations, with the goal of assigning functional changes to specific DNA or amino acid substitutions. If the optimal evolutionary strategy is followed (i.e., the libraries are generated to maximize the number of clones with single amino acid substitutions), there will usually be only one amino acid substitution per generation, and functional changes from generation to generation can therefore be assigned to individual mutations. When more than one mutation occurs, however, the interpretation may not be straightforward. In addition, synonymous mutations can affect properties such as total activity, through changes in expression levels. We share the predicament of those who study natural sequences.

We therefore developed a convenient method for distinguishing those mutations responsible for the changes in function from those that either have no effect (neutral) or are deleterious.[22] This method is based on backcrossing with the wild type, as first proposed by Stemmer.[12] First we recombine the evolved and wild-type genes to create the library containing all the possible mutation combinations. If we recombine equimolar amounts of the two genes, then each mutation has a 50% chance of showing up in any given gene in the recombined population. Because the probability that any one mutation will appear on a given gene is 1/2, the fraction of genes containing all n functional mutations is $(1/2)^n$. Thus, simply by screening that population to determine what fraction exhibits the evolved phenotype (e.g., thermostability), we can estimate the number of functional mutations from the percentage of clones exhibiting that phenotype. Finally, we can identify the functional mutations by sequencing a few of the genes that code for proteins with the evolved phenotype. In those genes, the neutral substitutions will still appear more or less with 50% frequency, but the beneficial mutations will be present in every gene. Deleterious mutations should not appear at all.

Huimin Zhao applied this method to an evolved thermostable subtilisin E gene to identify which of its 10 DNA mutations conferred the enzyme's higher stability. Upon screening the recombined library, he found that ~23% were thermostable, which immediately told us there were two mutations involved. Sequencing revealed those mutations to give rise to two amino acid substitutions, which Huimin confirmed by site-directed mutagenesis to confer the observed thermostability. This simple experiment was the equivalent of making all possible combinations of 2 of the 10 mutations (of which there are 450) by site-directed mutagenesis and testing them for thermostability.

Positions of Amino Acid Substitutions in Evolved Enzymes

The 10 amino acid substitutions that increase the specific activity of subtilisin E in 60% DMF all occurred at sites on the surface of the enzyme and surrounding the substrate binding pocket (Figure 5a).[6] In addition, they all appeared in variable loops rather than in conserved elements of secondary structure. Turning the enzyme 90° (Figure 5b) clearly shows that several of the beneficial mutations lie a considerable distance from the active site and bound substrate. Mutations appearing at positions 20 Å from the catalytic residues must exert their effects on catalysis through several layers of protein, and their individual small contributions to activity would still be very difficult to predict (and remain very difficult to explain). A few years later we found that the evolutionary "solutions" to increasing the activity of the pNB esterase in aqueous DMF (Figure 5c,d) were similar in many respects. Although these amino acid substitutions were distributed over a larger region of the enzyme, once again as much as 20 Å away from the catalytic residues,[9,20] they were all on the surface where they can presumably be accommodated without much disruption. The question of how they exert their influence on catalysis of course remains. It is notable that *none* of the pNB esterase mutations are in positions that contact the substrate, at least in our model of the wild-type enzyme. In both enzyme examples, the substitutions show no obvious pattern: large residues are replaced by small, small by large, charged by uncharged, and uncharged by charged.

We hope that comparing the X-ray crystal structures of the wild-type and mutant enzymes will help us to discern some mechanisms by which the evolved enzymes are activated or stabilized, but these structures are not yet available. It will not surprise me, however, when even the structures prove insufficient to elucidate the adaptative mechanisms; the effects of individual mutations are very small and may not influence the structure, at least within

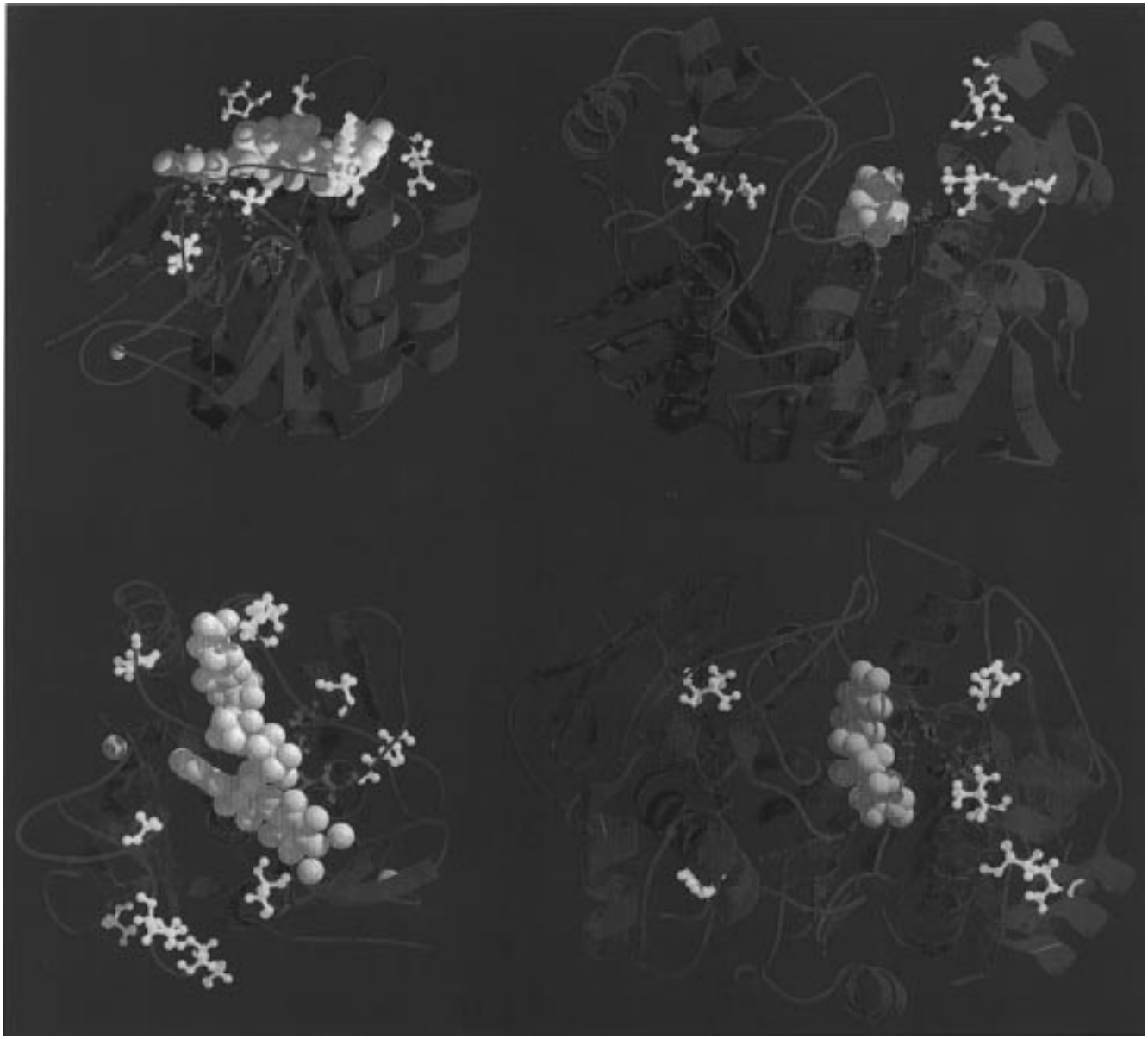

FIGURE 5. Amino acid substitutions (yellow) found to increase hydrolytic activity of subtilisin E (a (top left), b (bottom left)) and pNB esterase (c (top right), d (bottom right)) in aqueous DMF. Catalytic residues are shown in red, substrates in gray. Beneficial mutations in subtilisin E are located on the surface of the protein, in variable loops surrounding the active site and substrate binding pocket. Beneficial mutations in pNB esterase are also primarily on the surface. Mutations in both enzymes are as far as 20 Å from the catalytic sites.

the resolution of the method. Furthermore, other important effects of mutations on enzyme flexibility or dynamics will not be apparent in a static structure. The inherent complexity of proteins, including the ability of amino acid substitutions to exert effects over long distances, underscores the utility of the evolutionary design approach as well as the importance of evolving the whole gene, rather than a limited set of residues believed to influence the property of interest.

Conclusions

Evolution is a very powerful design strategy. Given that we now have the tools to implement evolution in the laboratory, the justifications for continuing to try to tame proteins "rationally" sound rather hollow. In fact, directed evolution is being picked up very quickly by protein engineering groups eager to obtain new functional proteins. The strategies we have developed are robust and can be applied to a wide variety of design problems with a significant probability of success. The limitations, however, of the approach I have outlined are significant. Perhaps the most important is that it will be difficult, if not impossible, to evolve surprising new functions (i.e., those that cannot be created by single or double amino acid substitutions). To do this, we will need new methods for creating and searching libraries many orders of magnitude larger than the ones we work with now. Accessing the functionally richest regions of sequence space, either through judicious (and limited) use of rational design or by other methods proposed[21] or yet to be developed, will certainly help to solve seemingly intractable design problems.

I thank the many talented undergraduate, graduate, and postdoctoral students who have contributed great ideas and hard work. This research would not have been possible without the encouragement and support of many colleagues, including Harold Bright, Gene Peterson, Bob Campbell, Pim Stemmer, and Stephen Queener. This research is supported by the Office of Naval Research, the Department of Energy, and the Army Research Office.

References

(1) Ekland, E. H.; Szostak, J. W.; Bartel, D. P. Structurally Complex and Highly-Active RNA Ligases Derived from Random RNA Sequences. *Science* **1995**, *269*, 364−370.
(2) Kauffman, S. A. *The Origins of Order*; Oxford University Press, Inc.: Oxford, 1993.
(3) Eigen, M. The Physics of Molecular Evolution. *Chem. Scr.* **1986**, *26B*, 13−26.
(4) The differences between screening and selection and laboratory implementation of the two approaches are discussed in two recent reviews: screening (Zhao, H.; Arnold, F. H. Combinatorial Protein Design: Strategies for Screening Protein Libraries. *Curr. Opin. Struct. Biol.* **1997**, *7*, 480−485) and selection (Hilvert, D.; Kast, P. 3D Structural Information as a Guide to Protein Engineering using Genetic Selection. *Curr. Opin. Struct. Biol.* **1997**, *7*, 470−479).
(5) Arnold, F. H. Directed Evolution: Creating Biocatalysts for the Future. *Chem. Eng. Sci.* **1996**, *51*, 5091−5102.
(6) Chen, K.; Arnold, F. H. Tuning the Activity of an Enzyme for Unusual Environments: Sequential Random Mutagenesis of Subtilisin E for Catalysis in Dimethylformamide. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5618−5622.
(7) You, L.; Arnold, F. H. Directed Evolution of Subtilisin E in *Bacillus Subtilis* to Enhance Total Activity in Aqueous Dimethylformamide. *Protein Eng.* **1996**, *9*, 77−83.
(8) Wong, C.-H.; Wang, K.-T. New Developments in Enzymatic Peptide-Synthesis. *Experientia* **1991**, *47*, 1123−1129.
(9) Moore, J. C.; Arnold, F. H. Directed Evolution of a para-Nitrobenzyl Esterase for Aqueous−Organic Solvents. *Nat. Biotechnol.* **1996**, *14*, 458−467.
(10) Giver, L; Gershenson, A.; Freskgard, P. O.; Arnold, F. H. Directed Evolution of a Thermostable Enzyme, submitted for publication.
(11) Müller, H. J. The Relation of Recombination to Mutational Advance. *Mutat. Res.* **1964**, *1*, 2−9.
(12) Stemmer, W. P. C. Rapid Evolution of a Protein *in vitro* by DNA Shuffling. *Nature* **1994**, *370*, 389−391.
(13) Stemmer, W. P. C. DNA Shuffling by Random Fragmentation and Reassembly: *In vitro* Recombination for Molecular Evolution. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10747−10751.
(14) Shao, Z.; Zhao, H.; Giver, L.; Arnold, F. H. Random-Priming *in vitro* Recombination: an Effective Tool for Directed Evolution. *Nucleic Acids Res.*, in press.
(15) Zhao, H.; Giver, L.; Shao, Z.; Affholter, J. A.; Arnold, F. H. Molecular Evolution by Staggered Extension Process (StEP) in vitro Recombination. *Nat. Biotechnol.* submitted for publication.
(16) Crameri, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. C. Improved Green Fluorescent Protein by Molecular Evolution using DNA Shuffling. *Nat. Biotechnol.* **1996**, *14*, 315−319.
(17) Zhang, J.-H.; Dawes, G.; Stemmer, W. P. C. Directed Evolution of a Fucosidase from a Galactosidase by DNA Shuffling and Screening. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 4504−4509.
(18) Crameri, A.; Dawes, G.; Rodrigues Jr., E.; Silver, S.; Stemmer, W. P. C. Molecular Evolution of an Arsenate Detoxification Pathway DNA Shuffling. *Nat. Biotechnol.* **1997**, *15*, 436−438.
(19) Zhao, H.; Arnold, F. H. Optimization of DNA Shuffling for High Fidelity Recombination. *Nucleic Acids Res.* **1997**, *25*, 1307−1308.
(20) Moore, J. C.; Jin, H. M.; Kuchner, O.; Arnold, F. H. Strategies for the *in vitro* Evolution of Protein Function: Enzyme Evolution by Random Recombination of Improved Sequences. *J. Mol. Biol.* **1997**, *272*, 336−347.
(21) Crameri, A.; Raillard, S.-A.; Stemmer, W. P. C. DNA Shuffling of a Family of Genes from Diverse Species Accelerates Directed Evolution. *Nature* **1998**, *391*, 288−290.
(22) Zhao, H.; Arnold, F. H. Functional and Non-Functional Mutations Distinguished by Random Recombination of Homologous Genes. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 7997−8000.

AR960017F