3

Artificial Enzymes

James A. Stapleton, Agustina Rodriguez-Granillo, and Vikas Nanda

CONTENTS
3.1 Introduction .......................................................................................................................... 47
3.2 Directed Evolution ............................................................................................................... 49
3.3 Directed Evolution with Rational Library Design .............................................................. 52
3.4 Selection of Enzymes with No Natural Parent ..................................................................... 53
3.5 Catalytic Antibodies ........................................................................................................... 54
3.6 De Novo Design of Metalloenzymes ................................................................................... 55
3.7 De Novo Design of Enzymes without Metal Cofactors ....................................................... 57
3.8 Beyond Proteins—Artificial Ribozymes ............................................................................. 59
3.9 Other Nonprotein Artificial Enzymes .................................................................................. 62
3.10 Conclusions and Perspectives .......................................................................................... 64
References .................................................................................................................................... 65

3.1 Introduction

In its purest definition, nanotechnology describes the manipulation of matter at the atomic level to build systems from the bottom-up rather than the top-down (Drexler 1981). A materials synthesis strategy that pays attention to the placement of each atom has the potential to provide huge gains in specificity, efficiency, and affordability, and to process enormous quantities of matter through massive parallelization. The molecular machines from which biological systems are constructed do exactly this. People, blue whales, and redwood trees are not carved from larger blocks or injection molded but built from the ground-up, atom by atom.

In this chapter, we will discuss artificial enzymes, a fascinating and promising area of research that has seen significant progress in recent years. Enzymes are the subset of biological macromolecules that catalyze chemical reactions. Biological machines, which in addition to catalysis perform structural, regulatory, and other central roles in biology, are natural examples of “soft” nanotechnology from which we can learn much about design on the nanoscale. These machines self-assemble and function in water at environmental temperature and pressure. Proteins, the dominant class of catalytic biological molecules, are linear polymers of amino acids that fold (in a type of intramolecular self assembly) into precise but flexible three-dimensional structures determined by their amino acid sequence (Anfinsen et al. 1961). Water is the solvent for all life, and protein folding and function take extensive advantage of the range of hydrophobic/hydrophilic properties of the 20 natural amino acids. However, some proteins are stable and functional in organic environments (Klibanov 2001), indicating that water-insoluble or water-incompatible chemistry is not beyond their reach.
Enormous numbers of possible conformations are available to long polypeptide chains. The forces that govern folding are subtle, and the energetic difference between two competing folds can be minuscule. Predicting the three-dimensional structure that will be adopted by a given sequence is therefore extremely difficult and is one of the grand challenges of structural biology. Increasing computer power and clever algorithms have allowed researchers to make great progress in folding prediction in recent years, and the folds of most proteins of fewer than \( \sim 100 \) amino acids can now be predicted with reasonable accuracy (Zhang 2008). In considering artificial enzymes, we are perhaps more interested in the opposite problem: can we predict what sequences will adopt a given target three-dimensional structure? This is known as the protein design problem, and great progress has been made on this front as well. New sequences have been generated that adopt protein structures from nature (Dantas et al. 2003) as well as structures not yet found in nature (Kuhlman et al. 2003).

Enzymes increase by many orders of magnitude the rates of reactions that would otherwise proceed on glacial timescales. The mechanisms by which they achieve these speedups are often so delicate as to be nearly impossible to understand, much less replicate by design. Enzymes grab specific substrates and hold them in the perfect orientation for a reaction to occur. They might bring two reactants into just the right geometric proximity to allow a new bond to form or bend a molecule to strain a bond enough to encourage it to break. Side chains are preorganized in ideal positions to abstract protons or perform nucleophilic attack. “Second shell” side chains modify the properties of these “first shell” groups, tuning their pK\(_a\)s and enhancing the stability of the catalytically active state. The close fit between a binding cleft and its substrate, and the geometrical constraints placed on a bound substrate, confer incredible specificity on enzymes, which can distinguish between nearly identical potential reactants and cut unwanted side reactions to insignificant levels. Hydrophobic amino acids within a binding cleft can modify the local environment to enable chemistry usually impossible in water and reserved for organic solvents. In short, enzymes pick precise target molecules out of the complex cellular milieu and place them in the precise geometry and chemical environment that minimizes the activation energy of the specific desired reaction. When the reaction is complete, they release the products, ready for another cycle. Some enzymes do their work with such remarkable efficiency that the reactions they catalyze are limited by the rate at which reactant molecules can reach them by diffusion.

Nature provides us with countless examples of beautiful, complex nanomachines. The \( F_0 - F_1 \) ATP synthase (Figure 3.1a) transforms the energy contained in transmembrane proton gradients into high-energy chemical bonds. It harnesses the physical energy of the gradient via a mechanical rotary mechanism (Boyer 1997). Protons passing through a transmembrane channel in the \( F_0 \) region of ATP synthase ratchet the central stalk of the protein in 120° steps. The rotation of the asymmetric stalk (green) causes cyclic conformational changes in the three \( \alpha \) subunits (red) that, along with three \( \alpha \) subunits (blue), make up the \( F_1 \) region of the protein. In the “open” state, ATP from the previous cycle is released and new ADP and phosphate are bound. A 120° rotation of the stalk changes the conformation to the “loose” state, which binds the ADP and phosphate more tightly. A final turn induces the “tight” state, in which the ADP and phosphate combine to form ATP. The cycle continues with a transition back to the “open” state. Under certain conditions, the enzyme can also run in the reverse direction, consuming ATP to generate a proton gradient. Paul D. Boyer and John E. Walker shared half of the 1997 Nobel Prize in Chemistry for discovering the mechanism of ATP synthase.

The ribosome (Figure 3.1b), a molecular machine composed of protein and nucleic acid subunits, is another clear inspiration for nanotechnology. The ribosome reads the information encoded in mRNA and synthesizes polypeptides with specific amino acid
Artificial Enzymes

sequences. Amino acid building blocks are added one-by-one to the end of a growing polypeptide chain. Thanks to the ribosome, heritable genetic information is able to direct the creation of information-rich polymers that fold specifically and perform precise functions. Bulk synthetic chemical peptide synthesis methods fall far short of the efficiency and specificity (about 1 error per 10,000 amino acids [Ellis and Gallant 1982]) of the ribosome. The 1974 Nobel Prize in Medicine and the 2009 Nobel Prize in Chemistry were both awarded for studies of the ribosome.

Often, the activity or substrate preference of a natural enzyme can be modified drastically by mutating even a single active site amino acid by chemical intuition (reviewed by Toscano et al. [2007]). While enzymes with remarkably different properties can be generated by this type of rational active site redesign, the creation of artificial enzymes to perform any desired reaction under any process conditions will require more general approaches.

3.2 Directed Evolution

In the absence of a method to create designer enzymes from scratch, nature has been our only source of catalytic nanomachines. For centuries, domestication of biological organisms and their enzymes have provided humanity with a way to produce valuable products like beer and cheese. However, natural enzymes have evolved to meet the specific requirements of their host cells, and the demands of human industrial or health applications are often quite different than those of any natural environment. To modify natural enzymes to

FIGURE 3.1
(See companion CD for color figure.) Biological molecular machines. (a) The $\text{F}_1$ domain of the $\text{F}_0$–$\text{F}_1$ ATP synthase. Rotation of the central stalk (green) in 120° steps causes conformational changes in the $\beta$ domains (red), which cycle among three states with distinct affinities for ATP and ADP. (Produced from PDB 1E79) (b) The ribosome is a factory that reads information from mRNA and catalyzes the synthesis of polypeptide chains. It is composed of a large (blue) and a small (green) subunit that clamp together around an mRNA molecule (not shown). More than half of the ribosome, including the catalytic site, consists of RNA (orange). (Produced from PDBs 2WDK and 2WDL.)
function optimally within nonnatural environments or on nonnatural substrates, protein engineers have mimicked the evolutionary process by which these enzymes were created. Directed evolution is a laboratory-based method in which humans replace the natural selective pressures that determine the fitness of a protein with nonnatural, application-specific selective pressures. A diverse collection (called a “library”) of mutants based on a natural parent protein is generated, and each candidate is tested for the desired function. The candidates deemed the fittest under the conditions of the final application are replicated and mutated, providing mutants for a new round of directed evolution. This cycle is repeated until a suitable enzyme is found (Figure 3.2a).

Each directed evolution project requires a customized high-throughput selection or screen capable of identifying improved enzymes. In vivo selections and screens are convenient and popular but are limited by the efficiency of transforming mutant DNA into the host and by the tendency of cells to find alternative, undesired ways of avoiding the selection pressure. Extremely high throughput in vitro techniques such as ribosome display (Hanes and Pluckthun 1997) partially address these limitations, but have been restricted to selections for binding. The candidates deemed the fittest under the conditions of the final application are replicated and mutated, providing mutants for a new round of directed evolution. This cycle is repeated until a suitable enzyme is found (Figure 3.2a).

Recently, a phage-based technique has been developed that allows continuous directed evolution of proteins (Esvelt et al. 2011), as opposed to the discrete, labor-intensive rounds of traditional methods (Figure 3.2b). Phage-assisted continuous evolution (PACE) takes advantage of the rapid life cycle and high mutation rate of Escherichia coli-infecting bacteriophage. Fresh E. coli cells flow into a well-mixed volume termed a “lagoon,” where they remain for a residence time shorter than their division period before flowing to waste. Within the lagoon is a population of phage that is diluted by the continuous flow and must

---

**FIGURE 3.2**
(See companion CD for color figure.) Directed evolution and PACE schematics. (a) Traditional directed evolution consists of discrete rounds of sequential mutagenesis and selection. Mutagenesis of a parent gene (left) generates a library (top) with diverse mutations (red x’s). The mutant genes are transformed into bacteria (right), which transcribe and translate the genes into mutant proteins (colored circles). A screen or selection identifies the fittest mutants, the gene is isolated and amplified, and another cycle begins. (b) PACE automates the process, allowing dozens or hundreds of rounds to proceed without human intervention. Bacteria (tan ovals) flow into a “lagoon” (black-bordered rectangle) containing phage (blue rectangles). The phage infects the bacteria, which express the target gene contained within the phage genome (black ovals with red x’s indicating mutations), producing mutant protein (colored circles). Only mutant proteins with the desired activity induce expression of the pIII protein (blue stars) from a plasmid in the bacteria (green ovals). Newly produced phage without pIII are not infectious. A gene must produce copies of itself within infectious phage rapidly enough to avoid washing out of the lagoon.
infect fresh bacteria and replicate fast enough to avoid being washed out. The phage life cycle is on the order of 10 min, faster than that of the bacteria and shorter than the residence time within the lagoon. The phage are modified to contain a copy of the gene to be evolved, which is expressed upon infection of a bacterial host. The desired enzymatic activity is linked to the production of pIII, a phage protein required for infectivity. Therefore, phage harboring genes that encode active proteins replicate within cells to produce infectious phage. Genes that encode proteins lacking the desired activity also produce new phage, but these are unable to infect new host cells and harmlessly wash out of the lagoon. The high natural error rate of phage replication generates a diverse set of random mutations within the target gene and can be optionally enhanced by a mutagenesis plasmid. Esvelt and coworkers demonstrated the power of PACE by evolving new versions of the T7 RNA polymerase that recognize the T3 promoter and that initiate with ATP or CTP rather than GTP. PACE enabled up to 200 rounds of evolution to occur over 8 days with no human intervention. The researchers were able to follow the mutational paths taken in initially identical parallel runs: in the T3 promoter experiment, two lagoons accumulated different mutations before converging upon the same optimal set. While PACE is limited to the evolution of enzymatic activities that can be linked to pIII production, the benefits of continuous evolution will motivate imaginative protein engineers to think of clever ways to establish such a linkage for a wide variety of enzymes.

Directed evolution has the benefit of requiring no information about the structure or mechanism of the enzyme. The same “blind watchmaker” that built the parent enzyme adapts it to the demands of its new environment. However, because functional proteins are islands in the vastness of sequence space, it is unwise to stray too far from the parental sequence by introducing too many mutations at once. Directed evolution is, therefore, better suited for small optimizing tweaks than it is for introducing large changes like entirely new folds or functions. In a few cases, however, directed evolution has succeeded in generating enzymes with activities not present in the parent enzyme. One effective strategy has been to use multistep evolutionary paths in which one or more bridging substrates span the structural gap between a wild-type substrate and a desired substrate. For example, in the PACE study, the wild-type T7 RNA polymerase showed no activity with the T3 promoter, and selection for activity on that promoter did not support phage propagation. Selection for activity on a hybrid promoter consisting of the T3 promoter sequence with the T7 promoter base at the −11 position followed by selection on the full T3 promoter succeeded in generating the desired mutant. Similarly, the steroids testosterone and progesterone were used to bridge the structural gap between the natural substrate of human estrogen receptor α ligand-binding domain, 17β-estradiol, and the final target substrate corticosterone (Chen and Zhao 2005). However, while this strategy can be effective for altering the substrate specificity of a particular type of activity when clear structural intermediaries between the natural and target substrates exist, it may be impossible to extend it to the creation of enzymes that catalyze new reactions.

Another approach to generating new activity combined rational design and directed evolution to introduce β-lactamase activity into a glyoxalase II (GlyII; αβ/βα) metallo-hydrolase (Park et al. 2006). The C-terminal domain of the parent enzyme was removed to relieve steric constraints, and substrate- and metal cofactor-binding loops derived from a sequence alignment of metallo-β-lactamase (MBL) enzymes were inserted into the scaffold along with targeted mutations that introduced catalytic residues. This rationally designed scaffold served as the parent for random directed evolution by error-prone polymerase chain reaction (PCR) and DNA shuffling (Stemmer 1994), resulting in the isolation of a mutant that could support E. coli growth in the presence of a 1.0 μg/mL concentration of
the lactam antibiotic cefotaxime. The designed scaffold modifications in this study were inspired by MBL, which belongs to the same structural superfamily as GlyII and provided an example solution for the design of a β-lactamase. While the successful conversion of GlyII into a β-lactamase is remarkable, the application of this approach to the generation of novel artificial enzymes is limited by the requirement for a homologous natural enzyme.

3.3 Directed Evolution with Rational Library Design

The absence of any requirement for structural or functional knowledge is a strength of directed evolution. In addition to circumventing our ignorance of protein structure/function relationships, random mutagenesis and screening often finds beneficial mutations at positions far from the active site that could not have been predicted rationally. However, random approaches rely on the ability to screen large numbers of mutants, and finding a needle in the haystack requires considerable luck. In addition, while activities present at low levels in the starting protein can be improved, generating entirely new activity is very difficult. Researchers are developing a variety of strategies to address these limitations by using rational methods to enrich mutant libraries in active mutants.

A simple but extremely successful rational library design method has been the application of structural data to the selection of crossover points during recombination-based library generation. “Sexual” recombination of homologous genes takes advantage of sequence diversity already vetted by nature for compatibility with a particular fold and function. However, random recombination can introduce clashes and disrupt important contacts, lowering the fraction of library members that fold successfully. The SCHEMA (Voigt et al. 2002) and Recombination as a Shortest Path Problem (RASPP; Endelman et al. 2004) protocols analyze structural contacts to determine positions at which recombination is least likely to disrupt the structure. The resulting designed libraries are enriched in folded proteins and have been shown to outperform libraries generated by random DNA shuffling. These methods have been used to generate high-quality libraries of P450 heme proteins (Otey et al. 2006) and fungal cellulases (Heinzelman et al. 2009).

Multiple sequence alignments of homologous proteins can be used in library generation to provide information about which amino acids are allowed at each position, narrowing down the size of the sequence space to be searched. Bias from the evolutionary history of these sequences can be removed statistically (Halabi et al. 2009) or avoided entirely by selecting competent sequences from synthetic pools (Jäckel et al. 2010). Compatible diversity at each site is then built into synthetic degenerate oligonucleotides, which are assembled by PCR to yield a diverse collection of mutant genes. Designed libraries can also be constructed so as to preserve the correlations between amino acid identities at multiple positions. One such library (Lippow et al. 2010), which maintained the linkages between neighboring amino acid identities in a computationally redesigned active site, was shown to be enriched in active mutants relative to a control library with no interposition information.

In an example of how artificial enzymes could revolutionize the chemical industry, scientists at Codexis created an enzyme to replace a high-pressure, rhodium-catalyzed asymmetric hydrogenation step in the synthesis of sitagliptin, an antidiabetic pharmaceutical (Savile et al. 2010). A homology model (model of a protein with unknown 3D structure based on sequence similarity to a known protein) of a transaminase enzyme with no activity for the prositagliptin ketone indicated positions in the binding pocket that could
be targeted for mutagenesis. Site saturation mutagenesis and screening identified a variant with four mutations that had low activity toward the substrate. Additional rounds of directed evolution under industrially realistic conditions resulted in an activity improved by four orders of magnitude and an enantiomeric excess of >99.95%, along with the ability to function in the very unnatural environment of a bioreactor: 50% organic solvents (required to keep the substrates in solution), 40°C, and 250 mM substrate. Even under these harsh conditions, the enzyme remained stable for more than 24 h, demonstrating that protein as a nanotechnological substrate is by no means limited to natural environments.

Statistical methods and machine learning are increasingly applied to isolate the effects of individual mutations on stability and function and predict optimal combinations of mutations. Statistical approaches can be particularly valuable when no high-throughput screen is available, and are becoming more attractive as the economics of DNA sequencing and synthesis improve. By analogy to the quantitative structure–activity relationship method popular in drug development, an algorithm based on protein sequence–activity relationships (ProSARs) has been developed and applied to enzyme engineering (Fox et al. 2003). In ProSAR, mutants are screened and sequenced, and the effect of each individual mutation is resolved by partial least-squares regression. New mutations are added to the pool as beneficial mutations are identified and deleterious mutations are removed from consideration. ProSAR was used to efficiently adapt halohydrin dehalogenase mutants to the demands of an industrial process for the production of the starting material for the cholesterol drug, Lipitor (Fox et al. 2007). A similar DNA synthesis-based strategy involves synthesizing small collections of mutants containing combinations of promising mutations identified via analysis of alignments of homologous sequences and then teasing apart the contributions of each mutation to the performance of a small number of mutants. Testing a total of fewer than 100 custom-synthesized mutants over two rounds of library construction provided enough data to construct proteinase K variants with 20-fold improvements in thermostability (Liao et al. 2007). A similar approach identified five mutations to prolyl endopeptidase that increased the stability of the enzyme under gastric conditions, again requiring the synthesis and screening of fewer than 100 mutants (Ehren et al. 2008).

A particularly exciting new extension of these ideas is the use of next-generation sequencing technologies to generate fitness landscapes of the WW-domain protein (Fowler et al. 2010), an RNA ligase ribozyme (Pitt and Ferre-D’Amare 2010), and the chaperone Hsp90 (Hietpas et al. 2011). Deep sequencing was used to identify the sequences of those members of a mutant library deemed active by a functional selection. The abundance of each sequence in the selected pool was taken as a measure of the fitness of that sequence. The copious data that result from this method form a detailed sequence/activity fitness landscape that identify sites critical for folding and function and inform further rounds of library design.

### 3.4 Selection of Enzymes with No Natural Parent

An interesting new approach to library construction uses rational or computational design over the entire protein to restrict the library to sequences likely to fold into stable structures while preserving as much diversity as possible. This approach leaves the catalytic mechanism up to chance, merely restricting the set of possible amino acids at each site to those likely to be compatible with a desired three-dimensional structure. While this approach generates libraries too large to exhaustively screen, active proteins can be isolated when
the library design results in a high percentage of folded proteins or when an extremely high-throughput selection can be applied.

Binary patterning is a simple but surprisingly effective method for generating diverse collections of protein sequences that adopt a defined three-dimensional fold. Hydrophobic/hydrophilic interactions are dominant drivers of protein folding: the interiors of folded proteins typically contain hydrophobic amino acids that pack within the core to hide from the solvent. In contrast, surface amino acids tend to be polar or charged and interact favorably with water. A library designed to dictate only the hydrophobic or hydrophilic character at each amino acid position contained proteins of diverse primary sequence, most of which folded into the desired four-helix bundle structure (Kamtekar et al. 1993). Many of these proteins had ordered, native-like cores (Wei et al. 2003a,b). Screening of these binary-patterned libraries identified heme-binding proteins (Rojas et al. 1997) that bound carbon monoxide (Moffet et al. 2001) and proteins with peroxidase, esterase, and lipase enzymatic activity (Das and Hecht 2007; Patel et al. 2009).

The same group transformed 27 single-knockout auxotrophic E. coli strains with a library of $1.5 \times 10^9$ patterned helical bundles and isolated transformants that rescued growth of four of the strains on minimal media (Fisher et al. 2011). Activity could not be measured in cell lysates or purified samples of these proteins, but the very low levels of activity expected given the slow-growing phenotype would likely be below the detection limit of such assays. The authors go to admirable lengths to rule out alternate explanations, including demonstrating that mutation of key residues in the synthetic proteins abolished rescue of growth. The authors then showed that a quadruple knockout could be rescued by cotransformation with plasmids encoding four selected synthetic proteins. While the complexity of the cell prevents ruling out all alternative explanations, it seems likely that at least some of these simple, 102-residue helix bundle proteins, which were neither designed nor evolved but directly selected from a designed library, possess minimal enzymatic activity. This is especially remarkable considering the complexity of the enzymes they replace in this study. Given the tendencies of four-helix bundles, it is likely that the selected enzymes bind metals or other cofactors upon which they rely for function. Targeted mutagenesis to introduce cavities into a selected bundle resulted in a pocket capable of binding small aromatic molecules (Das et al. 2011). Scaffolds with designed cavities of this type may allow the selection of enzymes with substrate binding clefts similar to those observed in natural enzymes.

An alternative strategy is to select from a less-restricted library with an extremely high-throughput method. In vitro methods are capable of selecting desired proteins from libraries of $10^{12}$ or more mutants. One of these, mRNA display, has been used to select ATP-binding proteins from random sequences of 80 amino acids (Keefe and Szostak 2001) and, more recently, to select enzymes capable of ligating two RNA molecules (Seelig and Szostak 2007). In the latter study, the starting library consisted of a zinc-finger protein scaffold with two randomized loops of 12 and 9 amino acids. Selection from this random library was followed by mutagenesis and recombination, finally yielding zinc-dependent enzymes that accelerate the reaction by as much as $2 \times 10^6$-fold.

### 3.5 Catalytic Antibodies

Modern transition state theory and computer simulations suggest that the dominant factor in enzymatic catalysis is the lowering of the activation barrier by stabilization of the
Artificial Enzymes

transition state (Benkovic and Hammes-Schiffer 2003; Garcia-Viloca et al. 2004). Therefore, molecules that specifically bind and stabilize a transition state analog (TSA) of a desired reaction are good candidates to act as enzymes for that reaction. This strategy forms the basis for catalytic antibodies: Designing and synthesizing a stable TSA of the desired reaction and then using this molecule as an antigen in immunizations of animals raises antibodies capable of stabilizing the transition state of the desired reaction. Between approximately $10^8$ and $10^{11}$ different specificities are present in a human antibody repertoire (Hanson et al. 2005). These varied libraries should contain a catalytic antibody for virtually any chemical reaction. Catalytic antibody generation is a knowledge-driven method (Golynskiy and Seelig 2010), since construction of a “good” TSA requires a detailed understanding of the reaction mechanism.

The first catalytic antibodies, or “abzymes,” were generated 25 years ago to catalyze the hydrolysis of esters and carbonates. TSAs were used as haptens to produce monoclonal antibodies with the ability to enhance the rate of the reactions $\sim 10^3$-fold (Pollack et al. 1986; Tramontano et al. 1986). Since then, artificial antibodies have been generated that catalyze a plethora of chemical transformations, including hydrolysis of amides and esters, cyclization, decarboxylation, lactonization, peroxidation, and reactions for which no natural or artificial enzyme exists (Nevinsky et al. 2002). However, even the most tailored abzymes cannot outperform highly evolved natural enzymes. Abzymes achieve maximum rate enhancements of $2.3 \times 10^8$ s$^{-1}$ versus $7 \times 10^{19}$ s$^{-1}$ for natural enzymes (Golynskiy and Seelig 2010).

There are several drawbacks in the design of abzymes that could explain their lower catalytic efficiency. Abzymes are specifically designed to bind to the TSA, which could result in enzymes that bind too tightly to the transition state, blocking catalysis or product release (Golynskiy and Seelig 2010). Also, transition state stabilization is only one of many strategies natural enzymes use to accelerate reactions. Abzymes are restricted to the single immunoglobulin fold (Golynskiy and Seelig 2010), and therefore have low flexibility and plasticity (Belogurov et al. 2009) and solvent-exposed active sites (Xu et al. 2004). Abzymes have an advantage, however, as in vivo therapeutic agents, since antibodies are less likely than other artificial enzymes to elicit an immune response in the body (Golynskiy and Seelig 2010).

3.6 De Novo Design of Metalloenzymes

If protein active sites were constrained to use only the chemical groups offered by the amino acids, they would not be able to catalyze all the reactions required to sustain life (Bertini et al. 2007). As much as 30% of all natural enzymes are thought to incorporate metal cofactors (Ragsdale 2006). The chemical diversity of the inorganic elements allows these so-called metalloenzymes to perform a wider range of biochemical functions by facilitating reactions such as bond forming and breaking, electron transfer, and radical chemistry (Ragsdale 2006).

The design of artificial metalloenzymes is challenging. In addition to providing a stable protein scaffold, the design must place the metal cofactor in the correct geometry and within the proper environment to obtain the desired function. As opposed to traditional metal catalysts in which the first coordinating shell (the chelating ligands) is the major component that defines the catalytic activity and specificity, in enzymatic
catalysis, second-shell and even more distant interactions can be as important as first-shell interactions (Rosati and Roelfes 2010). Artificial metalloenzymes can be designed de novo or by creating functional active sites within an existing protein scaffold. Although the latter approach has been more widely used because of the greater choices of protein scaffolds, we will focus on the de novo design of artificial metalloenzymes because with the recent advances in computational and structural biology this field has lately seen significant success (Lu et al. 2009).

The first artificial metalloenzyme was designed without the aid of computers, and consisted of four amphiphatic α-helices bound to a heme group (Sasaki and Kaiser 1989). This “heliochrome” had aniline hydroxylase activity similar to that of natural heme proteins. Since then, most of the de novo metalloenzymes that have been synthesized are based on heme groups bound to α-helical bundles (Lu et al. 2009). As discussed earlier in Section 3.4, Hecht and coworkers used combinatorial libraries of binary-patterned sequences with a periodicity of polar and hydrophobic residues similar to that of the heliochrome to de novo design a number of four-helix bundles that bind heme and catalyze peroxidase chemistry (Das and Hecht 2007; Kamtekar et al. 1993; Rojas et al. 1997; Wei et al. 2003a,b). Heme oxygenase activity was also engineered into de novo designed four-helix bundles (Monien et al. 2007).

Computational methods were applied in the de novo design a series of di-iron proteins called the due ferri (DF) proteins, inspired by natural dimetal proteins (Di Costanzo et al. 2001; Lombardi et al. 2000; Maglio et al. 2003; Marsh and DeGrado 2002; Summa et al. 2002). These proteins are four-helix bundles that bind a dinuclear iron cluster; the original protein was termed DF1 and consisted of two helix-loop-helix motifs (Lombardi et al. 2000; Figure 3.3). By rationally modifying the active site of one of these proteins,
Artificial Enzymes

57

DFtet, to accommodate the substrate, they were able to engineer a variant capable of catalyzing the oxidation of 4-aminophenol in the presence of atmospheric oxygen with a $10^3$-fold rate enhancement (Kaplan and DeGrado 2004). However, this redesign, which included the mutation of an Ala and a Leu to two Gly residues in two of the four chains, significantly destabilized the protein. More recently, the interhelical turn of DF1 was modified with the goal of overcoming the conformational destabilization introduced by the Gly mutations (Faiella et al. 2009). This new design, which also incorporated Gly mutations in the active site pocket, was termed DF3 and was able to catalyze the oxidation of 4-aminophenol and 3,5-ditert-butyl-catechol. The new design also exhibited improved thermodynamic stability with respect to previous variants and remained active for at least 50 cycles (Faiella et al. 2009).

3.7 De Novo Design of Enzymes without Metal Cofactors

While pragmatic evolution-based methods have been very successful in solving limited classes of problems, the ability to design a sequence that will fold and function as predicted remains the Holy Grail. This goal provides the ultimate test of our understanding of protein folding and enzymatic mechanism, and of our ability to control physics and chemistry at the nanoscale.

Thanks to the continued increases in computer speed, computation has emerged as a promising method for taming the complexity of enzyme design. Thus far, computational studies have focused on designing new active sites for incorporation into existing protein scaffolds. Over a decade ago, an active site was designed into a thioredoxin scaffold to create a new enzyme that hydrolyzed p-nitrophenyl acetate (PNPA; Bolon and Mayo 2001). Using a strategy similar to that underlying catalytic antibody generation, the active site was designed to stabilize a high-energy intermediate along the reaction pathway. A designed enzyme containing only three mutations relative to the wild type was able to catalyze PNPA hydrolysis with kinetics similar to those of early abzymes.

Computational redesign of a substrate-contacting loop changed the specificity of human guanine deaminase $2.5 \times 10^6$-fold in favor of a target substrate, ammelide (Murphy et al. 2009). The new loop, which at four residues in length was two residues shorter than the original loop, placed an asparagine residue in position to form hydrogen bonds with a docked ammelide molecule. A crystal structure of the designed enzyme revealed that the configuration of the designed loop matched the design to a alpha-carbon root mean square deviation (Cα-RMSD) of 1 Å. Point mutants confirmed that correct placement of the designed asparagine residue was important for activity. However, the activity of the designed enzyme with ammelide was seven orders of magnitude lower than that of the wild-type enzyme with guanine.

In a series of recent breakthrough reports, David Baker and his colleagues have described a novel protocol for the creation of artificial enzymes. In contrast to approaches in which the scaffold is chosen first, the first step of the Baker protocol (Figure 3.4) is to design and computationally model disembodied idealized active sites for the target reaction. Like abzymes, these theoretical enzymes, or “theozymes,” are designed to stabilize the transition state of the reaction. Quantum-mechanical calculations are then used to identify the most promising designs. Once a large collection of candidate theozymes has been generated, the RosettaMatch program attempts to graft each of these constellations
FIGURE 3.4
(See companion CD for color figure.) Overview of de novo computational enzyme design protocol for the Rosetta enzymes. The first step (1) is to choose a reaction that the new enzyme will catalyze, and identify the transition state(s) and key intermediate(s) of the reaction pathway. Possible functional groups that might stabilize the transition state are identified by chemical intuition. QM calculations (2) are used to guide and optimize the positioning of different side chains and functional groups around the transition state, generating different possible theozymes. Next, an ensemble of active sites (3) is created by varying the side chain rotamers, and these active sites are matched to complementary protein scaffolds (4). The resulting promising models are identified (5) and tested experimentally (6). The coordinates for the theozymes and the enzyme models were kindly provided by David Baker.
of amino acids onto each of a set of a few hundred scaffold protein structures taken from the Protein Data Bank. This matching step is extremely computationally intensive and is made possible by a distributed network of volunteers who provide access to their personal computers through a project called Rosetta@home (Das et al. 2007). The set of theozyme/scaffold matches is narrowed down through further computational and intuitive filtering. Finally, around 100 candidates are synthesized and tested for activity.

The first enzymes designed using this method were retro-aldolases that break a bond in a substrate not found in nature (Jiang et al. 2008) and an enzyme that performs the Kemp elimination, a reaction for which no natural enzyme is known (Rothlisberger et al. 2008). Both of these reactions break bonds, and therefore bind only one substrate molecule. An enzyme that catalyzes the Diels–Alder reaction, which combines two substrate molecules into one product, followed shortly (Siegel et al. 2010). The Rosetta enzyme design procedure has been described in detail and is available for use in any laboratory (Richter et al. 2011).

Eventually, it is hoped enzyme design will incorporate artificial active sites into de novo scaffolds designed to be optimal for the active site chemistry and the application environment. In an early example, a binary-patterned helix served as a designed scaffold for an artificial enzyme. The 32-residue peptide catalyzed its own replication by the ligation of two 16-residue peptides corresponding to its own N- and C-terminal halves (Lee et al. 1996; Saghatelian et al. 2001). The peptide was designed to fold into a helix presenting a binding face. When bound to the catalyst, the chemically activated termini of the substrate peptides were perfectly oriented for ligation, reducing the entropic cost of the reaction. Functional group preorganization, in which the folding energy of the protein scaffold balances the cost of fixing the substrate and key active site residues in orientations optimal for the desired reaction, is one of the most important catalytic strategies used by natural enzymes, and is promising target for mimicry by artificial enzymes.

### 3.8 Beyond Proteins—Artificial Ribozymes

So far, we have reviewed the major class of artificial enzymes: those composed of protein. However, another important class of biological enzymes is not protein-based but is instead based on nucleic acids. These are termed ribozymes or deoxyribozymes, depending on whether they are composed of RNA or DNA. Here, we will focus only on artificial ribozymes.

In the early 1980s, Cech and Altman independently discovered that RNA molecules not only carried genetic information but also had catalytic properties (Guerrier-Takada et al. 1983; Kruger et al. 1982). This breakthrough discovery was recognized with the Nobel Prize in chemistry in 1989. Since then, there have been increasing efforts to create artificial ribozymes with novel catalytic properties.

The first artificial ribozyme was published in 1990 (Robertson and Joyce 1990) and consisted of an RNA molecule evolved and selected in vitro to specifically cleave single-stranded DNA (as opposed to the parental ribozyme that cleaved RNA substrates). The general methodology to create artificial ribozymes (Ellington and Szostak 1990; Tuerk and Gold 1990) consists of synthesis of a DNA molecule with constant and random regions, amplification by PCR, and in vitro transcription to produce the initial random RNA pool that will be used for the first round of selection during the artificial evolution experiment. Unlike proteins, nucleic acid-based enzymes can be directly amplified and sequenced,
greatly simplifying the selection protocol. After a number of selection cycles, selected ribozymes are identified and tested for the desired catalytic activity.

The reactions catalyzed by artificial ribozymes are many and range from RNA processing, such as RNA cleavage, ligation, branching, phosphorylation, and capping, to peptide bond formation, alcohol oxidation, and the Diels–Alder reaction (Silverman 2009). The creation of artificial ribozymes that are able to form a 5′ to 3′ phosphodiester bond (Bartel and Szostak 1993; Ekland et al. 1995; Ikawa et al. 2004; Jaeger et al. 1999; Landweber and Pokrovskaya 1999; Robertson and Ellington 1999; Rogers and Joyce 1999) and true RNA-dependent RNA-polymerases that are able to polymerize a complete turn of an RNA helix (Johnston et al. 2001; Lawrence and Bartel 2003, 2005), together with the crystal structure of an RNA ligase (Robertson and Scott 2007), provided proof that although no known natural ribozyme can catalyze the polymerization of RNA, RNA can indeed catalyze a key step required for its own replication, supporting the “RNA world” theory (Joyce 2007). In addition, in vitro evolution was used to create a cross-catalytic system in which two RNA ligase ribozymes catalyze each other’s synthesis (Lincoln and Joyce 2009).

In nature, there is no RNA enzyme capable of catalyzing the aminoacylation of the 3′ terminus of tRNA; this job is instead done by a family of protein enzymes called aminoacyl-tRNA synthetases. In a series of studies (Bessho et al. 2002; Goto et al. 2008a,b; Kawakami et al. 2008a,b; Lee et al. 2000; Murakami et al. 2003, 2006; Niwa et al. 2009; Ohta et al. 2007; Saito et al. 2001), Suga and coworkers used repeated cycles of in vitro evolution experiments to create artificial aminoacyl-tRNA synthetase-like ribozymes, termed “flexizymes,” that are able to synthesize a wide array of acyl-tRNAs charged with artificial amino acids and hydroxy acids. Flexizymes not only support the existence of a primitive translation catalytic system consisting of RNA molecules only but also provide an artificial platform to express nonstandard peptides containing both proteinogenic and nonproteinogenic amino acids for therapeutic applications (Goto et al. 2011; Morimoto et al. 2011).

RNA molecules, like proteins, adopt defined three-dimensional structures that delineate binding sites and catalytic centers. Guided by molecular modeling, Ikawa et al. used known structural motifs, or modules that formed the reaction site to construct an RNA scaffold into which a random region was inserted (Ikawa et al. 2004). With this approach, they were able to create an artificial RNA ligase that accelerated the reaction 10⁶-fold over the uncatalyzed reaction. The artificial ribozyme exhibited a higher product yield than previously reported RNA ligases, suggesting that most of the RNA machine was properly folded in a catalytically active way.

The traditional approach to ribozyme development is to use an in vitro selection method, in which repeated rounds of selection are preformed and evolution occurs in a stepwise manner. However, analogous to the PACE system discussed in Section 3.2 for protein-directed evolution, an alternative is to use continuous evolution, which can occur hundreds of times more quickly (Wright and Joyce 1997). Joyce and coworkers were the first to continuously evolve catalytic RNA, and they were able to apply it to two types of RNA ligases (Voytek and Joyce 2007; Wright and Joyce 1997). Despite the efficiency and power of this method, it is extremely limited in the type of reactions that may be catalyzed since it requires a ribozyme with a sufficiently fast reaction rate (Voytek and Joyce 2007).

In the traditional in vitro selection process, there is no room to select for advanced enzymatic properties such as multiple turnover, since selection is generally based on formation of a covalent linkage between the RNA and a substrate labeled with a “capture” tag (Silverman 2009). One way of getting around this limitation is to use IVC strategies (Tawfik and Griffiths 1998), in which the sequence of the ribozyme (genotype) and its catalytic activity (phenotype) become “linked” within individual droplets in a water-in-oil
emulsion. Recently, a novel selection approach based on this strategy was developed to engineer an RNA polymerase capable of synthesizing RNAs of up to 95 nucleotides (Wochner et al. 2011). The method is termed compartmentalized bead-tagging, and consists of encapsulating a genetic library of ribozymes attached to magnetic beads in individual droplets, allowing transcription to occur within the droplets to create the ribozymes, triggering primer extension by addition of primer/template duplexes in a second emulsion, and detecting the extent of primer extension by a combination of rolling circle amplification of the extended primers and fluorescence-activated cell sorting (FACS; Figure 3.5). This method, in combination with rational RNA engineering, yielded an RNA polymerase with greater polymerase activity, fidelity, and generality than the parental ribozyme. The new ribozyme was able to synthesize an enzymatically active ribozyme from an RNA template (Wochner et al. 2011), and can polymerize sequences half of its own length, bringing us closer to the goal of a completely self-replicating ribozyme.

**FIGURE 3.5**
(See companion CD for color figure.) Compartmentalized bead-tagging method for the selection of artificial ribozymes. (1) Hairpin oligonucleotides (green) and biotinylated genes (red) from a library are attached to streptavidin-coated magnetic beads (blue circle). (2) Transcription takes place within a first water-in-oil emulsion, producing ribozymes that subsequently ligate to the hairpin. (3) The emulsion is broken and primer (black) and template (cyan) duplexes are attached to the magnetic beads. (4) In a second emulsion, ribozymes are released from the beads and primer extension can proceed. (5) Primer extension amplification is achieved by rolling circle amplification of a DNA minicircle. (6) To facilitate signal detection, fluorescent-labeled probes are hybridized to the DNA. (7) The active ribozymes are isolated by FACS and amplified by PCR amplification. (Modified from Wochner, A. et al., Science, 332, 209, 2011.)
The rate enhancements achieved with artificial ribozymes are comparable with the ones observed with abzymes, with maximum values of up to $10^{-10} \text{ s}^{-1}$ (Suga et al. 1998). However, although artificial ribozymes cannot compete with protein enzymes, they are able to outperform their natural counterparts (Silverman 2009).

### 3.9 Other Nonprotein Artificial Enzymes

Although nature has chosen to build its nanomachines out of protein and nucleic acids, other types of polymers are also capable of protein-like three-dimensional folds and functions. Alternative backbones and side chains (Figure 3.6) may prove to have advantages over proteins in future applications. Research on these types of biomimetic folding polymers, known as foldamers, is progressing rapidly.

Nature has elected to use only $\text{l}$-amino acids at the expense of their mirror-image enantiomers, the $\text{d}$-amino acids. In principle, for each natural protein, a corresponding sequence of $\text{d}$-amino acids will fold into a mirror-image protein with activity against

![Comparison of the chemical backbones of various foldamers.](image)

**FIGURE 3.6**

Comparison of the chemical backbones of various foldamers. (a) $\text{l}$-peptide, (b) $\text{d}$-peptide, (c) peptoid, (d) $\beta^2$-peptide, and (e) $\beta^3$-peptide.
mirror-image substrates. D-peptides are resistant to degradation by proteases, making them promising drug candidates. D-amino acids have been computationally modeled (Nanda and DeGrado 2006) and incorporated into t-peptides to improve stability (Rodriguez-Granillo et al. 2011). Full proteins composed entirely of D-amino acids could be very useful for the synthesis of enantiomers and diastereomers (Forster and Church 2007), or as a safety mechanism to prevent escaped synthetic biological systems from interfering with natural life.

Natural proteins are composed of 20 canonical amino acids, which among them cover a considerable amount of chemical space. However, artificial proteins that include noncanonical amino acids could potentially perform an even wider range of functions. Extensive work has been done to enable the global or site-specific incorporation of noncanonical amino acids into proteins and has been extensively reviewed (e.g., Antonczak et al. 2011). One particularly interesting project (Neumann et al. 2010) used a modified ribosome that recognized nucleotide quadruplet codons to synthesize artificial proteins containing noncanonical amino acids. Unlike earlier strategies based on suppression of the amber stop codon, this method can allow the incorporation of multiple noncanonical amino acids into the same protein.

Peptoids, or N-substituted glycines, are achiral, protease-resistant peptide analogs. A key advantage of peptoids is that they can be polymerized by a convenient and economical submonomer synthetic method (Burkoth et al. 2003) that allows incorporation of any of thousands of commercially available primary amines. An engineered ribosomal system (Kawakami et al. 2008b) has also been developed that could enable selections for enzymatic peptoids via isolation and amplification of encoding mRNA. Programming of polymer function by sequence-level design has been demonstrated with designed peptoids. In one example, a two-helix bundle was designed that selectively bound zinc with nanomolar affinity (Lee et al. 2008). Thiol and imidazole side chains (inspired by the cysteines and histidines used to bind zinc in proteins) were positioned to bind zinc only of the peptoid assumed the target structure. In another study, two peptoid polymers were reported to form two-dimensional crystalline sheets in aqueous solution (Nam et al. 2010). When mixed at a 1:1 ratio, the two 36-mers spontaneously formed a 2.7 nm-thick bilayer. Assembly was driven by the burial of hydrophobic side chains and pairing of positively and negatively charged side chains and did not require a phase interface as a template. Fusing an achiral small molecule catalyst to a structured peptoid resulted in enantioselective catalysis, which depended on the handedness of the peptoid used (Maayan et al. 2009).

Another promising class of foldamers is the β-peptides. As opposed to standard α amino acids, in which the amino group is bonded to the α carbon, in β amino acids, the amino group is bonded to the β carbon. The side chain can branch off of the α carbon (β2-peptides) or the β carbon (β3-peptides). β3-peptide foldamers have been shown to adopt helical secondary structure (Appella et al. 1997) and cooperative noncovalent quaternary structure (Qiu et al. 2006) but are more difficult than peptoids to synthesize.

Both β-peptides (Porter et al. 2002) and peptoids (Chongsiriwatana et al. 2008) can mimic the fold and function of antimicrobial peptides, with greatly reduced susceptibility to in vivo degradation. Despite the lack of a natural structural knowledge base against which to fit parameters, computational methods developed for protein structure prediction and design are being extended for use with β-peptides (Shandler et al. 2010) and peptoids (Butterfoss et al. 2009) to facilitate future designs. Computational design was recently used to build a self-assembling β-peptide hexameric bundle (Korendovych et al. 2010).

Enzyme-like catalysts are also emerging from the field of supramolecular chemistry, which adopts the noncovalent interaction-based specificity strategy that is so successful
for biological catalysts. Recently, a catalyst reminiscent of DNA polymerase was shown to effectively polymerize δ-valerolactone (Takashima et al. 2011). The catalyst consisted of a cyclodextrin heterodimer. One cyclodextrin ring contained the catalytic active site and the other acted as a molecular clamp that secured the growing chain. The catalyst was only active when the polymer chain was threaded through the center of the clamp.

These studies show that the gap between protein science and nanoscale chemistry is closing rapidly. As this gap narrows, we will no longer be dependent on nature’s examples of nanobiotechnological artificial enzymes, and we will be free to abandon protein and nucleic acids in favor of whatever chemical scaffold is best suited for the task at hand.

3.10 Conclusions and Perspectives

Enzymes are replacing traditional catalysts and playing increasing roles in a wide variety of industrial and medicinal applications, both in vivo (Keasling 2010) and in vitro (Zhang et al. 2011). The ability to create artificial enzymes to perform any desired chemical transformation would revolutionize the chemical and health industries. Through the process of evolution, nature has found remarkably clever and efficient ways of solving complex problems. Biomimicry has been a valuable strategy in engineering, product design, and architecture. In nanotechnology as well, close study of natural examples of functional nanomachines will be instrumental to our understanding and progress. The knowledge gained from the study of natural and artificial proteins will directly apply to the development of biologically inspired nonprotein nanomachines.

Artificial enzyme design is progressing rapidly, but daunting problems remain to be worked out. Limitations in computational algorithms and processor power are slowly relaxing, but we remain constrained by our poor understanding of the subtle mechanisms by which proteins do their work. It is interesting that the computational strategy of designing theozymes to stabilize the transition state has yielded enzymes that have similar activity to catalytic antibodies, which are raised to bind TSAs. It may turn out that “soft” bionanomachines like proteins are too floppy to perform effective catalysis by transition state stabilization alone. Proteins may have taken advantage of their softness by evolving subtle catalytic mechanisms that harness structural motion and transitions within the protein to accelerate reactions (Hammes-Schiffer and Benkovic 2006; Henzler-Wildman et al. 2007a,b). Indeed, the “lock-and-key” model of enzyme catalysis has fallen out of style in favor of the “induced fit” model, in which substrate binding changes the protein into an active conformation. Despite rapid improvements in computational power, such subtle mechanisms will remain extremely difficult to design for the foreseeable future. Indeed, efficient enzymes may prove to be so complex and finely tuned as to be chaotic and fundamentally undesignable.

The success of the protein design, evolution, and selection studies highlighted in this review provides encouragement for the future. It seems possible that while sequences that encode highly efficient enzymes are extremely rare, and to this point have only been discovered by millennia of evolution, sequences that encode poor catalysts are not especially rare. These sequences can be selected from relatively small constrained libraries or designed computationally. Poor catalysts, while themselves not useful, can serve as starting points for optimization by directed evolution, which generally requires a starting point with some amount of the desired activity. If this activity has not been found in nature,
Artificial Enzymes

synthetic enzymes with minimal activity could be very valuable. The question is whether evolutionary pathways through sequence space exist connecting these relatively abundant poor catalysts to the extremely rare sequence that encodes a highly optimized enzyme with efficiency approaching those of natural proteins. Comparison of the three-dimensional structures of a 102-residue helical bundle and the citrate synthase it functionally replaced (Fisher et al. 2011) would suggest that a tremendous amount of evolution would be needed to transform the former into the latter. However, the incredible complexity developed (or perhaps accumulated) by nature may not be necessary for human-defined tasks. The elimination of unnecessary complexity is a central tenet of synthetic biology (Endy 2005), which seeks to transform biology into an engineering discipline. Future work will determine whether subtleties such as protein motion are required for efficient catalysis, or whether they represent evolutionarily acquired but ultimately unnecessary baggage that can be eliminated by de novo design.

References


Artificial Enzymes


