# Protein Design by Directed Evolution

Christian Jäckel, Peter Kast, and Donald Hilvert

Laboratory of Organic Chemistry, ETH Zurich, CH-8093 Zurich, Switzerland;
email: jaeckel@org.chem.ethz.ch, kast@org.chem.ethz.ch, hilvert@org.chem.ethz.ch

Annu. Rev. Biophys. 2008. 37:153–73

The *Annual Review of Biophysics* is online at
biophys.annualreviews.org

This article's doi:
10.1146/annurev.biophys.37.032807.125832



1936-122X/08/0609-0153$20.00

## Key Words

molecular diversity, random mutagenesis, screening, selection, enzymes, computational design

## Abstract

While nature evolved polypeptides over billions of years, protein design by evolutionary mimicry is progressing at a far more rapid pace. The mutation, selection, and amplification steps of the evolutionary cycle may be imitated in the laboratory using existing proteins, or molecules created de novo from random sequence space, as starting templates. However, the astronomically large number of possible polypeptide sequences remains an obstacle to identifying and isolating functionally interesting variants. Intelligently designed libraries and improved search techniques are consequently important for future advances. In this regard, combining experimental and computational methods holds particular promise for the creation of tailored protein receptors and catalysts for tasks unimagined by nature.

## Contents





## INTRODUCTION

The living world has been shaped by Darwinian evolution. Recursive cycles of mutation and selective amplification of the fittest enable populations of organisms to adapt to changes in their environments over time. Beneficial traits that emerge and provide a survival advantage are passed on, whereas deleterious mutations and their hosts gradually disappear.

Evolutionary principles have also been successfully exploited by humans over millennia to breed plants and animals with desirable characteristics. More recently, biologists and chemists have begun to harness evolution to modify the properties of individual molecules, rather than whole organisms, in a directed fashion (4, 5, 72). In evolution experiments in the laboratory, the gene encoding a macromolecule of interest is (partially) randomized. Appropriate screening or selection methods may then be used to identify members of the resulting ensembles (libraries) of mutants that have particular properties, such as the ability to bind a small molecule or to catalyze a chemical transformation. Through iterative cycles of mutagenesis and amplification of selected library members, beneficial mutations accumulate as in genuine Darwinian evolution (**Figure 1**). In this way, populations of macromolecules may be deliberately evolved toward new goals.

Biopolymers are particularly suited for evolutionary approaches because of their inherently combinatorial nature. They are constructed by individually linking many standard building blocks. Evolutionary strategies provide a powerful way to probe large numbers of possible permutations, affording statistically meaningful insights into the complex, often subtle interactions that influence folding and function of such molecules (58). Such methods are also increasingly applied to design, offering means of tailoring existing biopolymers and of creating oligonucleotides and polypeptides, with new activity, from scratch. Although impressive progress has been made toward these goals (4, 5, 72, 83), evolution of biomacromolecules in the laboratory remains a challenging undertaking because functional sequences are rare compared with the almost unimaginably large number of theoretical sequences to be sampled.

In this review, we take stock of recent advances exploiting directed molecular evolution for the purposes of design. After briefly surveying the experimental tools most widely utilized, we highlight several studies illustrating the range of design strategies possible and the results that can be achieved. The design of functional enzymes through directed evolution is a particular focus. Natural enzymes are powerful catalysts but not necessarily suited to applications outside their normal biological contexts. Thus, rapid, systematic production of optimized designer enzymes could alter both industry and medicine in the future. To that end, merged combinatorial and rational approaches hold particular promise.

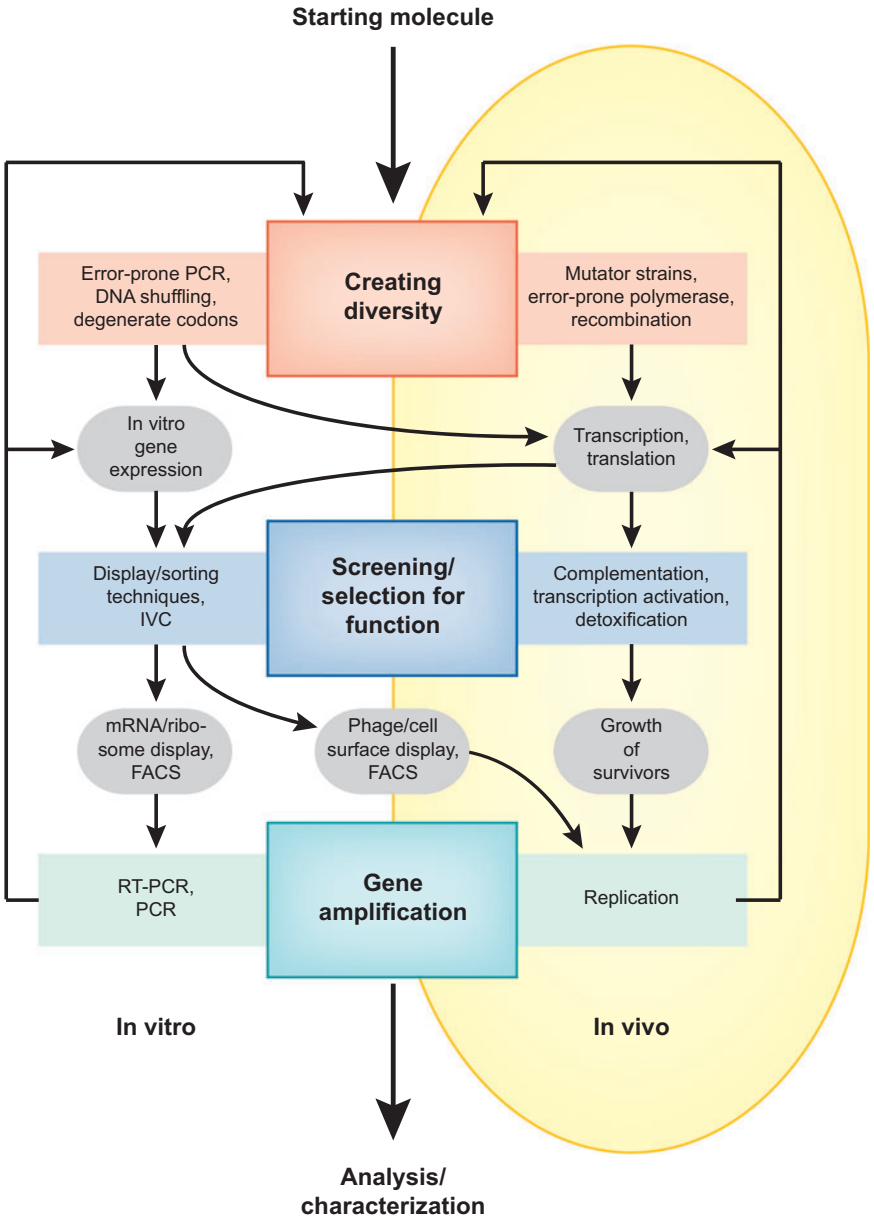


**Figure 1**

Schematic overview of the principal processes, strategies, and techniques of directed evolution. Today, numerous experimental methods are available to perform the fundamental processes of true Darwinian evolution (*central boxes*) in the laboratory, either in vivo within microorganisms or entirely in vitro in the test tube. Arrows indicate possible routes for connecting individual evolutionary steps. Abbreviations: PCR, polymerase chain reaction; RT-PCR, reverse transcription PCR; IVC, in vitro compartmentalization; FACS, fluorescence-activated cell sorting.

## TOOLS FOR DIRECTED EVOLUTION

A large arsenal of methodologies is available for carrying out evolution experiments in the laboratory, both within cells and in cell-free systems. **Figure 1** presents an overview of relevant strategies.

### Library Construction

Creating molecular diversity is the initial step in any directed evolution experiment. Until roughly 20 years ago, the humoral immune system was the readiest, most convenient source of protein libraries. In response to antigenic challenge, this powerful

protein-generating machinery produces antibody receptors that bind foreign molecules with high affinity and selectivity. By current estimates, the primary repertoire of the mammalian immune system contains at least 100 million different antibodies, each with a distinctive binding site (3). These further diversify by somatic mutation as the immune response to an antigen becomes more specific and refined (56). Although the immune system exists to protect the body against disease-causing microorganisms and toxins, its microevolutionary mechanisms can also be exploited to produce, essentially on demand, tailored monoclonal antibodies for diagnostics and medical applications (84). Immunization with appropriately designed template molecules (e.g., transition state analogs for particular reactions) has even afforded antibodies with a wide range of catalytic activities (28).

While the immune system is still unrivaled in biology in its ability to fashion protein receptors for virtually any natural or synthetic molecule, versatile new methods for creating protein libraries have emerged in the past few decades. Specifically, advances in chemical synthesis and molecular biology have made gene construction and modification routine. It is now practicable to create mutant proteins through targeted or random mutagenesis of the gene encoding an existing protein. The altered gene is introduced into a bacterial cell on a plasmid, a small circular DNA molecule. The progeny of such a cell, called a clone, produces the particular protein mutant specified by the modified gene. In numerous proteins, the structural and functional roles of individual amino acids have been dissected by replacing one codon with another. New properties are also engineered into proteins in this way. Furthermore, rather than producing and characterizing mutants one by one, large libraries of variants can be created in a single step using degenerate codons encoding all or subsets of the 20 proteinogenic amino acids for mutagenesis (9).

In principle, there is no a priori limit to the number of codons that can be randomized simultaneously. Multiple sites in an existing gene can be profitably targeted if relevant structural or functional information is available and if efficient means of sifting through the population are also at hand. It is even possible to use fully randomized gene sequences for discovery of new functions. Nevertheless, the sheer number of combinatorial possibilities rapidly outstrips most libraries even for small proteins. For instance, conventional building blocks alone could yield more than $10^{130}$ different 100-amino-acid-long polypeptides. Although gene libraries containing up to $10^{15}$ different members are accessible with current synthetic protocols, they constitute but a tiny fraction of sequence space, and technical hurdles abound. Substantially smaller ensembles are the norm.

A variety of complementary techniques introduce point changes at random sites within genes (see Reference 86 for a recent review). These include application of bacterial mutator strains and subjecting DNA to UV radiation or chemical mutagens. Error-prone PCR, which produces random mutations during PCR by reducing DNA polymerase fidelity, is perhaps the most widely used method. In contrast to more targeted approaches, such techniques do not require prior knowledge of structure or mechanism.

Among the most powerful strategies for the acquisition of new function in natural evolution is genetic recombination. Homologous recombination can similarly benefit iterative sequence evolution in vitro. DNA shuffling (68), which involves random fragmentation of a gene or a pool of related genes, followed by reassembly in a self-priming PCR reaction, helps bring together independent, beneficial mutations in a single gene, thereby streamlining exploration of sequence space. Variations on this basic method (49, 66, 78) have improved its implementation, facilitating creation of novel hybrid proteins.

The ideal strategy for gene diversification necessarily depends on the scientific questions asked. Frequently, a combination of randomization methods is required, especially if a gene sequence is subjected to multiple evolutionary rounds of optimization. Overall, however, chemists and biologists can now reliably create protein libraries that rival the complexity of the primary immunological repertoire. Moreover, these libraries are not restricted to the immunoglobulin scaffold. How can this diversity be exploited for protein design?

## Finding a Needle in the Haystack

In practice, identifying the relatively few molecules in a population that do something interesting, and separating them from the large number of variants that do not, is the greatest challenge in directed evolution. There are two fundamentally different approaches to address this problem. Libraries are assayed either by scrutinizing each member individually (screening) or by applying conditions that test all members simultaneously and allow only desirable clones to appear (selection) (59, 72). The choice of method depends on how frequently useful variants occur in the complex mixtures.

Depending on the difficulty of the task at hand—and the distance the starting structure is from the desired endpoint in sequence space—screening can be exploited with great effectiveness. Many sensitive chemical, biological, and immunological assays have been developed that enable rapid evaluation of the catalytic or binding activities of thousands of clones. These techniques are often based on the detection of colored or fluorescent reaction products, but high-performance liquid chromatography (HPLC), mass spectrometry, or NMR spectroscopy can also be employed if a suitable chromophore is not available (59). Although most screens accommodate analysis of $10^2$ to $10^4$ library members, much higher throughputs can be achieved with techniques like fluorescence-activated cell sorting (FACS), which can sample approximately $10^5$ clones per second (1). Because of its flexibility and given advances in automation and miniaturization, the importance of screening for laboratory evolution is likely to increase in the future.

**HPLC:** high-performance liquid chromatography

**FACS:** fluorescence-activated cell sorting

In classical selection experiments in vivo, which are often based on the ability of an encoded protein to complement a metabolic defect or provide resistance to some toxin, conditions are rigged so that only molecules with the desired function survive. As a consequence, many more library members ($>10^8$) can be analyzed simultaneously in a single experiment than is possible in a typical screen. Furthermore, by adjusting either substrate (35) or catalyst (46) concentration, the stringency of selection can be fine-tuned to differentiate enzyme variants encompassing a broad range of activities.

Although in vivo selection is often thought to be restricted to catalysts of natural biotransformations or closely related reactions, the set of selectable activities can be significantly expanded via modified yeast three-hybrid assays (38) or strategies that exploit product binding to other transcriptional regulators to give a selectable cellular readout (77). An even broader range of functions and environmental conditions can be accessed through in vitro formats. Phage display (34), cell surface display (21), ribosome display (25), and mRNA-protein fusions (62) are some of the powerful technologies that have been developed to couple genotype and phenotype beyond the boundaries of the cell. In vitro compartmentalization is another useful method (45, 71). Genes are encapsulated together with the protein biosynthesis apparatus in water-in-oil and water-in-oil-in-water emulsions, where they can be effectively subjected to selection for novel functions using many of the same techniques applied to cell-based systems.

Although the development of a viable selection system is generally more challenging than the setup of a screening procedure, the effort can pay off if the desired activity occurs only very rarely in large combinatorial libraries. The chief advantage of in vitro

relative to in vivo selection strategies is that much larger library sizes are accessible (up to $10^{15}$ versus $10^6$–$10^8$). Selection for binding functions is also usually straightforward. Selection for catalytic turnover can be more difficult, however, which is a disadvantage.

## Design versus Redesign

Nature had billions of years to create and optimize the polypeptides that make life as we know it possible. In adopting evolutionary strategies for the design of new proteins in the laboratory, where does one start? In principle, there are two options. On the one hand, these molecules can be created de novo, selecting functional entities directly from collections of random sequence. For this strategy, called protein design, completely naïve libraries can be used, in analogy to the presumed origins of biological proteins, or chemical information can be exploited to bias the design of such libraries to increase the probability of finding functional sequences. On the other hand, in protein redesign an existing protein molecule, which already folds to a stable structure and may possess suitable binding or catalytic properties, can be used as a starting template. Modifying its properties over multiple rounds of mutagenesis and screening/selection mimics the process of divergent evolution. As outlined below, both approaches have advantages and disadvantages. In both cases, modern computational methods may inform and extend experimental approaches. In silico screening can access much larger fractions of sequence space than is typically possible by experiment, facilitating identification of promising leads.

# PROTEIN DESIGN

## Selection from Deep Sequence Space

Experiments with random sequence libraries suggest that relatively little sequence information is required to form stable, cooperatively folding protein structures (17). Libraries of 70- to 90-residue polypeptides composed primarily of random combinations of glutamine, leucine, and arginine were constructed and screened for folded molecules. A surprisingly high fraction of the library members, corresponding to approximately 5% of the population of in-frame proteins, was found to have significant helical content and to resist both chemical and thermal denaturation as well as proteolytic degradation. Some nonnative biophysical properties, such as hyperstability in guanidinium chloride at 90°C and poor solubility in the absence of denaturants, were also exhibited. By reducing the leucine content of the ensemble to approximate the hydrophobicity of natural proteins, soluble helical proteins that undergo cooperative thermal denaturation were readily recovered (16). Judging from the limited chemical shift dispersion observed in one-dimensional $^1$H-NMR spectra and the rapid solvent exchange of all amide protons, however, these proteins are relatively loosely packed, much like molten globules.

While such findings have intriguing implications for the emergence of primordial proteins, polypeptides possessing a specific function are expected to occur substantially less frequently in sequence space. Indeed, much larger, more complex random libraries, together with powerful selection methods, have been needed to isolate proteins that bind small molecules or other proteins. For example, random libraries containing nearly $10^{13}$ members were used to isolate high-affinity streptavidin binding proteins, which are 88 amino acids in length (82). Seven rounds of in vitro selection by mRNA display yielded only 20 different binders. The isolated proteins contain a biotin-mimicking HPQ consensus motif and, in the best cases, recognize streptavidin with low nanomolar affinity, which is comparable to conventional antibody-antigen complexes. Experiments employing simpler libraries and shorter sequences gave substantially weaker binders.

In related experiments, mRNA display was used to isolate folded and functional proteins that recognize ATP with high affinity

and selectivity (33). Starting from an unbiased library of $6 \times 10^{12}$ proteins containing 80 randomized positions, four distinct families of binders were enriched by affinity selection on ATP-agarose. Functional proteins are evidently surprisingly common in random sequence space, occurring at a frequency as high as 1 in $10^{11}$ library molecules. Similar frequencies have been observed for the isolation of ATP binding RNA aptamers from oligonucleotide libraries, suggesting that the greater functional diversity of 20 amino acids, compared to four nucleotides, effectively compensates for the dramatic increase in combinatorial possibilities for polypeptides.

The initially identified ATP binders were diversified by mutagenic PCR and subjected to additional rounds of in vitro selection and amplification (33). After a total of 18 cycles, only derivatives of a single ancestor molecule survived. The strongest binder recognized ATP with a $K_d$ value of 100 nM, while significantly lower affinities were observed for other nucleotide analogs. Not unexpectedly, given its origin, the active protein adopts a fold that has not been found in nature. Crystallographic analysis shows that it consists of a three-stranded antiparallel β-sheet with two nonadjacent α-helices (**Figure 2**) (70). In addition to having a small but native-like hydrophobic core, the protein is stabilized by $Zn^{2+}$ coordination to two invariant CXXC motifs. ATP recognition occurs in a broad cleft at the surface of the protein via a large number of van der Waals, π-stacking, and electrostatic interactions.

These experiments demonstrate the feasibility of isolating structurally and functionally novel proteins from completely unbiased random sequences. Although relatively large libraries ($>10^{12}$ members) may be required, powerful selection methods such as mRNA display enable adequate sampling of sequence space and subsequent optimization of promising leads. Such experiments represent heroic undertakings, however, and extensions to molecular recognition of other ligands, much less catalysis of chemical transformations, are likely to be nontrivial.

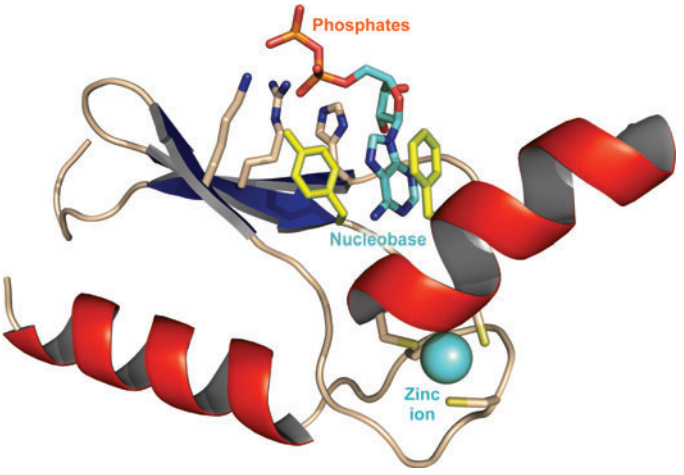


**Figure 2**

Artificial nucleotide binding protein with bound ADP (PDB code: 1UW1). The protein fold is stabilized by the coordination of a zinc ion to four cysteine residues. Ligand recognition involves π-stacking interactions between the nucleobase and two aromatic residues (*yellow*) and Coulombic interactions between the phosphates and positively charged amino acids.

## De Novo Proteins Biased to Specific Folds

Including basic chemical or structural information in the design of combinatorial sequence libraries can greatly increase the probability of finding folded proteins. For example, the sequence preferences of helices and sheets, or the tendency of hydrophobic residues to be buried in the protein interior, can be exploited for the creation of topologically predefined α-helical and β-sheet proteins (27).

In one prominent example, degenerate gene libraries were designed for four-helix-bundle proteins by specifying the sequence locations, but not the identities, of hydrophobic and hydrophilic residues (30). Although the prescribed polar/nonpolar binary pattern did not define detailed packing interactions, 60% of the library genes could be expressed in a bacterial host. The resulting proteins were resistant to intracellular proteolysis,

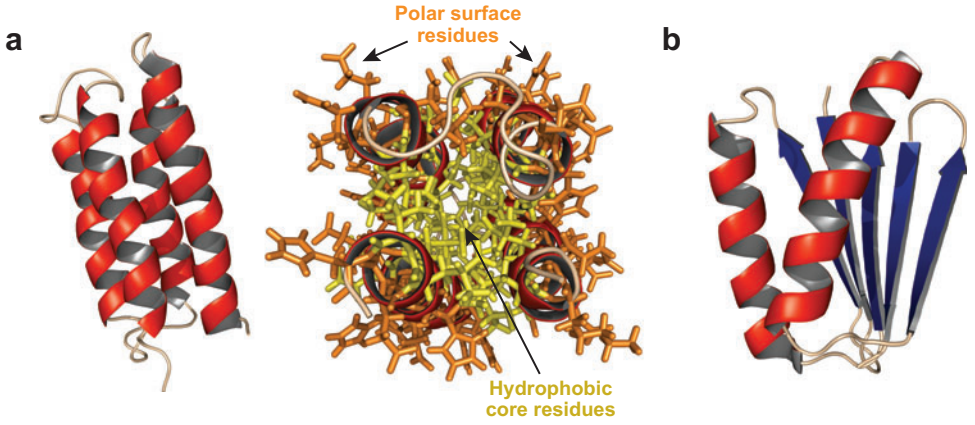


**Figure 3**

De novo protein design. (*a*) Four-helix-bundle protein designed by binary patterning (PDB code: 1P68; *left*, side view; *center*, view along the helix axis). Well-packed side chains in the hydrophobic core of the protein are shown in yellow, and polar surface residues are depicted in orange. (*b*) Computationally designed α/β-fold of Top7 (PDB code: 1QYS).



suggesting formation of stable globular structures. Subsequent biophysical analysis confirmed that they adopt compact helical folds, albeit with fluctuating structures, reminiscent of the Gln/Leu/Arg polypeptides described above. By lengthening the helices, the fraction of molecules that behave like molten globules could be substantially reduced. Structural studies on one of these second-generation proteins verified that it adopts an antiparallel four-helix-bundle fold, as designed, with nonpolar side chains buried and polar side chains exposed to solvent (**Figure 3*a***) (81). The backbone and buried residues were well ordered, showing that binary patterning is a viable strategy for generating native-like tertiary structures.

Although the binary-patterned helical bundles were not subjected to genetic selection or high-throughput screening, they exhibit a range of serendipitous functions. For instance, roughly half were found to bind heme (63) and displayed characteristic activities associated with this cofactor, such as carbon monoxide binding and peroxidase chemistry (27). Patterned proteins have also been shown to hydrolyze activated esters with efficiencies similar to many catalytic antibodies (80). The reported activities, while generally modest in comparison with true enzymes, reflect the chemical potential inherent in all polypeptides. They suggest that patterned combinatorial libraries could be a rich source of starting structures for evolving even more sophisticated receptors and catalysts.

The search for amino acid sequences able to adopt a specified tertiary fold has also been successfully tackled by computation. Powerful algorithms now allow efficient screening of virtual libraries that far exceed what is possible experimentally. For example, the primary sequence of a 28-amino-acid-long zinc finger domain, consisting of a ββα motif, was completely redesigned by computational screening of $>10^{27}$ sequences (15). The design of proteins with unnatural folds, including helical coiled-coil proteins with unprecedented right-handed twists (26) and a 93-amino-acid protein with a more complex mixed α/β-fold (36), illustrates the enormous potential of the computational approach. In the case of the α/β molecule, fragments from known protein structures were used to convert a two-dimensional cartoon of the target fold into a three-dimensional model. A general computational strategy that iterates

between sequence design and structure prediction was then used to optimize backbone and side chain conformations. The sequence with the lowest calculated energy after many rounds of design and optimization, Top7, has no significant homology to any known protein sequence. However, biophysical characterization and X-ray crystallography revealed an extremely stable and well-ordered structure that closely matches the intended design (**Figure 3*b***).

The computational design process mimics evolution in its iterative nature, with the calculated stability of the fold serving as the selection criterion for passing a design on to the next round of optimization. Because of their enormous throughput, such methods provide access to regions of sequence space that would be difficult to chart using purely experimental techniques. Moreover, as algorithms improve, function is sure to follow the success with structure, as indicated by the de novo generation of primitive diiron proteins that catalyze a phenol oxidase reaction (31). When combined with laboratory evolution, proteins with novel and useful properties should emerge. Experiment is essential in this context, because one can never be certain in advance that a predetermined fold will be compatible with the desired function.

## PROTEIN REDESIGN

Although certain novel features may require de novo design, it is not always necessary to reinvent the wheel to evolve new proteins with interesting physical or practical properties. Natural evolution has provided us with a vast array of proteins possessing defined structures and diverse functions, including enzymes, channels, transporters, receptors, and molecular machines. In analogy to the process of divergent evolution, mutagenesis and selection can be used to alter or fine-tune the properties of these biomacromolecules.

An impressive number of enzymes have already been optimized using directed evolution techniques (54, 57, 74). Stability, tolerance to nonnatural conditions, substrate specificity and product selectivity, and enantioselectivity are some of the properties that have been modified through iterative cycles of mutagenesis and screening/selection. In some cases, surprisingly few mutations suffice to alter function in dramatic fashion. In others, improvements are achieved gradually over many small steps. In one striking example of the latter, the substrate recognition features of Cre recombinase were evolved over 126 evolution cycles to generate a variant that efficiently and selectively excises integrated HIV proviral DNA from the human genome (64). Either way, provided a suitable assay is in place, this type of experiment can now be considered almost routine.

Existing proteins also represent excellent templates for investigating fundamental questions about structure-function relationships. Large leaps away from a starting sequence can be made to gain insight into evolutionary origins, the sequence determinants of folding, and general biophysical properties. In addition, existing proteins can be endowed with completely new activities unrelated to their starting function.

### Simplified Proteins

Although biological proteins are built from a standard set of 20 amino acids, it is generally assumed that their primordial ancestors were constructed from considerably smaller numbers of building blocks. Hypotheses regarding the evolution of the modern genetic code from simpler systems coding for relatively few amino acids (48) find support in studies demonstrating the feasibility of generating native-like proteins from reduced alphabets.

The reconstruction of an SH3 β-barrel domain from a limited set of 14 different residues is an early example of protein simplification (60). Larger and structurally more complex α/β enzymes have also been successfully simplified using combinatorial mutagenesis and genetic selection. For example,

amino acid usage in orotate phosphoribosyltransferase was reduced to a 13-amino-acid set (2), and "reverse engineering" of triosephosphate isomerase demonstrated that 77% of the amino acid sequence could be replaced by a seven-letter amino acid alphabet (67). Little effect on the biophysical characteristics or catalytic activity of these proteins was observed.

A thermostable chorismate mutase from *Methanococcus jannaschii* has yielded the most dramatically simplified enzyme to date. The wild-type catalyst adopts a homodimeric, α-helical bundle fold (**Figure 4*a***) (42). Because it accelerates an essential step in the biosynthesis of aromatic amino acids, genetic selection (32) could be used to assess the tolerance of the secondary structural elements to sequence change. Keeping the loop sequences and six highly conserved active-site residues constant, all the helices in the molecule were replaced by binary-patterned modules of randomized sequence in a two-stage selection procedure (73). Wild-type residues were replaced according to their polarity with restricted sets of four polar residues (Glu, Asp, Asn, Lys) or four nonpolar residues (Leu, Ile, Met, Phe). Simplified yet active enzymes were identified by their ability to complement the chorismate mutase deficiency of a genetically engineered bacterial host. These simplified catalysts possess biophysical and kinetic characteristics similar to those of the parent protein, even though only eight different amino acids constitute more than 90% of their sequence. Sequence analysis of active clones revealed packing constraints and second-shell interactions that are critical for producing a functional active site in this helical bundle scaffold, which would have been difficult to discern by crystal gazing alone.

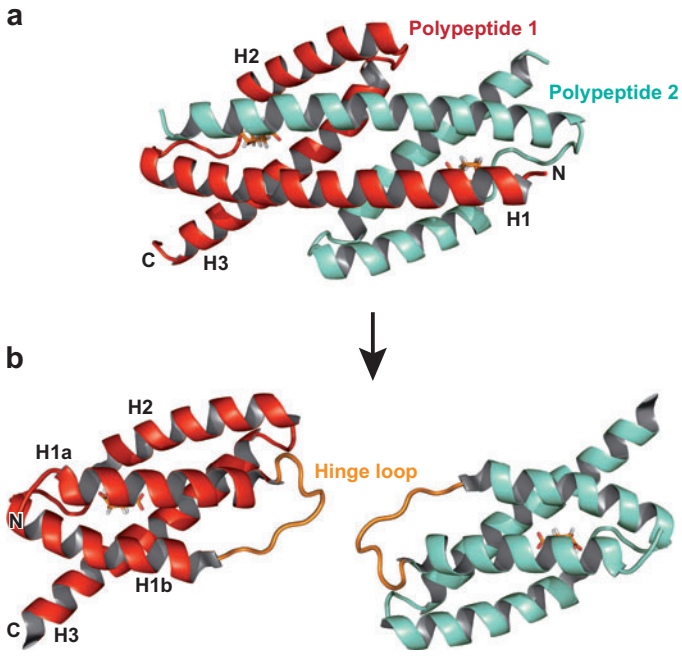


**Figure 4**

Alteration of protein quaternary structure. (*a*) The prototypic helical bundle chorismate mutase (PDB code: 1ECM) consists of two identical polypeptides. Residues from both subunits are required to form functional active sites. (*b*) The dimer was converted into monomers by insertion of an eight-amino-acid hinge loop into the middle of helix H1. In the monomers (PDB code: 2GTV), the N-terminal half of helix H1 (H1a) folds back onto itself and displaces the corresponding segment from the second polypeptide. An oxabicyclic transition state analog, which binds at the active site, is shown in stick representation.

Although it was possible to isolate binary-patterned enzymes, such active helical bundle chorismate mutases are extremely rare in sequence space. Extrapolating from the selection statistics, it was estimated that a library of nearly $10^{24}$ members would have been needed to obtain functional mutases had every position in the protein been randomized simultaneously. Although the experiments are not strictly comparable, the size of such a library is many orders of magnitude larger than the library described above that was used to identify noncatalytic ATP binding proteins from random sequences (33), underscoring the high demands that catalysis places on a protein scaffold.

For further simplification of the binary-patterned mutase, loop residues and an active-site glutamine were systematically replaced with amino acids from the restricted alphabet (79). These experiments yielded an active enzyme constructed entirely from a set of only nine amino acids. It is fully functional in the cell, replacing wild-type chorismate mutase in bacterial metabolism. Moreover, the

nine-amino-acid protein was shown through biophysical characterization to capture significant features of its natural counterpart, although it appears to be somewhat less stable and more flexible. More than half of the 20 proteinogenic amino acids can evidently be removed from the biosynthetic tool kit without sacrificing significant function.

The generation of active biocatalysts from dramatically reduced amino acid alphabets provides strong support for the idea that primordial enzymes were made from only a handful of building blocks. Nevertheless, numerous questions remain. Many alternative alphabets could conceivably afford active catalysts. How few amino acids would be needed? What are the best sets? To what extent does the answer depend on the choice of starting fold? And, finally, does reduced sequence diversity have generally adverse consequences for folding or function? Answering these questions experimentally promises to be a daunting task. However, as we have seen, computational methods enable rapid evaluation of enormous numbers of sequences. They have the potential to guide the design of more focused experiments and can be expected to facilitate analysis of the subtle patterns of variation and covariation in functional sequences.

## New Quaternary Structures

Many natural proteins possess multiple subunits (23). Their higher-order quaternary structure provides a variety of benefits relative to the corresponding monomers, ranging from increased stability and the availability of intersubunit binding sites to the possibility of cooperative interactions and allosteric regulation. Insights into the evolutionary origins of such species have been gained from efforts to convert multimers into monomers and monomers into multimers.

For some oligomeric proteins, quaternary structure appears to be the product of a process dubbed domain swapping, which involves replacement of a unit of structure, or domain, in one polypeptide with the same domain from an identical polypeptide (6). The result is an intertwined dimer or higher-order oligomer. The helical bundle chorismate mutase discussed in the previous section is an example of such a protein. Insertion of randomized hinge loops into the middle of the long H1 helix that spans the dimer (**Figure 4*a***), and selecting for functional variants by genetic complementation, enabled rational redesign of mutase topology (43).

**mMjCM:** monomeric *Methanococcus jannaschii* chorismate mutase

Not surprisingly, the sequence and length of the inserted loop profoundly influence quaternary structure. While no active chorismate mutases were found with six-residue segments, hinge loops containing either four or seven amino acids converted the *Escherichia coli* enzyme into monomers with marginal stability but near wild-type activity. In contrast, a five-residue insert afforded stable but weakly active hexamers. Subjecting the latter to further rounds of random mutagenesis and selection resulted in a >600-fold increase in catalytic efficiency and, in the process, converted the hexamer to a trimer (46, 75). Apparently, the transition between many different oligomeric states—monomers, dimers, trimers, and hexamers—can be induced by a small number of seemingly innocuous changes in primary sequence, at least in the case of these helical bundle proteins.

A stable monomeric mutase with near native activity (mMjCM, or monomeric *Methanococcus jannaschii* chorismate mutase) was obtained through analogous selection experiments from libraries containing an eight-amino-acid-long hinge loop inserted into the thermostable enzyme from *M. jannaschii* (**Figure 4**) (44). Only a relatively small fraction of all possible loop sequences (<0.05%) produced active monomers, so biophysical and structural studies were carried out to investigate how the selected sequence (Ala-Arg-Trp-Pro-Trp-Ala-Glu-Lys) influences protein structure. Although highly active catalytically, the engineered monomer unexpectedly exhibits all the characteristics of a nonnative molten globule state, including poor NMR signal dispersion, rapid

hydrogen/deuterium exchange, noncooperative thermal denaturation, and binding of hydrophobic fluorophores (76). Binding of a transition state analog at the active site leads to some ordering, making determination of the structure of the complex by NMR spectroscopy possible (52). The monomer assumes a helix-bundle fold that closely matches the original design (**Figure 4**), with an active site geometry similar to that of its dimeric *E. coli* homolog. However, it retains unusual flexibility on the millisecond timescale.

The discovery of a disordered enzyme with near native activity is unprecedented. This finding contravenes the traditional view that efficient catalysis requires a well-defined three-dimensional structure. Indeed, it appears that enzyme catalysis can be added to the growing list of possible functions ascribed to intrinsically disordered proteins in biology (20). The observations with mMjCM dovetail with speculations that modern-day enzymes may have evolved from catalytic molten globules (19). They also speak to the idea that promiscuous activities may fuel natural enzyme evolution (29). Insofar as structural plasticity confers relaxed specificity, it is conceivable that molten globules were useful starting points for the evolution of new catalysts in nature. Analogously, they may also be superior to highly evolved enzymes as templates for laboratory evolution experiments.

## Topological Redesign

Circular permutation of an existing enzyme is one of the simplest ways to create a new fold. This type of topological rearrangement involves an internal relocation of the N and C termini of the protein rather than a complete reshuffling of secondary structural elements. It is believed to be important for the evolution of natural proteins (24) and has been mimicked in the laboratory to study folding pathways and alter enzyme activity (55).

One model for the origin of circularly permuted proteins in nature involves a gene duplication event, followed by in-frame fusion of the two copies, and subsequent shortening of the new 5′ and 3′ ends of the fused gene (41). The feasibility of this permutation-by-duplication mechanism was established in a laboratory evolution experiment with *Hae*III methyltransferase (51). Two copies of the wild-type M. *Hae*III gene were fused, and separate libraries of 5′- and 3′-truncated genes were generated by randomly introducing start and stop codons, respectively. The functional proteins that were selected in vivo recapitulated hypothetical evolutionary pathways linking different families of known DNA methyltransferases. In addition to showing that new protein topologies can arise via multistep gene rearrangements, this work emphasizes the inherent modularity of proteins. This modular nature likely facilitates the emergence of new folds.

Combinatorial shuffling of polypeptide segments can also yield novel folded domains. For example, native-like proteins have been generated directly by combinatorial assembly of segments from nonhomologous proteins using phage display (61). In other experiments, a nonhomologous random recombination method was used to alter helical connectivities in the engineered monomeric mutase, mMjCM (10). The gene encoding the enzyme was fragmented and the fragments were religated in random order. Following in vivo selection for chorismate mutase activity, functional variants that contained many insertions, deletions, and rearrangements were obtained (**Figure 5**). The topologically altered enzymes included circular permutants and rearranged enzymes with secondary structure connectivities that had not been seen previously. Intriguingly, one of the topological variants created in this laboratory experiment bears a striking resemblance to the fold discovered subsequently in nature for the exported chorismate mutase of *Mycobacterium tuberculosis* (47). Nevertheless, it was not possible to swap secondary structural elements in mMjCM with analogous elements from the unrelated enzyme fumarase. The only functional chimeras obtained from attempts to

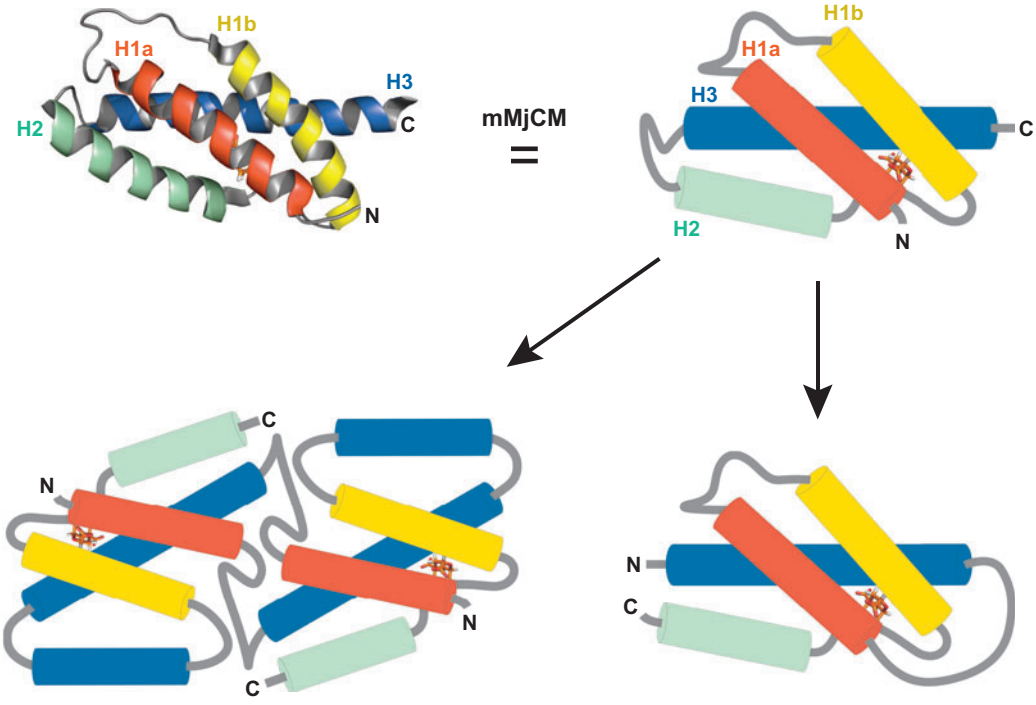


**Figure 5**

Topologically rearranged variants of the monomeric chorismate mutase (mMjCM). Structural models of some of the modified proteins obtained by nonhomologous random recombination are depicted. Loops are shown in gray; helices H1a, H1b, H2, and H3 are distinguished by different colors as labeled. A transition state analog in stick representation shows the approximate location of the active sites.

recombine mMjCM and fumarase had fumarase sequences localized at the termini or in one loop.

The experiments on the topological redesign of enzymes show that secondary structural elements can be combined in many different ways, much like LEGO® bricks, to generate diverse tertiary folds. In addition to identifying functionally important structural elements, new protein folds not (yet) discovered in nature can be created.

## New Activities in Old Enzymes

In the previously cited examples of protein redesign, the native activity of the polypeptide chain was preserved and exploited as a means of evaluating successful folding of modified structures. Introduction of new activities into existing scaffolds represents a potentially more demanding task because it requires identification of new constellations of functional groups, positioned precisely to orchestrate a sequence of delicately timed chemical events.

Toward the goal of creating catalysts with new functions, minor promiscuous side reactions of an existing enzyme can often be enhanced through directed evolution at the expense of the original transformation (12). The reprogramming of the nonselective enzyme γ-humulene synthase, which produces 52 different sesquiterpenes from farnesyl diphosphate, to produce individual products with high selectivity is an impressive example of the potential of this approach (85).

Transplanting catalytic activity between enzymes that share the same fold and perform

mechanistically similar chemical functions can also be relatively straightforward. In many cases, only a few mutations, often focused at the active site, are needed to gain new activity (74). The functional interconversion of various enzymes within the enolase superfamily is illustrative (22). In other cases, more extensive modification may be necessary.

Simple homology is no guarantee of success, as we discovered in attempts to convert a chorismate mutase into an isochorismate pyruvate lyase. These natural enzymes catalyze unrelated pericyclic reactions (18), but sequence comparisons (37) and structural studies (87) showed that they are closely related. Although their similarity is particularly high in the active site, where 7 of 12 residues are conserved, not a trace of isochorismate pyruvate lyase activity could be detected in chorismate mutase variants mutated to resemble more closely the lyase active site (K. Höland, K.J. Woycechowsky, C. Jäckel & D. Hilvert, unpublished data). Distant mutations apparently have a much stronger influence on active site chemistry in this system than could have been anticipated from comparisons of the nearly identical tertiary and quaternary structures of the two enzymes.

The conversion of glyoxalase II, a metallohydrolase, to a metallo-β-lactamase further illustrates the difficulties that are sometimes encountered in transplanting unrelated activities between homologous enzymes (50). Although a catalyst was obtained that increased resistance of *E. coli* to cefotaxime roughly 100-fold, extensive modifications of the template were required. The incorporation and adjustment of the requisite functional elements entailed insertion, deletion, and substitution of several active site loops, as well as selection of diverse point mutations to fine-tune activity. The final variant bore only 59% amino acid identity to the starting template.

The benefit of large numbers is evident in the selection and evolution of an active RNA ligase from a partially randomized, noncatalytic zinc finger scaffold (65). A library containing $>10^{12}$ independent sequences was generated by randomizing two loops in the retinoid-X receptor, and successful catalysts were isolated by mRNA display. They catalyze the $Zn^{2+}$-dependent ligation of a 5′-triphosphorylated RNA oligonucleotide to the terminal 3′-hydroxyl group of a second RNA, a reaction for which no natural enzyme is known. Although the catalytic activity is modest in absolute terms, multiple turnovers and rate enhancements of $>10^6$ were achieved. Related schemes can be imagined for selecting enzymes that promote other bond-making and bond-breaking reactions without explicit knowledge of structure or mechanism.

## Computational Redesign

Like de novo protein design, the adaptation of existing proteins to new tasks stands to profit greatly from advances in computational methods. Known scaffolds can be effectively redesigned using algorithms that have been developed to facilitate the systematic exploration of amino acid sequence and side chain rotameric preferences (14, 40, 53, 88). Once a starting structure has been chosen from the Protein Data Bank, the design process often begins by simulated docking of the target ligand—be it a protein, small molecule, or transition state—into a potential active site in the protein. A subset of residues in and around this pocket is then randomized, and the design algorithm iteratively searches the conformational space of the side chains and ligand to minimize the energy of each possible sequence. From the list of sequences with the lowest energy predicted by the design program, a small number of proteins are produced and their activities are measured.

Novel receptors for proteins and small molecules have been successfully designed in this way. The computational redesign of the ligand specificities of several members of the periplasmic binding protein superfamily is particularly notable (39). Binding of structurally diverse molecules such as trinitrotoluene (TNT), L-lactate, or serotonin with

low nanomolar to micromolar affinities establishes the high degree of control over biomolecular recognition that can be achieved with these methods.

The first steps toward nascent enzymes have also been taken. For example, primitive iron and copper binding sites, which interact with $O_2$ and $H_2O_2$, have been designed into proteins such as thioredoxin (8) and maltose binding protein (7). In other work, variants of thioredoxin were designed that hydrolyze *p*-nitrophenyl acetate, using histidine as a nucleophile (11). Although the first-generation catalysts only achieved a ~25-fold rate acceleration compared with the reaction with 4-methylimidazole in solution, they demonstrate the feasibility of conferring activity on otherwise inert scaffolds. Extrapolating from the successful recapitulation of natural active site geometries (88), automated sequence design has great potential to guide efforts to generate and optimize catalytic efficiency in such molecules and thus complement experimental evolutionary approaches.

Although the field of computational protein (re)design is relatively young, and convincing cases of efficient catalyst design have yet to be reported, much can be expected in the future. With improvements in the algorithms for creating and ranking the best combinatorial designs, its promising beginnings augur well for the ultimate success of this strategy.

## PERSPECTIVES AND PROSPECTS

Rational protein design is an enormous challenge. The innate complexity of these biopolymers has made a detailed understanding of sequence-structure-function relationships elusive. Not only is structural information distributed nonrandomly across protein sequence, these molecules can also assume many energetically similar conformational states. In the specific case of enzymes, our grasp of the dynamic, synergistic interactions between substrate and protein functional groups is still insufficient to explain the incredible catalytic prowess they display.

In nature, proteins were not designed. Rather, they arose over many steps of Darwinian evolution. These same evolutionary strategies can be imitated in the laboratory, on the human timescale, to tailor existing proteins and isolate new ones, as examples presented in this review attest. Indeed, methods exist that mimic each step of the evolutionary cycle (**Figure 1**). While improvements in the synthesis and handling of large gene libraries can still be envisaged, the problem of generating molecular diversity is essentially solved. The development of suitable selection and screening methods for searching through these libraries is a particularly critical step, but again many alternatives are available and new assay formats are being continually invented. Conceivably, given its iterative nature, the entire evolutionary process could even be automated to speed the discovery cycle.

The principal practical success of directed evolution has been in the area of enzyme engineering. The utility of this approach for tailoring the stability, specificity, and reaction conditions of existing protein catalysts is now well established (54, 57, 74). The challenge for the future will be to extend evolutionary methods to more complex problems, such as creation of catalysts for reactions that lack biological enzymes; production of new macromolecular structures including molecular machines and nonbiological polymers; and evolution of signal transduction systems, biosynthetic pathways, and artificial cells. The first steps in these directions have been taken (5), but we still have far to go.

For protein design, future progress will depend critically on optimizing strategies for productively exploring sequence space. Even though large libraries can be made and searched, there are far too many possibilities for every molecule to be physically made, much less evaluated individually. A key goal must therefore be to increase the frequency of useful/desirable variants in a population to a searchable number (1 in $10^{15}$) and then

sort, select, and evolve to make them relatively common (1 in 10). To that end, binary patterning (27), alternative amino acid alphabets (79), interaction networks (69), and knowledge of structure and mechanism can effectively guide the design of biased libraries for engineering protein properties. The wealth of information emerging from genome sequencing and structural genomics projects can also be exploited for targeted mutagenesis to improve the probability of finding mutant sequences with desirable attributes.

Using computational methods (13) as an adjunct to directed evolution may ultimately be the most powerful way forward. The two share the general evolutionary feature of iterative rounds of mutation and selective amplification of functional variants, but computation can accommodate libraries much larger than is experimentally practicable. Although still somewhat primitive, design algorithms will someday map regions of sequence space now out of experimental reach. Allied, computation and evolution will be extremely powerful. Computation will point the way; experimental evolution will navigate and optimize it.

Directed evolution allows experimentalists to explore protein functions that are poorly understood and to create macromolecules without counterparts in nature. In the future, intelligent application of evolutionary methods will clarify the links between sequence and structure and between structure and function. Medical and industrial applications will surely follow.

## SUMMARY POINTS

1. Natural Darwinian evolution can be effectively imitated in the laboratory toward tailoring protein properties for diverse applications, including catalysis.
2. Many experimental technologies exist to produce large and diverse protein libraries and for screening or selecting folded, functional variants.
3. Folded proteins arise surprisingly frequently in random sequence space, an occurrence exploitable in evolving selective binders and catalysts.
4. While more conservative than direct selection from deep sequence space, redesign of existing proteins by random mutagenesis and selection/screening is a particularly robust method for optimizing or altering extant properties. It also represents an excellent starting point for achieving more ambitious engineering goals.
5. Allying experimental evolution with modern computational methods is likely the broadest, most fruitful strategy for generating functional macromolecules unknown in nature.

## FUTURE ISSUES

1. Creative new strategies are needed for productively biasing random sequence libraries to increase the probability of finding variants with interesting functional properties.
2. Despite impressive advances, automated screening of virtual sequence libraries is not yet completely reliable and remains far from routine. More accurate mathematical parameters to calculate and rank the energies of simulated structures are needed. Improving the outcomes of such procedures will require the development of algorithms that better describe the electrostatics of buried residues and dynamics of (catalytic) side chains, enable backbone flexibility, and predict bound water molecules.

3. Extension of evolutionary strategies to more complex problems, such as the creation of catalysts for completely new reactions, production of macromolecular machines and encoded nonnatural polymers, and engineering networks of interacting proteins, constitutes an exciting challenge for the future.

## DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

The authors thank Dr. Kenneth Woycechowsky for stimulating discussions. This work was generously supported by the ETH Zurich, the Schweizerischer Nationalfonds, and through a fellowship to CJ from the Deutsche Forschungsgemeinschaft.



## LITERATURE CITED

1. Aharoni A, Griffith AD, Tawfik DS. 2005. High-throughput screens and selections of enzyme-encoding genes. *Curr. Opin. Chem. Biol.* 9:210–16
2. Akanuma S, Kigawa T, Yokoyama S. 2002. Combinatorial mutagenesis to restrict amino acid usage in an enzyme to a reduced set. *Proc. Natl. Acad. Sci. USA* 99:13549–53
3. Alt FW, Blackwell TK, Yancopoulos GD. 1987. Development of the primary antibody repertoire. *Science* 238:1079–87
4. Arnold FH. 1998. Design by directed evolution. *Acc. Chem. Res.* 31:125–31
5. Arnold FH. 2001. Combinatorial and computational challenges for biocatalyst design. *Nature* 409:253–57
6. Bennett MJ, Choe S, Eisenberg D. 1994. Domain swapping: entangling alliances between proteins. *Proc. Natl. Acad. Sci. USA* 91:3127–31
7. Benson DE, Haddy AE, Hellinga HW. 2002. Converting a maltose receptor into a nascent binuclear copper oxygenase by computational design. *Biochemistry* 41:3262–69
8. Benson DE, Wisz MS, Hellinga HW. 2000. Rational design of nascent metalloenzymes. *Proc. Natl. Acad. Sci. USA* 97:6292–97
9. Besenmatter W, Kast P, Hilvert D. 2004. New enzymes from combinatorial library modules. *Methods Enzymol.* 388:91–102
10. Bittker JA, Le BV, Liu JM, Liu DR. 2004. Directed evolution of protein enzymes using nonhomologous random recombination. *Proc. Natl. Acad. Sci. USA* 101:7011–16
11. Bolon DN, Mayo SL. 2001. Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. USA* 98:14274–79
12. Bornscheuer UT, Kazlauskas RJ. 2004. Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways. *Angew. Chem. Int. Ed.* 43:6032–40
13. Butterfoss GL, Kuhlman B. 2006. Computer-based design of novel protein structures. *Annu. Rev. Biophys. Biomol. Struct.* 35:49–65
14. Dahiyat BI, Mayo SL. 1996. Protein design automation. *Protein Sci.* 5:895–903
15. Dahiyat BI, Mayo SL. 1997. De novo protein design: fully automated sequence selection. *Science* 278:82–87

16. Davidson AR, Lumb KJ, Sauer RT. 1995. Cooperatively folded proteins in random sequence libraries. *Nat. Struct. Biol.* 2:856–64
17. Davidson AR, Sauer RT. 1994. Folded proteins occur frequently in libraries of random amino acid sequences. *Proc. Natl. Acad. Sci. USA* 91:2146–50
18. DeClue MS, Baldridge KK, Künzler DE, Kast P, Hilvert D. 2005. Isochorismate pyruvate lyase: a pericyclic reaction mechanism? *J. Am. Chem. Soc.* 127:15002–3
19. DeGrado WF. 1993. Catalytic molten globules. *Nature* 365:488–89
20. Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell. Biol.* 6:197–208
21. Georgiou G, Stathopoulos C, Daugherty PS, Nayak AR, Iverson BL, Curtiss R 3rd. 1997. Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nat. Biotechnol.* 15:29–34
22. Gerlt JA, Babbitt PC, Rayment I. 2005. Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. *Arch. Biochem. Biophys.* 433:59–70
23. Goodsell DS, Olson AJ. 2000. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* 29:105–53
24. Grishin NV. 2001. Fold change in evolution of protein structures. *J. Struct. Biol.* 134:167–85
25. Hanes J, Plückthun A. 1997. In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA* 94:4937–42
26. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. 1998. High-resolution protein design with backbone freedom. *Science* 282:1462–67
27. Hecht MH, Das A, Go A, Bradley LH, Wei Y. 2004. De novo proteins from designed combinatorial libraries. *Protein Sci.* 13:1711–23
28. Hilvert D. 2000. Critical analysis of antibody catalysis. *Annu. Rev. Biochem.* 69:751–93
29. James LC, Tawfik DS. 2003. Conformational diversity and protein evolution: a 60-year-old hypothesis revisited. *Trends Biochem. Sci.* 28:361–68
30. Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. 1993. Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262:1680–85
31. Kaplan J, DeGrado WF. 2004. De novo design of catalytic proteins. *Proc. Natl. Acad. Sci. USA* 101:11566–70
32. Kast P, Asif-Ullah M, Jiang N, Hilvert D. 1996. Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: the importance of electrostatic catalysis. *Proc. Natl. Acad. Sci. USA* 93:5043–48
33. Keefe AD, Szostak JW. 2001. Functional proteins from a random-sequence library. *Nature* 410:715–18
34. Kehoe JW, Kay BK. 2005. Filamentous phage display in the new millennium. *Chem. Rev.* 105:4056–72
35. Kleeb AC, Edalat MH, Gamper M, Haugstetter J, Giger L, et al. 2007. Metabolic engineering of a genetic selection system with tunable stringency. *Proc. Natl. Acad. Sci. USA* 104:13907–12
36. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science* 302:1364–68
37. Künzler DE, Sasso S, Gamper M, Hilvert D, Kast P. 2005. Mechanistic insights into the isochorismate pyruvate lyase activity of the catalytically promiscuous PchB from combinatorial mutagenesis and selection. *J. Biol. Chem.* 280:32827–34
38. Lin H, Cornish VW. 2002. Screening and selection methods for large-scale analysis of protein function. *Angew. Chem. Int. Ed.* 41:4402–25
39. Looger LL, Dwyer MA, Smith JJ, Hellinga HW. 2003. Computational design of receptor and sensor proteins with novel functions. *Nature* 423:185–90

40. Looger LL, Hellinga HW. 2001. Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. *J. Mol. Biol.* 307:429–45
41. Luger K, Hommel U, Herold M, Hofsteenge J, Kirschner K. 1989. Correct folding of circularly permuted variants of a beta alpha barrel enzyme in vivo. *Science* 243:206–10
42. MacBeath G, Kast P, Hilvert D. 1998. A small, thermostable, and monofunctional chorismate mutase from the archaeon *Methanococcus jannaschii*. *Biochemistry* 37:10062–73
43. MacBeath G, Kast P, Hilvert D. 1998. Probing enzyme quaternary structure by combinatorial mutagenesis and selection. *Protein Sci.* 7:1757–67
44. MacBeath G, Kast P, Hilvert D. 1998. Redesigning enzyme topology by directed evolution. *Science* 279:1958–61
45. Mastrobattista E, Taly V, Chanudet E, Treacy P, Kelly BT, Griffiths AD. 2005. High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chem. Biol.* 12:1291–300
46. Neuenschwander M, Butz M, Heintz C, Kast P, Hilvert D. 2007. A simple selection strategy for evolving highly efficient enzymes. *Nat. Biotechnol.* 25:1145–47
47. Ökvist M, Dey R, Sasso S, Grahn E, Kast P, Krengel U. 2006. 1.6 Å crystal structure of the secreted chorismate mutase from *Mycobacterium tuberculosis*: novel fold topology revealed. *J. Mol. Biol.* 357:1483–99
48. Osawa S, Jukes TH, Watanabe K, Muto A. 1992. Recent evidence for evolution of the genetic code. *Microbiol. Rev.* 56:229–64
49. Ostermeier M, Shim JH, Benkovic SJ. 1999. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol.* 17:1205–9
50. Park H-S, Nam S-H, Lee JK, Yoon CN, Mannervik B, et al. 2006. Design and evolution of new catalytic activity with an existing protein scaffold. *Science* 311:535–38
51. Peisajovich SG, Rockah L, Tawfik DS. 2006. Evolution of new protein topologies through multistep gene rearrangements. *Nat. Genet.* 38:168–74
52. Pervushin K, Vamvaca K, Vögeli B, Hilvert D. 2007. Structure and dynamics of a molten globular enzyme. *Nat. Struct. Mol. Biol.* 14:1202–6
53. Pokala N, Handel TM. 2005. Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. *J. Mol. Biol.* 347:203–27
54. Powell KA, Ramer SW, del Cardayré SB, Stemmer WP, Tobin MB, et al. 2001. Directed evolution and biocatalysis. *Angew. Chem. Int. Ed.* 40:3948–59
55. Qian Z, Lutz S. 2005. Improving the catalytic activity of *Candida antarctica* lipase B by circular permutation. *J. Am. Chem. Soc.* 127:13466–67
56. Rajewsky K, Förster I, Cumano A. 1987. Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science* 238:1088–94
57. Reetz MT. 2006. Directed evolution of enantioselective enzymes as catalysts for organic synthesis. *Adv. Catal.* 49:1–69
58. Reidhaar-Olson JF, Sauer RT. 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* 241:53–57
59. Reymond J-L, ed. 2005. *Enzyme Assays: High-Throughput Screening, Genetic Selection, and Fingerprinting*. Weinheim: Wiley-VCH
60. Riddle DS, Santiago JV, Bray-Hall ST, Doshi N, Grantcharova VP, et al. 1997. Functional rapidly folding proteins from simplified amino acid sequences. *Nat. Struct. Biol.* 4:805–9
61. Riechmann L, Winter G. 2000. Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. *Proc. Natl. Acad. Sci. USA* 97:10068–73

62. Roberts RW, Szostak JW. 1997. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. USA* 94:12297–302
63. Rojas NRL, Kamtekar S, Simons CT, McLean JE, Vogel KM, et al. 1997. De novo heme proteins from designed combinatorial libraries. *Protein Sci.* 6:2512–24
64. Sarkar I, Hauber I, Hauber J, Buchholz F. 2007. HIV-1 proviral DNA excision using an evolved recombinase. *Science* 316:1912–15
65. Seelig B, Szostak JW. 2007. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* 448:828–31
66. Sieber V, Martinez CA, Arnold FH. 2001. Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* 19:456–60
67. Silverman JA, Balakrishnan R, Harbury PB. 2001. Reverse engineering the $(\beta/\alpha)_8$ barrel fold. *Proc. Natl. Acad. Sci. USA* 98:3092–97
68. Stemmer WP. 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* 370:389–91
69. Süel GM, Lockless SW, Wall MA, Ranganathan R. 2003. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* 10:59–69
70. Lo Surdo P, Walsh MA, Sollazzo M. 2004. A novel ADP- and zinc-binding fold from function-directed in vitro evolution. *Nat. Struct. Mol. Biol.* 11:382–83
71. Tawfik DS, Griffiths AD. 1998. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* 16:652–56
72. Taylor SV, Kast P, Hilvert D. 2001. Investigating and engineering enzymes by genetic selection. *Angew. Chem. Int. Ed. Engl.* 40:3310–35
73. Taylor SV, Walter KU, Kast P, Hilvert D. 2001. Searching sequence space for protein catalysts. *Proc. Natl. Acad. Sci. USA* 98:10596–601
74. Toscano MD, Woycechowsky KJ, Hilvert D. 2007. Minimalist active-site redesign: teaching old enzymes new tricks. *Angew. Chem. Int. Ed.* 46:3212–36
75. Vamvaca K, Butz M, Walter KU, Taylor SV, Hilvert D. 2005. Simultaneous optimization of enzyme activity and quaternary structure by directed evolution. *Protein Sci.* 14:2103–14
76. Vamvaca K, Vögeli B, Kast P, Pervushin K, Hilvert D. 2004. An enzymatic molten globule: efficient coupling of folding and catalysis. *Proc. Natl. Acad. Sci. USA* 101:12860–64
77. van Sint Fiet S, van Beilen JB, Withold B. 2006. Selection of biocatalysts for chemical synthesis. *Proc. Natl. Acad. Sci. USA* 103:1693–98
78. Voigt CA, Martinez C, Wang Z-G, Mayo SL, Arnold FH. 2002. Protein building blocks preserved by recombination. *Nat. Struct. Biol.* 9:553–58
79. Walter KU, Vamvaca K, Hilvert D. 2005. An active enzyme constructed from a 9-amino acid alphabet. *J. Biol. Chem.* 280:37742–46
80. Wei Y, Hecht MH. 2004. Enzyme-like proteins from an unselected library of designed amino acid sequences. *Protein Eng. Des. Sel.* 17:67–75
81. Wei Y, Kim S, Fela D, Baum J, Hecht MH. 2003. Solution structure of a de novo protein from a designed combinatorial library. *Proc. Natl. Acad. Sci. USA* 100:13270–73
82. Wilson DS, Keefe AD, Szostak JW. 2001. The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl. Acad. Sci. USA* 98:3750–55
83. Wilson DS, Szostak JW. 1999. In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* 68:611–47
84. Winter G, Milstein C. 1991. Man-made antibodies. *Nature* 349:293–99
85. Yoshikuni Y, Ferrin TE, Keasling JD. 2006. Designed divergent evolution of enzyme function. *Nature* 440:1078–82

86. Yuan L, Kurek I, English J, Keenan R. 2005. Laboratory-directed protein evolution. *Microbiol. Mol. Biol. Rev.* 69:373–92
87. Zaitseva J, Lu J, Olechoski KL, Lamb AL. 2006. Two crystal structures of the isochorismate pyruvate lyase from *Pseudomonas aeruginosa*. *J. Biol. Chem.* 281:33441–49
88. Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, et al. 2006. New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* 15:2785–94

# Contents

Annual Review of Biophysics
Volume 37, 2008

**Index**



**Errata**

An online log of corrections to *Annual Review of Biophysics* articles may be found at http://biophys.annualreviews.org/errata.shtml