Published in final edited form as:
Curr Opin Chem Biol. 2011 April ; 15(2): 201–210. doi:10.1016/j.cbpa.2010.11.020.

Optimizing Non-natural Protein Function with Directed Evolution

Eric M Brustad[1] and **Frances H Arnold**[1,2]

[1]Division of Chemistry & Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

[2]Department of Bioengineering, California Institute of Technology, Pasadena, CA 91125, USA

Abstract

Developing technologies such as unnatural amino acid mutagenesis, non-natural cofactor engineering, and computational design are generating proteins with novel functions; these proteins, however, often do not reach performance targets and would benefit from further optimization. Evolutionary methods can complement these approaches: recent work combining unnatural amino acid mutagenesis and phage selection has created useful proteins of novel composition. And weak initial activity in a computationally designed enzyme has been improved by iterative rounds of mutagenesis and screening. A marriage of ingenuity and evolution will expand the scope of protein function well beyond Mother Nature's designs.

Introduction

Protein engineering has emerged as a powerful vehicle for the development of important protein-based tools, including biocatalysts, biosensors, and therapeutics. The ability to tailor proteins for a specific purpose is the ultimate test of our understanding of their structure and function; it is unfortunately a test that is all too often failed. The complex and subtle interplay of interactions that dictates fold and function presents a daunting obstacle to rational protein design. Evolution is Nature's solution to the design problem. Scientists have learned how to implement evolutionary strategies to engineer new proteins, exploiting natural protein scaffolds as starting points for breeding improved versions. A not-surprising testament to the power of natural selection is that the most reliable approach to optimizing protein function is by iterative rounds of mutagenesis and screening or selection, i.e. directed evolution [1-3].

Protein engineers are rapidly moving beyond merely improving on natural proteins, which carry out their biological functions with breath-taking efficiency but may not be suitable for the myriad other applications on the protein chemist's wish list. A general lesson that has come out of laboratory evolution efforts is that it is easier to improve an existing, albeit low, activity than it is to discover a new one. There is evidence that at least one mechanism by which a novel enzyme evolves is from an existing 'promiscuous' activity, which is captured and optimized (often after gene duplication) by natural selection [4]. Laboratory evolution

Corresponding author: Arnold, Frances H (frances@cheme.caltech.edu).

Conflicts of Interest

The authors are aware of no conflicts of interest regarding the preparation and submission of this manuscript.

experiments show that such side activities in fact can often be increased and optimized. Generating a function where it does not already exist, not even at a low level, however, is a bigger challenge that will benefit from the use of new technologies that improve the design process or that provide new chemical functionalities not present in natural systems.

The past decade has demonstrated that synthetic and computational tools can be exploited to generate proteins with novel structural and chemical properties. Protein sequence space is no longer limited to the functionality presented by the canonical genetic code, as a wide variety of unnatural amino acids (UAAs) can now be incorporated into proteins *in vitro* and *in vivo* [5]. In addition, considerable work has been applied to engineer artificial metallocenters and non-natural cofactors into proteins [6]. With this rapidly expanding chemical diversity available for protein engineering, the functional space may be limited ultimately by the imagination of the protein scientist. While these strategies provide the ability to add new chemical functionality, overcoming certain limitations will be key to their wide use in protein engineering. Unnatural amino acid technologies, for example, often suffer from decreased protein yields due in part to decreased protein stability or inefficient translation, and the introduction of new transition metals is often limited by cross-reactivity with other components of the cellular milieu, necessitating extensive protein purification before the metal can be introduced [6].

Computational approaches to protein design are also building on our growing understanding of protein structure and function. Entire protein folds and proteins that catalyze reactions not present in Nature have been designed *in silico* and constructed [7,8•,9,10••]. Nonetheless, while considerable achievements have been made, the results are less impressive when they are compared to the products of natural evolution. For example, the abilities of naturally evolved enzymes can be appreciated by comparing the rate constant for the enzyme-catalyzed reaction to that of the spontaneous uncatalyzed reaction (k_{cat}/k_{uncat}). Enzymes demonstrate rate enhancements as high as 10^{19}, and values of 10^{12} are not uncommon [11]. The best computationally designed enzymes, however, exhibit enhancements ($<10^{6}$) that are many orders of magnitude lower.

Directed evolution complements these methods by (**1**) improving new protein designs through mutation and selection, with the ultimate goal of providing molecules with useful synthetic and therapeutic properties and by (**2**) demonstrating the limitations and drawbacks of these methods, stimulating improvements into next generation protein designs. Here we will highlight selected work demonstrating how these advances in biotechnology are providing new opportunities for the directed evolution of non-natural protein functions.

Directed Evolution with Unnatural Amino Acids

The past twenty years have seen remarkable advances in methods for incorporating unnatural amino acids (UAAs) into proteins. A variety of techniques have been developed, including native protein ligation, global amino acid replacement, *in vitro* translation or direct cellular injection of chemically aminoacylated tRNAs, and the generation of orthogonal tRNA/aminoacyl-tRNA synthetase pairs for the site-specific incorporation of amino acids *in vivo*. The foundations of this research have been thoroughly detailed elsewhere; we point the reader to several excellent recent reviews [5,12,13].

For the most part, experiments involving the introduction of UAAs into proteins have focused on new chemical diversity that yields a specific function, with little effort placed on further tailoring the protein for that function. Nevertheless, the novel chemical functionality has already been used in interesting applications. Some recent highlights include work by Grunewald et al., who demonstrated that incorporation of an immunogenic *p*-nitrophenylalanine residue into a protein could overcome self-tolerance in autologous host

proteins, providing a new tool for immunotherapy [14•]. Injection of mice with a *p*-nitrophenylalanine-modified mouse tumor necrosis factor (*m*TNF) led to the formation of cross-reactive antibodies that recognized both the modified and wild type *m*TNF. Mills et al. have shown that a genetically-encoded fluorescent coumarin amino acid can be introduced into the antigen-binding region of a Fab specific for CD40 ligand [15]. Changes in fluorescence can be used to monitor ligand binding, perhaps providing a general means for genetically encoding antibody-based biosensors. Tirrell, Schuman and colleagues have demonstrated that 'click' reactive azidohomoalanine and homopropargylglycine amino acids can be introduced globally into proteins in living cells by exploiting the promiscuity of host methionyl-tRNA synthetases. By combining pulse chase amino acid incorporation and labeling with various 'click' reactive fluorophores, they were able to image newly synthesized proteins at various time points in rat hippocampal neurons [16•]. In addition, with a modified methionyl tRNA synthetase specific for longer 'click' amino acid homologues, they were able to achieve cell-specific protein labeling in a mixed cell population [17•].

Although introduction of UAAs provides new chemical functionality, it is also disruptive: proteins containing UAAs often exhibit reduced native activity and lower stability. As technologies for *in vivo* UAA incorporation advance, we can consider using directed evolution to improve proteins containing UAAs. The first example of laboratory evolution with UAAs was work performed by Wong [18] and later by Bacher and Ellington [19]. Using chemical mutagenesis or spontaneous mutation in cell culture, respectively, they demonstrated that bacterial cells grown in the presence of a cytotoxic fluorinated tryptophan homologue would, over generations, evolve tolerance to the new amino acid. As shown by Bacher et al. [19], these cells were able to grow, albeit more slowly, with the fluorinated homologue as their sole tryptophan source, suggesting almost complete proteomic replacement of tryptophan. Overexpression of a model protein, glutathione S-transferase (GST), in the evolved bacterial strain in the presence of the modified tryptophan yielded only the fluorinated protein, as determined by mass spectrometry.

Tirrell and colleagues have performed similar experiments looking at global replacement of an amino acid in individual proteins. Using a minimal growth medium supplemented with 5′,5′,5′-trifluoroleucine (TFL), they were able to completely replace the canonical leucine with TFL in chloramphenicol acetyl transferase (CAT, 13 leucines) [20] and GFP (19 leucines) [21]. This came at significant cost, however, to protein function. They then asked whether native activity could be restored by directed evolution. Global leucine replacement in CAT resulted in a decrease in T_{50} (temperature at which half the native activity is lost after 30 minute incubation) of 9°C. Three rounds of error-prone PCR random mutagenesis and screening for improved retention of activity at high temperatures generated a CAT TFL variant whose stability matched that of the wild type enzyme and whose k_{cat}/K_m was only slightly decreased. Global incorporation of TFL in GFP led to protein misfolding and loss of fluorescence. Eleven rounds of error-prone PCR and FACS screening for improved fluorescence produced a GFP TFL mutant with fluorescence and folding properties similar to wild type GFP. Of the 19 leucine residues in GFP, 6 were mutated to other canonical amino acids, suggesting that there are specific locations at which TFL cannot be accommodated. It is interesting to note that expressing the same variant with leucine rather than TFL also improved folding and expression compared to wild type GFP. This suggests that the directed evolution experiments increased overall protein stability, which permits UAA incorporation.

It is particularly interesting to generate proteins with functions that are not available to or surpass those of natural proteins. Exciting work by Liu, Schultz and colleagues has laid the groundwork for the discovery and evolution of novel binding sequences containing UAAs

using phage display. In this work, a pIII hyperphage phagemid expression system was adapted for use in *E. coli* strains (X - *E. coli*) expressing orthogonal tRNA/aminoacyl-tRNA synthetase pairs that were evolved for incorporation of UAAs into proteins in response to the TAG stop codon (figure 1a) [22•]. Initial tests in a TAG-containing human-derived scFv demonstrated that phage production depended on the presence of the UAA; however, significant optimization was needed to limit expression bias due to inefficient incorporation of the UAA.

To test unnatural amino acid selection, a known post-translationally modified sulfotyrosine (sY)-containing antibody (412d) was examined (figure 1a - 1c) [22•,23••]. 412d is specific for the HIV coat protein gp120, and sulfation of two tyrosine residues within 412d (412d-sYsY, figure 1b) is required for antigen binding. Liu et al. recently reported the evolution of an orthogonal tRNA synthetase capable of directly incorporating sY into proteins [24]. Four rounds of phage selection against immobilized gp120 were performed using a doped library containing low levels of 412d-sYsY (1/2000) in a library of 412d sequences randomized at both sY positions and 4 other sites using a degenerate NNK codon (codes for 20 amino acids and the TAG stop codon). Sequencing of clones after each round of selection showed convergence to the wild type 412d sequence and new sequences also enriched in sY [22•]. Selection for tyrosine-containing 412d (412d-YY) yielded no such enrichment. In a follow-up experiment, 412d libraries were constructed in which both sY residues were fixed using UAA mutagenesis and nearby residues were randomized (figure 1b) [23••]. After 4 rounds of selection a variety of new sequences were isolated showing divergence from the wild type 412d (figure 1b). These sequences would not be available in nature due to strict sequence determinants for post-translation tyrosine sulfation. One isolated variant, 131-3, binds gp120 with an affinity (0.51 nM) similar to that of wild type 412d (0.63 nM). Kinetic analysis of Fab binding demonstrated that improvements in k_{on} over 412d were achievable; however, no decrease in k_{off} was observed (figure 1b). This inability to improve on the natural 412d antibody might reflect a limitation of the selection: Fab-antigen complex half-lives for wild type and evolved variants were determined to be at least 3 fold greater than the 30 minute experimental washing procedure used during phage selection (figure 1a), suggesting that longer washing steps may be required to select for tighter-binding antibody sequences [23••].

Phage selection is of course not limited to mimicking naturally occurring post-translational modifications. Selections were performed using a boronic acid-containing amino acid and a human-derived germline scFv randomized at 6 positions within CDR3 [25•]. Boronic acids demonstrate strong affinity for hydroxyl moieties, in particular 1,2-diols (figure 1d). Selection against an immobilized polyhydroxylated resin provided sequences enriched for boronic acids. The ability to select for sequences containing boronic acid UAAs may open the door to generating therapeutic proteins containing ‘chemical warheads’ specific for various biologically relevant sugars, or artificial lectins.

Limitations of site-specific UAA incorporation include limitations in the number of different UAAs that can be incorporated at one time, reduced translational efficiency at TAG codons and limited ribosomal incorporation of alternative amino acid structures (e.g. reduced incorporation efficiency of C-alpha substituted- and *D*-amino acids). Many of these limitations occur at the level of the ribosome. Chin and colleagues recently used directed evolution to develop a series of orthogonal bacterial ribosomes•mRNA pairs by mutating the Shine-Delgarno sequence of an mRNA and the corresponding complementary region of ribosomal RNA [26]. Using this system, orthogonal ribosomal evolution experiments can be undertaken without disrupting the endogenous cellular translational machinery. Using clever selection methodology, Neumann et al. evolved ribosomes that more efficiently decode TAG stop codons as well as four-base frameshift codons [27,28••]. This work is significant

as one requirement for UAA incorporation is the need for 'blank' codons that permit site-specific incorporation of UAA without disturbing the codon usage of the 20 naturally occurring amino acids. Neumann et al. have demonstrated the utllity of engineered ribosomes by efficiently incorporating two different unnatural amino acids into a single protein in high yield [28••]. In this report, simultaneous incorporation of a cross-reactive azide-containing amino acid and an alkyne-containing amino acid into a model protein permitted the formation of a covalent 'staple' crosslink between both UAAs using 'click' chemistry.

Non-natural Cofactors and Directed Evolution

Novel chemical diversity can also be introduced into proteins through artificial cofactors. Since early work by Kaiser and Whitesides [29,30], there has been considerable attention paid to developing artificial cofactors, particularly metallocenters, to generate new activities. Metalloprotein design is very broad and beyond the scope of this review, but several recent reviews highlight the multiple achievements in this area [6,31,32]. Of interest for directed evolution are *in vivo* or hybrid approaches that allow the new cofactor to be introduced in a way that leaves the protein amenable to mutagenesis and high-throughput screening. Recent work suggests that this is not far off, as improved metalloprotein construction together with rational protein engineering has produced enzymes with high selectivity and multiple hundreds of turnovers [33,34]. Directed evolution will provide an additional systematic approach to optimizing these systems.

One approach to making artificial metalloproteins uses biotin-streptavidin technology to create a hybrid catalyst consisting of the protein and a metal center, as pioneered by Whitesides [30] and expanded by Reetz, Ward and others [30,32,35]. Non-natural transition metal complexes attached by chemical linkers to biotin are immobilized within avidin or streptavidin by exploiting these proteins' extremely strong affinity ($Kd \sim 10^{14}\ M^{-1}$) for biotin. The inherent catalytic activity of the transition metal complex in association with the three-dimensional protein active site provides a platform for stereospecific organometallic catalysis. Streptavidin is thermostable, and both streptavidin and avidin have known crystal structures and are amenable to mutagenesis and bacterial expression, making them good candidates for directed evolution.

Several labs have explored a variety of reactions using this approach, including the asymmetric Rh-catalyzed hydrogenation of alkenes [34,36], Pd-catalyzed allylic alkylation [37], Mn-salen-catalyzed asymmetric sulfoxidation [38], and Rh- or Ru-catalyzed transfer hydrogenation [33,39]. Ward and colleagues have combined chemical optimization of transition metal-biotin conjugates with site-directed mutagenesis to produce hybrid enzymes with high conversion and enantioselectivity for various reactions [35,36,37]. A recent report describes immobilizing one of the four streptavidin monomers to a biotinylated solid support, which accelerates the production of homogeneous metalloproteins and enables moderate throughput screening [39]. By saturation mutagenesis of specific residues in the active site of streptavidin, the authors generated Ru-based transfer hydrogenation catalysts with up to 98% ee for the reduction of *p*-methylacetophenone [33]. Reetz and colleagues have optimized a high-throughput screen for the hydrogenation of α-acetamido-acrylic acid esters [34]. Iterative saturation mutagenesis of streptavidin and screening produced mutant biotinylated diphosphine-Rh complex hybrid catalysts with improved enantioselectivity (both *R* and *S*).

Artificial metalloprotein catalysts often require a considerable quantity of purified protein to exhibit measurable activity [34]. In addition, reactions of new metallocofactors with other cellular components may negatively affect catalyst stability or generate non-specific reaction

centers that reduce stereoselectivity. This makes it impossible to screen hundreds of mutant enzymes, as large-scale expression and purification are required to produce enough homogeneous protein for examination. In the end, new approaches that permit efficient *in vivo* generation of artificial metallocenters may be useful. As a start, Schultz and colleagues have recently demonstrated the genetic incorporation of a variety of novel unnatural metal-chelating amino acids [40,41]. It will be interesting to see this technology adapted to introduce pre-organized transition metal-containing amino acids into proteins *in vivo*.

Exciting new opportunities for artificial enzyme design and evolution are enabled by new technology for incorporation of non-natural metalloporphyrins into proteins. Traditional methods have required harsh *in vitro* unfolding, stripping and heme reconstitution protocols. Marletta and colleagues recently demonstrated that an engineered *E. coli* strain deficient in heme biosynthesis can take up exogenous artificial metalloporphyrins from defined media and incorporate them into heme proteins [42]. In a recent report, they introduced Ru(II)-CO mesoporphyrin IX into myoglobin and the H-NOX (heme nitric oxide/oxygen) binding protein from *Thermoanaerobacter tengcongensis***,** resulting in fluorescent biosensors for oxygen [43•]. Lelyveld et al. have recently introduced an alternative method for *in vivo* non-natural metalloporphyrin incorporation by exploiting the relaxed substrate specificity of a naturally occurring bacterial heme transport protein, ChuA. By overexpressing ChuA in iron-deficient media in the presence of Mn^{3+}-protoporphyrin IX, they produced a manganese-substituted cytochrome P450 that demonstrated an enhanced metal-induced T_1 relaxation of water compared to the native heme protein [44•]. This will significantly enhance the performance of protein-based, ligand-sensitive MRI contrast sensors made by directed evolution [45•]. These strategies for the *in vivo* generation of proteins with novel porphyrins have opened the door to their modification and optimization by directed evolution.

Computational Enzyme Design and Directed Evolution

While introducing unnatural amino acids or metallocenters provides a direct means to establish new chemistry, engineering new enzymatic reactivities using the building blocks nature has already provided represents one of the ultimate tests of our understanding of protein structure and function. Recent work from the Baker laboratory has demonstrated the progress made in computational design of new enzyme catalysts [8•,9,10••]. Unfortunately, the catalytic activities of these designed enzymes do not come close to the levels exhibited by natural enzymes. But that may not be true for long. Rational design can provide a starting point by arranging at least the minimum features for catalysis within an active site; it should be possible for laboratory evolution to improve upon these rudimentary designs.

A simplified description of the computational design methodology involves several key steps: (1) using computational algorithms to generate idealized active sites to stabilize an enzymatic transition state, (2) identifying protein scaffolds from a library of folds that can fit the desired transition state model (Rosetta Match), (3) optimizing the entire active site within selected scaffolds, and (4) the synthesis, expression and characterization of a number of top ranking designs. Using this approach, the Baker lab has designed novel enzymes that catalyze a variety of reactions including a retro-aldol condensation [9], the Kemp elimination [10••] and a stereoselective bimolecular Diels-Alder reaction [8•]. These works demonstrate that measurable non-natural enzyme activity can be engineered within protein scaffolds. The resulting activity, however, is generally extremely low, with rate enhancements over the uncatalyzed reaction reaching up to 10^5 in only the most successful examples. In some cases, rational mutational analysis has led to small improvements in activity [8•,9,46•]. For example, during the biochemical evaluation of a computationally designed retroaldolase, deletion of two hydrogen bonding residues that were believed to be

key for catalysis surprisingly led to a five-fold increase in enzyme activity [46•]. This work underscores design failures but also highlights how subsequent improvements can be achieved through mutation. If enough product is produced by the initial enzyme design to enable high-throughput screening, directed evolution can be applied to yield further improvements in activity.

Optimization of novel catalysis by directed evolution has been demonstrated for an enzyme computationally designed for the Kemp elimination of 5-nitrobenzisoxazole, a model proton transfer reaction (figure 2a) [10••]. Three elements were designed into the transition state model to promote catalysis (figure 2a and b): (1) A catalytic base (either a sole Asp/Glu residue, or a His residue polarized by an Asp or Glu residue (His-Asp dyad) was oriented in position to deprotonate the oxazole C-H moiety. (2) A hydrogen bond donor (Ser or Lys) was positioned to stabilize the accumulating negative charge that results on the phenolic oxygen. (3) An aromatic residue was added to form a favorable π-stacking interaction with the planar transition state (figure 2b). Of the 59 designs tested, 8 produced active enzymes with various active site structures. k_{cat}/K_mvalues ranged from ~ 6 to 160 $M^{-1}s^{-1}$, with rate accelerations as much as 10^5.

The product formed by the Kemp reaction has a unique absorbance spectrum, which can be used for high-throughput activity-based screening. This allowed the authors to use directed evolution to screen random mutants for improved activity. Construct Ke07, for which there was a crystal structure, was chosen as the starting parent. Seven rounds of iterative mutagenesis (error-prone PCR and DNA shuffling) and screening resulted in a variant with 8 additional mutations, a more than 200-fold improvement in k_{cat}/K_m (from ~12 to 2,600 $M^{-1}s^{-1}$) and a ~100-fold increase in k_{cat}/k_{uncat} (from 10^4 to 10^6) (figure 2b and c). Recently Khersonsky et al. have solved the crystal structures of several intermediates from the directed evolution process [47••]. From these, they were able to identify several determinants of the improved activity as well as limitations of the initial design. The crystal structure of Ke07 showed that the distance between the designed hydrogen bond donor, Lys222, and the catalytic Glu101 was shorter than anticipated (figure 2b), suggesting that the interaction of these two residues may prevent deprotonation of the oxazole ring by Glu101. An Ile7Asp mutation introduced during the 4th round of selection resulted in a salt bridge to Lys222, drawing it away from Glu101 and possibly allowing both residues to stabilize the transition state more efficiently (figure 2c). Other mutations have been associated with tuning the ligand-binding environment, altering the pKa of the catalytic base and improving its interaction with the substrate [47••].

While the results outlined above can be considered a tremendous success in the *de novo* design of enzymatic activity, there are also cautionary lessons. Many natural enzymes enhance the reaction rate (k_{cat}/k_{uncat}) by more than a factor of 10^{12} [11,48], whereas the most successful designed enzymes only boast a10^6-fold rate enhancement, even after further optimization by directed evolution. These values are similar to those established by catalytic antibodies decades earlier. A catalytic antibody for the Kemp elimination developed by Hilvert and colleagues showed a k_{cat}/K_m of 5,500 $M^{-1}s^{-1}$, eerily similar to the designed Kemp eliminase (2,600 $M^{-1}s^{-1}$) [49]. While there is clearly plenty of room for improvement, seven rounds of directed evolution of the designed Kemp eliminase provided only a 200-fold increase in k_{cat}/K_m; additional rounds of mutagenesis and screening yielded no further benefit. This, along with myriad failed efforts to used directed evolution to improve catalytic antibodies, underscores the fact that not all protein designs are evolvable. The limited improvements observed in designed proteins reflect their local, non-productive fitness landscapes marked by the lack of nearby fitness peaks (not every bad enzyme is near a good enzyme) or by local maxima where subsequent mutations are not advantageous (figure 3) [1]. The goal of course is to find those starting points that lie on large fitness

peaks, where improvement is not only possible, but easy to access. Mechanistic studies of designed proteins, updated crystal structures bound to transition state analogues, and improvements in computational methodology may lead to improvements in designed enzymes [48]. But, the most important task may well be to learn what makes one enzyme evolvable and another not.

Conclusions

Engineering novel functions in proteins presents an exciting challenge for protein engineering. However, efforts to date to generate novel catalysts have primarily demonstrated that we are getting good at making bad enzymes. Making good enzymes will require a whole new level of insight, or new methodologies altogether. Technologies such as those reviewed here can be viewed as creating new starting points where optimization through evolution might be possible. This is how Nature would approach the problem. Yet, while directed evolution offers a mechanism for protein optimization through the accumulation of beneficial mutations, there is no guarantee that the new proteins lie at the base of a fitness peak [1]. Lessons from these initial successes and failures should stimulate improvements in design approaches leading to new paths for discovery. Finally, advances in UAA incorporation in combination with new cloning strategies such as TAG codon scanning [50] may permit the random introduction of UAAs within a protein sequence. This way, directed evolution could improve protein fitness using functionality not available to natural proteins.

Acknowledgments

The authors thank Mary Farrow, Christopher Snow, Philip Romero, and Crystal Dilworth for thoughtful discussions during the preparation of this manuscript. EB is supported by a Ruth M Kirschstein NIH postdoctoral fellowship Award Number F32GM087102 from the National Institute of General Medical Sciences. We also acknowledge support from the Jacobs Institute for Molecular Engineering for Medicine and NIH 1-R01-DA028299-01.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

• of special interest

•• of outstanding interest

1. Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. Nat Rev Mol Cell Biol. 2009; 10:866–876. [PubMed: 19935669]
2. Turner NJ. Directed evolution drives the next generation of biocatalysts. Nat Chem Biol. 2009; 5:567–573. [PubMed: 19620998]
3. Jackel C, Hilvert D. Biocatalysts by evolution. Curr Opin Biotechnol. 2010; 21
4. Khersonsky O, Tawfik DS. Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu Rev Biochem. 2010; 79:471–505. [PubMed: 20235827]
5. Liu CC, Schultz PG. Adding new chemistries to the genetic code. Annu Rev Biochem. 2010; 79:413–444. [PubMed: 20307192]
6. Heinisch T, Ward TR. Design strategies for the creation of artificial metalloenzymes. Curr Opin Chem Biol. 2010; 14:184–199. [PubMed: 20071213]
7. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. Science. 2003; 302:1364–1368. [PubMed: 14631033]
8. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science. 2010; 329:309–313. [PubMed: 20647463] • This work

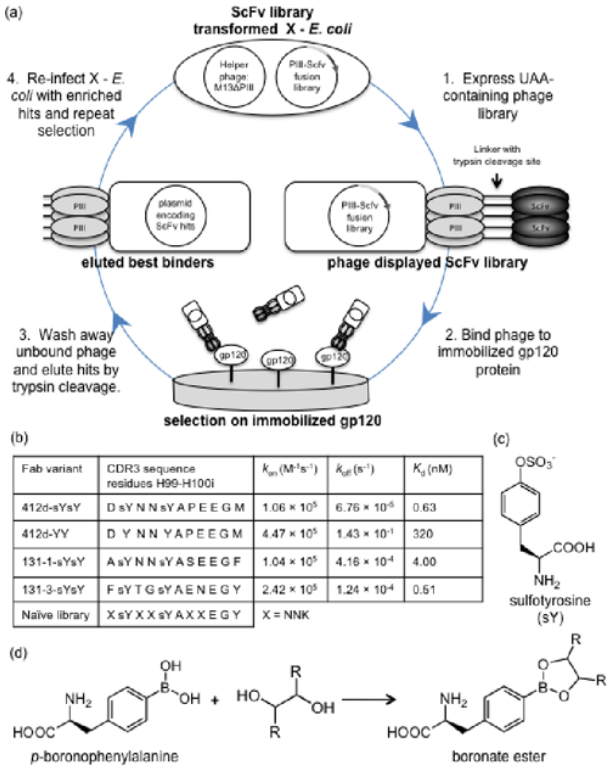
expands on previous computational enzyme design methodology to construct an artificial Diels-Alder catalyst between a model diene and dienophile. There are no known enzymes that catalyze a biomolecular Diels-Alder reaction, which is an extremely important tool for chemical synthesis, While the activity of the designed enzyme is low, it forms a single enantiomer out of a possible four suggesting that computational methodology may prove useful for the design of stereoselective enzymes.

9. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF 3rd, et al. De novo computational design of retro-aldol enzymes. Science. 2008; 319:1387–1391. [PubMed: 18323453]
10. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, et al. Kemp elimination catalysts by computational enzyme design. Nature. 2008; 453:190–195. [PubMed: 18354394] •• Computational design was used to create an enzyme that catalyzes the Kemp-elimination, a reaction not catalyzed by naturally occurring enzymes. Initial designs demonstrating limited activity were improved significantly (> 200 fold) using directed evolution.
11. Wolfenden R, Snider MJ. The depth of chemical time and the power of enzymes as catalysts. Acc Chem Res. 2001; 34:938–945. [PubMed: 11747411]
12. Voloshchuk N, Montclare JK. Incorporation of unnatural amino acids for synthetic biology. Mol Biosyst. 2010; 6:65–80. [PubMed: 20024068]
13. Wang Q, Parrish AR, Wang L. Expanding the genetic code for biological studies. Chem Biol. 2009; 16:323–336. [PubMed: 19318213]
14. Grunewald J, Tsao ML, Perera R, Dong L, Niessen F, Wen BG, Kubitz DM, Smider VV, Ruf W, Nasoff M, et al. Immunochemical termination of self-tolerance. Proc Natl Acad Sci U S A. 2008; 105:11276–11280. [PubMed: 18685087] • This work demonstrates that an artificial p-nitrophenylalanine amino acid can be introduced into autologous proteins to elicit an immune response, overcoming the immunological barrier of self-tolerance. This may facilitate the generation of therapeutic antibodies towards cancer or chronic degenerative diseases.
15. Mills JH, Lee HS, Liu CC, Wang J, Schultz PG. A genetically encoded direct sensor of antibody-antigen interactions. Chembiochem. 2009; 10:2162–2164. [PubMed: 19672911]
16. Dieterich DC, Hodas JJ, Gouzer G, Shadrin IY, Ngo JT, Triller A, Tirrell DA, Schuman EM. In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. Nat Neurosci. 2010; 13:897–905. [PubMed: 20543841] • This work demonstrates that ‘click chemistry’ reactive unnatural methionine homologues can be globally incorporated into newly expressed proteins in neurons allowing metabollic labeling with appropriate fluorophores. Pulse chase application of two different unnatural amino acids allowed selective labeling of proteins produced at sequential time periods.
17. Ngo JT, Champion JA, Mahdavi A, Tanrikulu IC, Beatty KE, Connor RE, Yoo TH, Dieterich DC, Schuman EM, Tirrell DA. Cell-selective metabolic labeling of proteins. Nat Chem Biol. 2009; 5:715–717. [PubMed: 19668194] • Cells expressing a mutant methionyl-tRNA synthetase can selectively incorporate a click reactive azidonorleucine amino acid into newly expressed proteins. Protein labeling with biophysical probes in a complex cellular mixture results in selective labeling of cells harboring the mutant tRNA synthetase.
18. Wong JT. Membership mutation of the genetic code: loss of fitness by tryptophan. Proc Natl Acad Sci U S A. 1983; 80:6303–6306. [PubMed: 6413975]
19. Bacher JM, Ellington AD. Selection and characterization of Escherichia coli variants capable of growth on an otherwise toxic tryptophan analogue. J Bacteriol. 2001; 183:5414–5425. [PubMed: 11514527]
20. Montclare JK, Tirrell DA. Evolving proteins of novel composition. Angew Chem Int Ed Engl. 2006; 45:4518–4521. [PubMed: 16763955]
21. Yoo TH, Link AJ, Tirrell DA. Evolution of a fluorinated green fluorescent protein. Proc Natl Acad Sci U S A. 2007; 104:13887–13890. [PubMed: 17717085]
22. Liu CC, Mack AV, Tsao ML, Mills JH, Lee HS, Choe H, Farzan M, Schultz PG, Smider VV. Protein evolution with an expanded genetic code. Proc Natl Acad Sci U S A. 2008; 105:17688–17693. [PubMed: 19004806] • This work combines phage display technology with site-specific unnatural amino acid incorporation methodology to permit selection of binding proteins containing

unnatural amino acids. The methodology is validated by selection for an anti-HIV containing antibody sequence that has been genetically encoded with a sulfotyrosine modification.

23. Liu CC, Choe H, Farzan M, Smider VV, Schultz PG. Mutagenesis and evolution of sulfated antibodies using an expanded genetic code. Biochemistry. 2009; 48:8891–8898. [PubMed: 19715291] •• Expanding on the work in reference [22•], unnatural amino acid phage display is used to select for new sulfotyrosine-modified antibody sequences that display binding characteristics on par with a natural antibody. In addition, the selection process is not limited by sequence requirements for post-translational tyrosine sulfation, suggesting that this method can be used to explore new sequence space that cannot be explored by nature.
24. Liu CC, Cellitti SE, Geierstanger BH, Schultz PG. Efficient expression of tyrosine-sulfated proteins in E. coli using an expanded genetic code. Nat Protoc. 2009; 4:1784–1789. [PubMed: 20010929]
25. Liu CC, Mack AV, Brustad EM, Mills JH, Groff D, Smider VV, Schultz PG. Evolution of proteins with genetically encoded “chemical warheads”. J Am Chem Soc. 2009; 131:9616–9617. [PubMed: 19555063] • Expanding on the work in reference [22•], the authors demonstrate that new chemical functionality can be introduced into phage-displayed proteins. This paper shows that an artificial boronic acid moiety can be introduced into antibody sequences to produce proteins that covalently bind sugars.
26. Rackham O, Chin JW. A network of orthogonal ribosome x mRNA pairs. Nat Chem Biol. 2005; 1:159–166. [PubMed: 16408021]
27. Wang K, Neumann H, Peak-Chew SY, Chin JW. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. Nat Biotechnol. 2007; 25:770–777. [PubMed: 17592474]
28. Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature. 2010; 464:441–444. [PubMed: 20154731] •• Expanding on previous work [26,27], the authors demonstrate that orthogonal ribosome • mRNA pairs can be used for ribosomal evolution. The directed evolution of ribosomal libraries carrying mutations focused around the tRNA decoding center of the 16S rRNA produced novel ribosomes that efficiently decode four-base frameshift codons as well as the TAG stop codon. This effectively raises the number of blank codons available for unnatural amino acid mutagenesis and enables the efficient incorporation of 2 different unnatural amino acids into a single protein.
29. Levine HL, Nakagawa Y, Kaiser ET. Flavopapain: synthesis and properties of semi-synthetic enzymes. Biochem Biophys Res Commun. 1977; 76:64–70. [PubMed: 869949]
30. Wilson ME, Whitesides GM. Conversion of a protein to a homogenous asymmetric hydrogenation catalyst by site-specific modification with a diphosphinethodium(I) moiety. J Am Chem Soc. 1978; 100:306–307.
31. Lu Y, Yeung N, Sieracki N, Marshall NM. Design of functional metalloproteins. Nature. 2009; 460:855–86. [PubMed: 19675646]
32. Reetz, MT. Directed evolution of stereoselective hybrid catalysts. In: Ward, TR., editor. Topics in Organometallic Chemistry: Bio Inspired Catalysts. Vol. 24. Springer; 2009. p. 63-92.
33. Pordea A, Creus M, Letondor C, Ivanova A, Ward TR. Improving the enantioselectivity of artificial transfer hydrogenases based on the biotin-streptavidin technology by combinations of point mutations. Inorganica Chimica Acta. 2010; 363:601–604.
34. Reetz MT, Rentzsch M, Pletsch A, Maywald M, Maiwald P, Peyralans JJP, Maichele A, Fu Y, Jiao N, Hollmann F, et al. Directed evolution of enantioselective hybrid catalysts: a novel concept in asymmetric catalysis. Tetrahedron. 2007; 63:6404–6414.
35. Creus M, Ward TR. Designed evolution of artificial metalloenzymes: protein catalysts made to order. Org Biomol Chem. 2007; 5:1835–1844. [PubMed: 17551630]
36. Klein G, Humbert N, Gradinaru J, Ivanova A, Gilardoni F, Rusbandi UE, Ward TR. Tailoring the active site of chemzymes by using a chemogenetic-optimization procedure: towards substrate-specific artificial hydrogenases based on the biotin-avidin technology. Angew Chem Int Ed Engl. 2005; 44:7764–7767. [PubMed: 16276543]

37. Pierron J, Malan C, Creus M, Gradinaru J, Hafner I, Ivanova A, Sardo A, Ward TR. Artificial metalloenzymes for asymmetric allylic alkylation on the basis of the biotin-avidin technology. Angew Chem Int Ed Engl. 2008; 47:701–705. [PubMed: 18181264]
38. Pordea A, Mathis D, Ward TR. Incorporation of biotinylated manganese-salen complexes into streptavidin: New artificial metalloenzymes for enantioselective sulfoxidation. Journal of Organometallic Chemistry. 2009; 694:930–936.
39. Creus M, Pordea A, Rossel T, Sardo A, Letondor C, Ivanova A, Letrong I, Stenkamp RE, Ward TR. X-ray structure and designed evolution of an artificial transfer hydrogenase. Angew Chem Int Ed Engl. 2008; 47:1400–1404. [PubMed: 18176932]
40. Lee HS, Schultz PG. Biosynthesis of a site-specific DNA cleaving protein. J Am Chem Soc. 2008; 130:13194–13195. [PubMed: 18788806]
41. Lee HS, Spraggon G, Schultz PG, Wang F. Genetic incorporation of a metal-ion chelating amino acid into proteins as a biophysical probe. J Am Chem Soc. 2009; 131:2481–248. [PubMed: 19193005]
42. Woodward JJ, Martin NI, Marletta MA. An Escherichia coli expression-based method for heme substitution. Nat Methods. 2007; 4:43–4. [PubMed: 17187078]
43. Winter MB, McLaurin EJ, Reece SY, Olea C Jr. Nocera DG, Marletta MA. Ru-porphyrin protein scaffolds for sensing O2. J Am Chem Soc. 2010; 132:5582–5583. [PubMed: 20373741] • These authors have developed an in vivo protocol for the introduction of artificial metalloporphyrins into proteins [40]. This methodology circumvents the harsh conditions used for traditional in vitro heme swapping protocols. The authors go on to demonstrate the biosynthetic incorporation of Ru-porphyrins into myoglobin and the thermostabile heme nitric oxide/oxygen binding domain from Thermoanaerobacter tengcongensis to generate sensors for molecular oxyge
44. Lelyveld VS, Brustad EM, Arnold FH, Jasanoff A. Metal-substituted protein MRI contrast agents engineered for enhanced relaxivity and ligand sensitivity. J Am Chem Soc. 2010 article in review. • By overexpressing a naturally occuring bacterial heme transport protein, these authors demonstrate that non-natural metalloporphyrins can be taken up by E. coli and co-translationally introduced into the heme binding domain of cytochrome P450 BM3 from B. megaterium. The manganese containing protein demonstrates improved metal induced water T1 relaxivity, which could improve the development of protein based MRI contrast sensors [45].
45. Shapiro MG, Westmeyer GG, Romero PA, Szablowski JO, Kuster B, Shah A, Otey CR, Langer R, Arnold FH, Jasanoff A. Directed evolution of a magnetic resonance imaging contrast agent for noninvasive imaging of dopamine. Nat Biotechnol. 2010; 28:264–270. [PubMed: 20190737] • These authors show that the heme center of cytochrome P450 BM3 from B. megaterium promotes T1 relaxation of axially ligated water molecules which can be observed by MRI. They go on to show that this MRI signal is sensitive to ligand binding and by applying directed evolution, evolve a series of MRI sensors responsive to the neurotransmitter dopamine. These proteins are used to image dopamine in living rat brains.
46. Lassila JK, Baker D, Herschlag D. Origins of catalysis by computationally designed retroaldolase enzymes. Proc Natl Acad Sci U S A. 2010; 107:4937–4942. [PubMed: 20194782] • This work pursues rigorous biochemical characterization of a computationally designed retro-aldolase [9]. This effort elucidates the biochemical basis of many of the successes as well as failures of computationally designed enzyme activit
47. Khersonsky O, Rothlisberger D, Dym O, Albeck S, Jackson CJ, Baker D, Tawfik DS. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. J Mol Biol. 2010; 396:1025–1042. [PubMed: 20036254] •• The authors have solved the crystal structure of a designed Kemp eliminase [10••] as well as series of progeny that were the result of directed evolution. This work provides structural evidence for the basis of weak initial enzyme activity as well as the basis for improved catalysis by the accumulated mutations.
48. Baker D. An exciting but challenging road ahead for computational enzyme design. Protein Sci. 2010
49. Thorn SN, Daniels RG, Auditor MT, Hilvert D. Large rate accelerations in antibody catalysis by strategic use of haptenic charge. Nature. 1995; 373:228–230. [PubMed: 7816136]
50. Daggett KA, Layer M, Cropp TA. A general method for scanning unnatural amino acid mutagenesis. ACS Chem Biol. 2009; 4:109–113. [PubMed: 19199564]



Fab variant	CDR3 sequence residues H99-H100i	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	K_d (nM)
412d-sYsY	D sY N N sY A P E E G M	1.06×10^5	6.76×10^{-5}	0.63
412d-YY	D Y N N Y A P E E G M	4.47×10^5	1.43×10^{-1}	320
131-1-sYsY	A sY N N sY A S E E G F	1.04×10^5	4.16×10^{-4}	4.00
131-3-sYsY	F sY T G sY A E N E G Y	2.42×10^5	1.24×10^{-4}	0.51
Naïve library	X sY X X sY A X X E G Y	X = NNK		

Figure 1.

Directed evolution of novel UAA-dependent binding proteins using phage display [22•,23••, 25•]. (a) General phage display protocol for selection of HIV gp120 specific ScFv proteins. X – *E. coli* refers to bacterial strains engineered for the incorporation of UAA into proteins in response to the TAG stop codon. (b) CDR3 sequences of gp120 specific sulfotyrosine (sY)-containing antibody, 412d, and new evolved variants are shown. Replacing sY with natural tyrosine reduces binding of the wild type (412d-YY) and evolved variants. (c) Structure of sulfotyrosine. (d) Novel reactive groups such as boronic acids can provide chemical warheads for binding molecules such as hydroxyl-rich sugars.

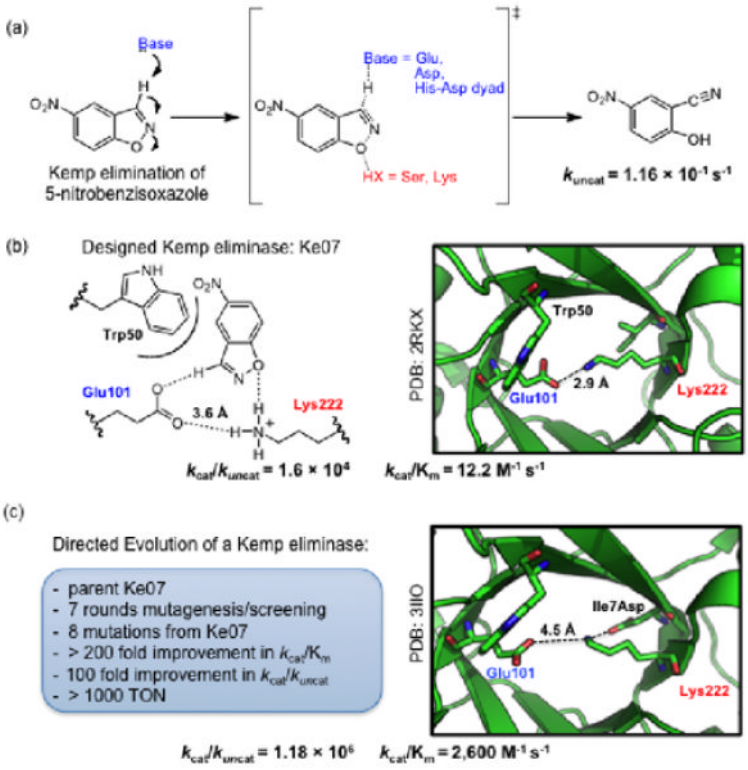


Figure 2.

Directed evolution of a computationally designed Kemp eliminase [10••,47••]. (a) Kemp elimination reaction mechanism (left), transition state (middle), and product (right) are shown. (b) A schematic of the minimal computationally designed active site of parent Kemp eliminase Ke07 is shown on the left. In the designed model, the distance between Glu101 and Lys222 is approximately 3.6 Å. The observed distance between these two residues is significantly shorter, according to the crystal structure (right, PDB: 2RKX). (c) 7 rounds of directed evolution starting from Ke07 improved catalytic efficiency more than 200 fold. A crystal structure for a variant from round 4 of evolution (4-1E/11H, PDB 3IIO) is shown on the right. The interaction between Glu101 and K222 is significantly diminished in this structure.

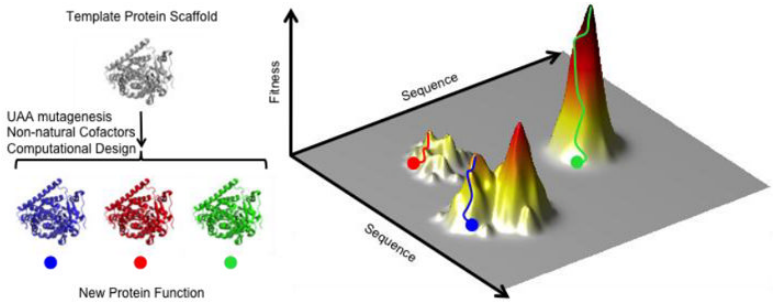


Figure 3.

Developing technologies such as unnatural amino acid (UAA) mutagenesis, non-natural cofactor engineering and computational design can facilitate the introduction of new function, distinct from those demonstrated in Nature, within a protein scaffold. Fitness is a measure of how well protein sequences perform the target function. Most proteins possess no activity (shown in grey on the fitness landscape). Successful designs (shown as circles) may show weak initial activities that can be starting points for further optimization by directed evolution. Directed evolution leading only to small improvements may indicate that an initial design lies near low fitness peaks (red trajectory) or that further improvements are limited by the presence of local optima (blue trajectory). An ideal design would yield a sequence that lies on a large fitness peak (green trajectory), where multiple rounds of mutagenesis and screening produce a highly functional protein.