Published in final edited form as:
Curr Opin Biotechnol. 2010 December ; 21(6): 734–743. doi:10.1016/j.copbio.2010.08.011.

Beyond directed evolution - semi-rational protein engineering and design

Stefan Lutz
Department of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, GA, 30322

Abstract

Over the last two decades, directed evolution has transformed the field of protein engineering. The advances in understanding protein structure and function, in no insignificant part a result of directed evolution studies, are increasingly empowering scientists and engineers to device more effective methods for manipulating and tailoring biocatalysts. Abandoning large combinatorial libraries, the focus has shifted to small, functionally-rich libraries and rational design. A critical component to the success of these emerging engineering strategies are computational tools for the evaluation of protein sequence datasets and the analysis of conformational variations of amino acids in proteins. Highlighting the opportunities and limitations of such approaches, this review focuses on recent engineering and design examples that require screening or selection of small libraries.

INTRODUCTION

Enzymes are highly versatile and proficient catalysts. Optimized by Darwinian evolution over millions of years, they can greatly accelerate chemical reactions while ensuring high substrate specificity, as well as exquisite enantio and stereoselectivity. These performance features make biocatalysts attractive candidates for asymmetric synthesis in the laboratory and industrial processes. However, there are often significant discrepancies between an enzyme's function in nature and the specific requirements for ex vivo applications envisioned by scientists and engineers. Enzyme engineering by directed evolution has become the strategy of choice for tailoring the catalytic, biophysical and molecular recognition properties of target proteins [1].

Traditionally, directed evolution relies on an iterative two-step protocol, initially generating molecular diversity by random mutagenesis and in vitro recombination, then identifying library members with improvements in desired phenotype by high-throughput screening or selection. The approach can be problematic as even protein libraries with millions of members still sample only a tiny fraction of the vast sequence space possible for an average protein. Biases in the experimental methods and the degeneracy of the genetic code further skew and restrict the library design [2]. Rather than addressing these problems through bigger libraries and more screening or selection, many researchers are moving beyond traditional directed evolution, instead advocating new strategies for designing smaller, higher quality libraries.

Often referred to as semi-rational, smart or knowledge-based library design, these approaches utilize information on protein sequence, structure and function, as well as computational predictive algorithms to preselect promising target sites and limited amino acid diversity for

Corresponding author: Lutz, S. (sal2@emory.edu).

protein engineering. The focus on specific amino acid positions translates into dramatically reduced library sizes while the consideration of evolutionary variability, topological constraints and mechanistic features to weigh in on amino acid identity can result in libraries with higher functional content. In addition to the sequence and structure-based design strategies, QM and MD calculations, as well as machine-learning algorithms have become invaluable tools to effectively explore the impact of amino acid substitutions on protein structure and stability. Together, these concepts offer promising predictors for altering protein features such as substrate specificity, stereoselectivity and stability by enzyme redesign (but leaving the catalytic machinery of the native biocatalyst intact), as well as the creation of new function by de novo design.

From a practical perspective, semi-rational protein engineering can significantly increase the efficiency of biocatalyst tailoring. Besides typically requiring fewer iterations to identify variants with the desired phenotype, the generation of small high-quality libraries can largely eliminate the need for high-throughput methods in library analysis. The smaller number of variants also creates new opportunities for the evaluation of library members by protocols not amendable to a high-throughput format. Finally, these design strategies provide an intellectual framework to predict and rationalize experimental findings, taking the field from discovery-based towards hypothesis-driven protein engineering. To highlight the rapidly growing number of successful enzyme engineering studies by semi-rational and computer-guided protein design, this review concentrates (with few exceptions) on recent studies that required libraries of less than 1000 members (Table 1).

SEQUENCE-BASED ENZYME REDESIGN

A popular strategy to more effectively navigate and identify “islands” of functionality in protein sequence space has been the use of evolutionary information. Multiple sequence alignments (MSAs) and phylogenetic analyses have become standard tools for the exploration of amino acid conservation and ancestral relationships among groups of homologous protein sequences and structures. Whether these statistics are derived from large natural sequence pools or through neutral drift experiments in the laboratory, the data is valued by protein engineers for identifying functional hot spots, assessing local amino acid variability and guiding back-to-consensus designs. Among the large number of data analysis software [3,4], two new internet-based computational tools are noteworthy.

The HotSpot Wizard server (http://loschmidt.chemi.muni.cz/hotspotwizard) combines information from extensive sequence and structure database searches with functional data to create a mutability map for a target protein [5]. The system’s performance “at the bench” was demonstrated as part of the engineering of haloalkane dehalogenase (DhaA) from *Rhodococcus rhodochrous* [6] (*vide infra*). Similarly, the commercial 3DM database (http://3dmcsis.systemsbiology.nl) integrates protein sequence and structure data from GenBank and the PDB to create comprehensive alignments of protein superfamilies [7]. These data collections can effectively be searched through a series of built-in filters for evolutionary features such as correlated mutations and conservation. Furthermore, information on the functional role of individual amino acid residues, as well as details on their mutability is accessible. As a bonus for the busy researcher, the 3DM database tracks the literature and automatically updates itself with new structural and functional information. Its potential benefits for protein engineering was demonstrated in three recent studies, probing model enzymes for changes in activity and substrate specificity [8,9], as well as improvements in enantioselectivity [10•]. The latter study, investigating the impact of specific residues on the enantioselectivity of *Pseudomonas fluorescens* esterase, is particularly interesting as control experiments demonstrate the qualitative advantages of evolution-guided library design. 3DM analysis on over 1700 members of the α/β-hydrolase fold family was applied to define

(evolutionarily) allowed amino acid substitutions in four positions near the active site. In lab experiments, the esterase library comprised of allowed substitutions clearly outperformed two controls which carried either random or not allowed substitutions, affording functional variants with higher frequency and superior catalytic performance.

Further advances in enzyme redesign are possible through the combination of sequence-based information with computational modeling tools. While performing molecular dynamics simulations to explain moderate functional gains in DhaA variants previously identified in random mutagenesis and DNA shuffling libraries [11], Damborsky and coworkers noticed that beneficial mutations did not affect the active site directly but instead improved catalytic performance through conformational changes in the enzyme's two access tunnels [12]. These observations prompted further in silico studies that focused specifically on the product release through these tunnels as a function of nearby amino acid substitutions which in turn led to the identification of five key residues located at the tunnel entries and inside the tunnels (Figure 1). Guided by HotWizard, specific substitutions were introduced in two of the positions while the impact of changing the remaining three amino acid residues on enhanced dehalogenase activity could easily be evaluated by small site-saturation mutagenesis libraries. The strategy was not only effective for isolating enzyme variants with significantly higher activity but it also proved critically important for identifying the most beneficial amino acid substitutions.

Another conceptionally very interesting and effective strategy for the identification of mutational hotspots was applied for the engineering of prolyl endopeptidase from *Shingomonas capsulata* [13]. Aiming to improve enzyme activity at low pH and increase protease resistance, Khosla and coworkers used a MSA of 100 peptidase homologs to capture the diversity at individual amino acid positions. The variability at each site was then quantified through the assignment of so called position and substitution scores and yielded a list of 30 specific, potentially beneficial mutations. Next, different combinations of these specific amino acid substitutions were introduced into 47 variants through whole-gene synthesis. Working with such a small focused library, although more challenging to prepare than traditional mutagenesis, enabled an in-depth functional evaluation of each candidate without the need for sophisticated selection or high-throughput screening methods. Analysis of the assay data via machine-learning algorithms [14,15] ranked 22 of the substitutions as beneficial and, following a second round of library preparation and functional evaluation, narrowed down the selection to five amino acid changes that slightly raised the catalytic activity of the endopeptidase while improving its resistance to pepsin digestion by 200-fold.

Rather than considering only the amino acid diversity in modern-day proteins, the REAP (Reconstructing Evolutionary Adaptive Paths) method travels back in evolutionary time to exploit sequence data from ancestral proteins for the creation of focused and functional-enriched enzyme libraries [16]. The method by Chen *et al*. uses phylogeny to identify mutations in gene sequences that emerged during functional divergence from a common universal ancestor. Arguing that the variation and conservation of individual amino acid positions at these pivotal moments in evolutionary history could pinpoint functional hot spots in a protein sequence, the authors explored the split in the phylogenetic tree of Family A DNA polymerases into viral and non-viral subgroups. Viral polymerases are known for their greater substrate promiscuity compared to the non-viral enzymes, hence differences in the sequences of the predicted ancestral polymerase at this evolutionary branching point could account for substrate specificity changes without infringing on catalytic performance. Of the approximately 100 sites of interest (~18% of protein sequence) from this evolutionary analysis, 35 positions were selected based on their proximity to the active site and additional biochemical information. Finally, the amino acid diversity at these positions was limited to natural variations found in the phylogeny which reduced the size of the experimental library to only 93 variants! Thirty of the first-round candidates were functional polymerases - eight of them with greater than

wild type activity for unnatural dNTPs and two of those with excellent performance for novel 3'-modified dNTPs as substrates.

STRUCTURE-BASED ENZYME REDESIGN

Protein function is usually intimately linked to three-dimensional structure, making the substitution of one or more amino acids in macromolecules a function of not just sequence context but also structural topology. The rapidly growing number of protein structures in the PDB and advances in homology modeling offer valuable assistance for protein engineers to more effectively locate key residues near active sites and at domain interfaces or hinge regions which can translate into superior library designs. Three recent novel and innovative examples of structure-based enzyme redesign are discussed below.

Reaching beyond the mutagenesis of residues near the active site, Reetz and coworkers explored the potential functional benefits of distal mutations through so-called induced allostery [17]. Arguing that amino acid substitutions in remote enzyme locations could translate into changes in structure and dynamics at the active site, the authors focused on the interface of the FAD and NADP-binding domains of the thermostable Bayer-Villiger monooxygenase from *Thermobifida fusca* [18,19]. Structure analysis identified two amino acid positions in a helical segment located near the two domains' hinge region which were subjected to saturation mutagenesis. Activity screening of these monooxygenase variants on a series of 2-substituted cyclohexanones suggested that substitutions in these distal positions not only increased catalytic activity but also significantly broadened the range of acceptable substrates. These performance changes were rationalized with the help of MD simulations which indicated changes in active site accessibility and backbone movement.

Simultaneously targeting several structural regions within an enzyme, an impressive example of multi-parameter optimization was the recently reported redesign of ω-transaminase from *Arthrobacter sp*. into an industrial biocatalyst for the production of the antidiabetic compound sitagliptin [20••]. Process specifications demanded a biocatalyst with high activity and >99.5% enantioselectivity but also required stability to harsh reaction conditions including elevated temperature (45–50°C), the presence of organic solvent (50% DMSO), and high substrate concentrations (200 g/L). Faced with the additional challenge that none of the screened wild type enzymes showed detectable activity for the bulky prositagliptin ketone substrate, Savile *et al*. applied a combination of structure-based enzyme redesign and directed evolution to gradually remodel the *Arthrobacter* enzyme into a biocatalyst that met the process specifications (Figure 2A). Molecular modeling and site-saturation mutagenesis were initially applied to expand the active site binding pocket and establish measurable turnover of prositagliptin ketone, setting the stage for subsequent rounds of directed evolution to optimize activity and performance under the desired reaction conditions. Of the 27 mutations found after eleven iterations, about half were located in the vicinity of the substrate binding site while a second hotspot was found at the protein-protein interface of the dimer.

If the substitution of a single amino acid is often disruptive to protein structure, it should come as no surprise that random chimeragenesis, the recombination of peptide fragments from parental proteins with low sequence identity, typically is highly degenerate. Computational predictive frameworks for guided chimeragenesis such as SCHEMA [21] have proven very effective to identify protein fragments that can be interchanged with minimal structural interference (Figure 2B). Following the successful application of the algorithm for engineering cytochrome P450s [22] and β-lactamases [23], the SCHEMA-guided recombination of three cellobiohydrolase genes to search for thermostable chimeras was reported more recently [24, 25]. Keeping the library size to a mere 48 chimeras (out of $3^8 = 6561$ possible combinations), Arnold and coworkers employed mathematical modeling to analyze the thermostabilities of

these initial candidates and score individual protein fragments for their contribution. Relying on the previously established additivity of a protein fragment's stabilizing effects, the design of a second set of chimeras was biased towards fragments with higher stability scores which led to the identification of highly active progeny with substantially elevated thermostability.

COMPUTATIONAL ENZYME REDESIGN

Advances in computational protein design algorithms have made in silico modeling a highly promising strategy for the tailoring of biocatalysts. Rather than relying on evolutionary information as a guide for sequence alterations and combinatorial library preparation at the bench, computational methods can effectively estimate the energetics of amino acid variations on the overall protein structure through the use of rotamer libraries and backbone reorganization, hence reducing experimental protein engineering to the evaluation of only a handful of rational designs. The capabilities of these predictive frameworks to reprogram the substrate specificity of enzymes was recently demonstrated by two groups.

Chen *et al*. used their *K** algorithm to switch the substrate specificity of the phenylalanine adenylation domain in the nonribosomal peptide synthetase enzyme gramicidin S synthetase A [26]. Targeting either Leu or one of the charged amino acids Arg, Lys, Glu or Asp as the new substrates, initial simulations concentrated on seven residues lining the binding pocket of the substrate side chain (Figure 3A). For Leu, mutations in two of the seven positions translated into a 20-fold higher k_{cat}/K_M for Leu and 30-fold decline in the catalytic efficiency for Phe, effectively switching the substrate specificity of the adenylation domain. Interestingly, the performance shift was largely due to changes in the apparent binding constants for the two substrates. Subsequently, mutations in other parts of the protein structure were explored to further boost enzyme activity. To identify sites of interest, the authors computationally assessed the protein-wide tolerance to mutations and selected variants with superior predicted protein stability, an effort that in the case of the Leu-specific adenylation domain identified three additional positions, each improving enzyme performance four-fold. Unfortunately, the redesign of the substrate binding pocket for the charged amino acids did not match the success with Leu. Although catalytic activity for the new substrates was detectable (in principle an infinite improvement over nondetectable activity in wild type enzyme), the overall performance was roughly four orders of magnitude lower than for the natural substrate.

An excellent demonstration that computational enzyme redesign does not need to be limited to individual amino acid substitutions but can be applied towards the replacement of entire loop regions was reported by Murphy *et al*. [27]. Drawing some inspiration from bacterial cytosine deaminases, the authors used the Rosetta Design algorithm [28] for remodeling of a loop region in the active site of human guanine deaminase (hGDA) with the goal to create a cytosine-specific hGDA variant for potential application in suicide gene therapy. In silico modeling of simultaneous variations in loop length and amino acid composition identified a new sequence, glycine-rich and two amino acids shorter than the native loop, which boosted activity for ammelide (a design-intermediate for cytosine) by 100-fold while diminishing guanine deamination by four orders of magnitude (Figure 3B).

COMPUTATIONAL DE NOVO ENZYME DESIGN

In the spirit of Richard Feynman's quote "what I cannot create, I do not understand", the ultimate enzyme engineering challenge is not so much about engineering but rational design. Instead of remodeling an existing enzyme, the creation of biocatalysts from scratch not only offers potential practical benefits in that it empowers scientists and engineers to build synthetic enzymes for any chemical transformation, it also presents a testing ground for our fundamental understanding of the intricacies of protein structure and function.

Synthetic biocatalysts with novel function

While the two synthetic biocatalysts for the Kemp elimination [29] and retroaldol reaction [30] are well established in the literature, the most recent de novo design of a Diels-Alderase (DA) by Baker and collaborators represents another milestone in enzyme engineering [31••]. The latest example expands the concept of computational enzyme design to carbon-carbon bond formation between two separate substrates, catalyzing an intermolecular Diels-Alder reaction which requires the concomitant binding of two substrates and their precise geometric orientation for productive cycloaddition. The DA design followed a similar strategy as was employed for the two previous biocatalysts, initially using QM simulations to create an comprehensive theozyme library (~10^{19} variants) which was fitted into a library of protein scaffolds by the RosettaMatch software (Figure 4). Approximately 10^6 feasible theozyme-scaffold pairs were identified which, after further optimization in RosettaDesign, resulted in 84 designs that were experimentally evaluated and two candidates with detectable DA activity. Following some additional fine-tuning of residues lining the active site, the catalytic efficiency of these two synthetic DAs matched the performance of catalytic antibodies raised for Diels-Alder cycloadditions. Furthermore, the novel enzymes catalyze multiple turnover cycloadditions and exhibited stereoselectivity and substrate specificity, all hallmarks of true enzymes.

Beyond the initial design of novel synthetic biocatalysts, studies to explore mechanistic details of these designer enzymes and address shortfalls in the original designs have highlighted some of the challenges and exciting opportunities for future de novo enzyme design. For example, the detailed kinetic and mutational analysis of a top-scoring retroaldol designer enzyme by Lassila *et al*. uncovered that the catalyst's 10^5-fold rate acceleration could largely be attributed to favorable hydrophobic interactions with the naphthyl moiety of the substrate [32•]. In contrast, the postulated catalytic Lys residue in the active site contributed only moderately (10-fold) while mutations in the binding pocket of an explicit water molecule, believed to participate in proton transfer, had little to no effect on catalysis.

Separately, Ruscio *et al*. investigated the proton-transfer reaction of an alternate retroaldolase design [33]. The molecular dynamics simulations suggested that temporal conformational fluctuations in the protein interfere significantly with the optimized active site model, hence compromising the catalytic efficiency of the designer enzyme. To overcome these limitations, the authors proposed the integration of NMR data as templates for protein design algorithms, yet the idea's implementation might be complicated by inherent software biases for crystallographic data [34]. Along the same lines, the exclusion of conformational heterogeneity in proteins and conformational changes associated with catalysis in current enzyme designs could also explain the similar catalytic performance of de novo enzymes and their corresponding catalytic antibodies [35,36]. Focusing largely on the chemical step in the catalytic cycle, the reliance on transition state models in the form of actual analogs or simulated theozymes was argued to not properly account for events such as substrate binding, product release, and conformational changes, hence caping the performance of present designs.

Directed evolution represents an obvious strategy to improve existing designer enzymes and to potentially "break" the performance cap. In the case of a Kemp eliminase, seven rounds of random mutagenesis and DNA shuffling improved the catalytic efficiency by two orders of magnitude [37]. The functional gains could be linked to a more optimal organization of side chains in the active site and reduced thermostability which seems to benefit catalysis due to the increased flexibility in the active site region. Interestingly, attempts to further improve catalytic performance through additional rounds of directed evolution were however unsuccessful.

Recreating existing enzymes by de novo design

Lu and coworkers applied de novo protein design to tackle a very different problem in biocatalysis [38]. Rather than seeking synthetic enzymes for new catalytic function, the researchers employed sequence-homology modeling and molecular dynamics simulations to assemble the presumptive active site metal complex of nitric oxide reductase (NOR) in whale-sperm myoglobin.

NORs are key players in bacterial denitrification and mammalian signal transduction pathways, yet detailed structural and mechanistic studies have been impeded by difficulties to obtain enzyme in sufficient quantity. Assembly of the proposed catalytic site, consisting of a heme and a putative Fe_B site, in the myoglobin scaffold established a robust model system to explore the spectroscopic properties and to validate the catalytic function of the hypothetical metal complex. Taking advantage of the existing heme prosthetic group in myoglobin, the designers focused on the computer-guided remodeling of a hydrophobic pocket near the heme cofactor to establish the new non-heme Fe^{2+} binding site (Figure 5). The redesigned myoglobin showed the predicted spectroscopic changes upon Fe^{2+} binding and crystallographic data confirmed the accuracy of the calculated structure model. More importantly, the nitric oxide reduction activity of the artificial metalloenzyme clearly rose above background. The design of a catalytically active NOR model has created a powerful predictive framework for future mechanistic studies as demonstrated in more recent experiments by the same group examining the presumptive functional role of a second, conserved Glu near the metal centers and evaluating the effects of metal ion substitution in the non-heme metal binding site [39].

CONCLUDING REMARKS

The methodological advances in semi-rational enzyme engineering and de novo enzyme design in recent years provide researchers with powerful and effective new strategies to manipulate biocatalysts. As the examples in this review demonstrate, the integration of sequence and structure-based approaches in library preparation has already proven a potent guide to enzyme redesign. In the case of computational de novo and redesign methods, current models still tend to lag behind laboratory-evolved variants in catalytic performance, yet these functional shortfalls can be rationalized by small (but costly performance-wise) imperfections in active site topology and a disregard for conformational changes. While some experimental optimization is possible by directed evolution, refinements in the design algorithm will likely yield further improvements in the accuracy of structure predictions and hence superior catalytic performance. Separately, the integration of protein dynamics in future simulations might deliver additional functional enhancement and at the same time provides an excellent testing ground for assessing its relevance to biocatalysis. In parallel with the fine-tuning of the predictive framework, comprehensive biochemical and biophysical studies will be invaluable to experimentally evaluate the functional contributions of individual design features which in turn provides critical feedback for improvements in future designs.

In summary, computational protein design has fundamentally changed the way protein engineers can manipulate biomacromolecules, yet it won't replace directed evolution as the method of choice for protein engineering. Instead, researchers should recognize the complementarity of the two strategies and embrace their integration for the more effective manipulation of enzymes. The utilization of in silico methods permits the generation of predictive frameworks for hypothesis-driven protein engineering which can significantly reduce the complexity of the system and translate into smaller, more focused and functionally-rich libraries. In turn, directed evolution still represents the most effective strategy to identify the top-performing candidates in these focused libraries, yet the new design strategies and technical advances have initiated a departure from the traditional protocols. Whole-gene library synthesis is replacing shuffling and mutagenesis protocols for library preparation while highly

specific low-throughput screening assays are increasingly applied in place of monumental screening and selection efforts of millions of candidates. Together, these exciting developments are poised to take protein engineering beyond directed evolution and towards practical, more efficient strategies for tailoring biocatalysts.

Acknowledgments

This work was supported in part by the National Institutes of Health (GM69958), the US National Science Foundation (CBET-0730312) and a grant from the Petroleum Research Fund by the American Chemical Society (PRF 47135-AC1). Thanks also to the members of the Lutz lab for helpful comments on the manuscript.

Abbreviations

MD	molecular dynamics
QM	quantum mechanics
MSA	multiple sequence alignment
PDB	protein database

References

1. Lutz, S.; Bornscheuer, U., editors. The Protein Engineering Handbook. Weinheim: Wiley-VCH; 2009.
2. Wong TS, Zhurina D, Schwaneberg U. The diversity challenge in directed protein evolution. Comb Chem High Throughput Screen 2006;9:271–288. [PubMed: 16724918]
3. Damborsky J, Brezovsky J. Computational tools for designing and engineering biocatalysts. Curr Opin Chem Biol 2009;13:26–34. [PubMed: 19297237]
4. Pei J. Multiple protein sequence alignment. Curr Opin Struct Biol 2008;18:382–386. [PubMed: 18485694]
5. Pavelka A, Chovancova E, Damborsky J. HotSpot Wizard: a web server for identification of hot spots in protein engineering. Nucleic Acids Res 2009;37:W376–383. [PubMed: 19465397]
6•. Pavlova M, Klvana M, Prokop Z, Chaloupkova R, Banas P, Otyepka M, Wade RC, Tsuda M, Nagata Y, Damborsky J. Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. Nat Chem Biol 2009;5:727–733. [PubMed: 19701186] Molecular dynamics simulation and structure-based enzyme design identified key functional residues in the enzyme's active site access tunnel. Substitutions in these positions resulted in variants with superior activity and functional gains were shown to originate from changes in solvent accessibility of the active site
7. Kuipers RK, Joosten HJ, van Berkel WJ, Leferink NG, Rooijen E, Ittmann E, van Zimmeren F, Jochens H, Bornscheuer U, Vriend G, et al. 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. Proteins 2010;78:2101–2113. [PubMed: 20455266]
8. Kuipers RK, Joosten HJ, Verwiel E, Paans S, Akerboom J, van der Oost J, Leferink NG, van Berkel WJ, Vriend G, Schaap PJ. Correlated mutation analyses on super-family alignments reveal functionally important residues. Proteins 2009;76:608–616. [PubMed: 19274741]
9. Joosten HJ, Han Y, Niu W, Vervoort J, Dunaway-Mariano D, Schaap PJ. Identification of fungal oxaloacetate hydrolyase within the isocitrate lyase/PEP mutase enzyme superfamily using a sequence marker-based method. Proteins 2008;70:157–166. [PubMed: 17654546]
10•. Jochens H, Bornscheuer UT. Natural diversity to guide focused directed evolution. ChemBioChem. 2010 in press. Validation of higher functional content in designer enzyme libraries derived from consensus sequence over random or non-consensus sequence pools
11. Bosma T, Damborsky J, Stucki G, Janssen DB. Biodegradation of 1,2,3-trichloropropane through directed evolution and heterologous expression of a haloalkane dehalogenase gene. Appl Environ Microbiol 2002;68:3582–3587. [PubMed: 12089046]

12. Banas P, Otyepka M, Jerabek P, Petrek M, Damborsky J. Mechanism of enhanced conversion of 1,2,3-trichloropropane by mutant haloalkane dehalogenase revealed by molecular modeling. J Comput Aided Mol Des 2006;20:375–383. [PubMed: 17016745]

13. Ehren J, Govindarajan S, Moron B, Minshull J, Khosla C. Protein engineering of improved prolyl endopeptidases for celiac sprue therapy. Protein Eng Des Sel 2008;21:699–707. [PubMed: 18836204]

14. Liao J, Warmuth MK, Govindarajan S, Ness JE, Wang RP, Gustafsson C, Minshull J. Engineering proteinase K using machine learning and synthetic genes. BMC Biotechnol 2007;7:16. [PubMed: 17386103]

15. Minshull J, Govindarajan S, Cox T, Ness JE, Gustafsson C. Engineered protein function by selective amino acid diversification. Methods 2004;32:416–427. [PubMed: 15003604]

16. Chen F, Gaucher EA, Leal NA, Hutter D, Havemann SA, Govindarajan S, Ortlund EA, Benner SA. Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. Proc Natl Acad Sci U S A 2010;107:1948–1953. [PubMed: 20080675]

17. Wu S, Acevedo JP, Reetz MT. Induced allostery in the directed evolution of an enantioselective Baeyer-Villiger monooxygenase. Proc Natl Acad Sci U S A 2010;107:2775–2780. [PubMed: 20133612]

18. Malito E, Alfieri A, Fraaije MW, Mattevi A. Crystal structure of a Baeyer-Villiger monooxygenase. Proc Natl Acad Sci U S A 2004;101:13157–13162. [PubMed: 15328411]

19. Fraaije MW, Wu J, Heuts DP, van Hellemond EW, Spelberg JH, Janssen DB. Discovery of a thermostable Baeyer-Villiger monooxygenase by genome mining. Appl Microbiol Biotechnol 2005;66:393–400. [PubMed: 15599520]

20••. Savile CK, Janey JM, Mundorff EC, Moore JC, Tam S, Jarvis WR, Colbeck JC, Krebber A, Fleitz FJ, Brands J, et al. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. Science 2010;329:305–309. [PubMed: 20558668] The combination of structure-guided enzyme design and directed evolution was used to reprogram the substrate specificity and to stabilize a omega-transaminase for application as an industrial biocatalyst

21. Otey CR, Landwehr M, Endelman JB, Hiraga K, Bloom JD, Arnold FH. Structure-guided recombination creates an artificial family of cytochromes P450. PLoS Biol 2006;4:e112. [PubMed: 16594730]

22. Li Y, Drummond DA, Sawayama AM, Snow CD, Bloom JD, Arnold FH. A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. Nat Biotechnol 2007;25:1051–1056. [PubMed: 17721510]

23. Meyer MM, Hochrein L, Arnold FH. Structure-guided SCHEMA recombination of distantly related beta-lactamases. Protein Eng Des Sel 2006;19:563–570. [PubMed: 17090554]

24. Heinzelman P, Snow CD, Wu I, Nguyen C, Villalobos A, Govindarajan S, Minshull J, Arnold FH. A family of thermostable fungal cellulases created by structure-guided recombination. Proc Natl Acad Sci U S A 2009;106:5610–5615. [PubMed: 19307582]

25. Heinzelman P, Snow CD, Smith MA, Yu XL, Kannan A, Boulware K, Villalobos A, Govindarajan S, Minshull J, Arnold FH. SCHEMA Recombination of a Fungal Cellulase Uncovers a Single Mutation That Contributes Markedly to Stability. J Biol Chem 2009;284:26229–26233. [PubMed: 19625252]

26. Chen CY, Georgiev I, Anderson AC, Donald BR. Computational structure-based redesign of enzyme activity. Proc Natl Acad Sci U S A 2009;106:3764–3769. [PubMed: 19228942]

27. Murphy PM, Bolduc JM, Gallaher JL, Stoddard BL, Baker D. Alteration of enzyme specificity by computational loop remodeling and design. Proc Natl Acad Sci U S A 2009;106:9215–9220. [PubMed: 19470646]

28. Das R, Baker D. Macromolecular modeling with rosetta. Annu Rev Biochem 2008;77:363–382. [PubMed: 18410248]

29. 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]

30. 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 retroaldol enzymes. Science 2008;319:1387–1391. [PubMed: 18323453]

31••. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, StClair 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] Describes the computational de novo design of a functional biocatalyst for the intermolecular Diels-Alder reaction

32•. 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] A meticulous functional analysis of designer enzymes to quantify the contributions of individual design features to overall catalysis and provide critical feedback for future enzyme designs

33. Ruscio JZ, Kohn JE, Ball KA, Head-Gordon T. The influence of protein dynamics on the success of computational enzyme design. J Am Chem Soc 2009;131:14111–14115. [PubMed: 19788332]

34. Schneider M, Fu X, Keating AE. X-ray vs. NMR structures as templates for computational protein design. Proteins 2009;77:97–110. [PubMed: 19422060]

35. Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol 2009;5:789–796. [PubMed: 19841628]

36. Marti S, Andres J, Moliner V, Silla E, Tunon I, Bertran J. Computational design of biological catalysts. Chem Soc Rev 2008;37:2634–2643. [PubMed: 19020677]

37. 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]

38. Yeung N, Lin YW, Gao YG, Zhao X, Russell BS, Lei L, Miner KD, Robinson H, Lu Y. Rational design of a structural and functional nitric oxide reductase. Nature 2009;462:1079–1082. [PubMed: 19940850]

39. Lin YW, Yeung N, Gao YG, Miner KD, Tian S, Robinson H, Lu Y. Roles of glutamates and metal ions in a rationally designed nitric oxide reductase based on myoglobin. Proc Natl Acad Sci U S A 2010;107:8581–8586. [PubMed: 20421510]

40. Lutz S. Biochemistry - Reengineering enzymes. Science 2010;329:285–287. [PubMed: 20647454]

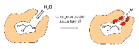

Figure 1.
Mutations at or near the substrate access tunnels can restrict water accessibility in haloalkane dehalogenase which benefits catalysis by shielding the substrate complex from bulk solvent.

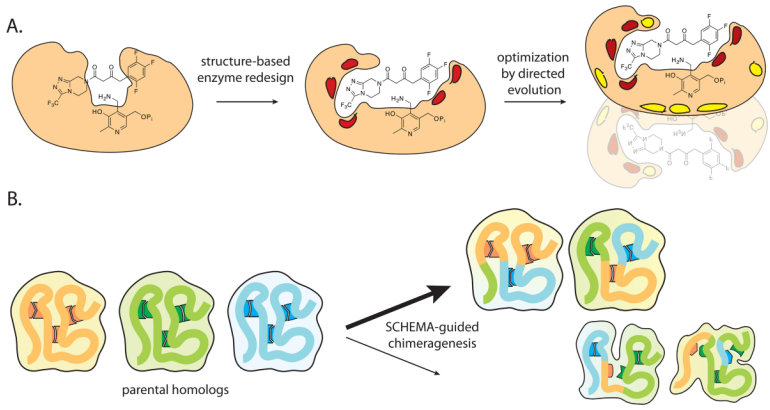

Figure 2.

Structure-based enzyme redesign. A) Engineering of an ω-transaminase for the industrial production of silagliptin (figure adapted from [40]). Mutations in the initial redesign (shown in red) concentrate on the active site while changes by directed evolution (in yellow) are skewed towards dimer stabilization. B) SCHEMA analysis of three parental cellobiohydrolases optimizes generation of functional chimera by minimizing recombination of protein fragments with incompatible topologies.

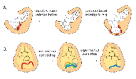

Figure 3.
Computational enzyme redesign. A) Mutations (shown in red) change the substrate specificity of phenylalanine adenylation domain to amino acids with small hydrophobic (leucine) and charged side chains (arginine). B) Engineering of loop region (red) in the active site of human guanosine deaminase creates an ammelide-specific deaminase.

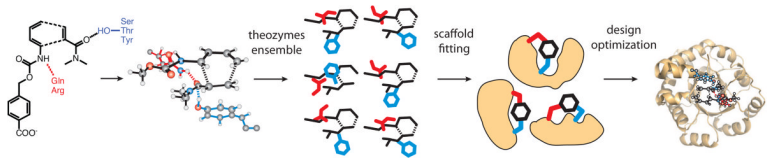

Figure 4.

De novo design of a Diels-Alderase. An ensemble of theozymes (computational models of the reaction's presumptive transition state including key amino acids) were matched against a library of protein scaffolds to identify suitable combinations. Subsequently, design solutions were refined in silico and tested experimentally (figure adapted from [40]).

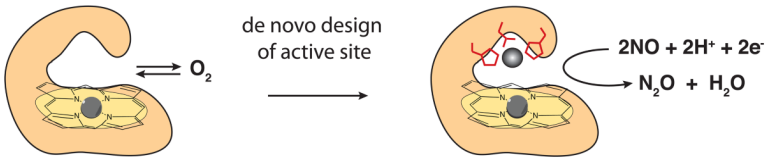

Figure 5.
Converting myoglobin into nitric oxide reductase. Remodeling of a hydrophobic pocket into a ferrous binding site transforms the oxygen-binding protein into a biocatalyst.

TABLE 1

Target	Project goal	Methodology			evaluated	Conclusions / Comments	Ref.
		Design (active site)	*Design (distal regions)*	*Experimental*	library size		
Enzyme redesign: sequence-based							
Thermus aquaticus DNA polymerase	Substrate specificity for unnatural nucleoside triphosphates	Reconstructed evolutionary adaptive path (REAP) analysis	none	Whole-gene synthesis	93	Single amino acid substitution required for efficient incorporation of unnatural NTP.	[16]
Pseudomonas fluorescens esterase	Improve enantioselectivity	3DM analysis at 4 specific amino acid positions	none	Site-saturation mutagenesis	~500	Yielded variants with improved activity (200-fold)and enantioselectivity (20-fold).	[10•]
Sphingomonas capsulata prolyl endopeptidase	Improve activity and pH/ protease stability	Hot-spot selection based on position & substitution scores from multiple sequence alignment. Machine-learning algorithm for library analysis		Whole-gene synthesis	91 (two rounds)	Variants with substitutions in 5 positions raise activity by 20% and improve protease resistance by 200-fold.	[13]
Rhodococcus rhodochrous haloalkane dehalogenase	Improve catalytic activity	MD simulations to identify mutational hotspots in access tunnels to active site		Site-directed and site-saturation mutagenesis	~2500	32-fold improved activity by restricting water access to active site.	[6•]
Enzyme redesign: structure-based							
Class II cellobiohydrolases (*Humicola insolens*,	Increase thermostability	SCHEMA structure-guided recombination of protein fragments from 3 CBH II cellulases		Whole-gene synthesis	48	Cellulase chimeragenesis raises operating temperature by up to 15°C. Quantitative predictions of thermostability.	[24]
Thermobifida fusca phenylacetone monooxygenase	Substrate specificity through induced allostery	none	X-ray structure analysis of hinge region and domain interface	Site-saturation mutagenesis	~400	Enhanced substrate promiscuity due to distal mutations in 2 adjacent positions near NADPH/ FAD interface.	[17]
Arthrobacter sp. omega-transaminase	Substrate specificity, thermostability and tolerance for organic solvents	MOE modeling (molecular modeling & structure-based analysis)	Directed evolution	primary: site-saturation mutagenesis (round 1&2) secondary: random mutagenesis, in vitro recombination (rounds 3–11)	~36,000 (over 11 rounds)	Redesigned enzyme meets project objectives for application in industrial process.	[20••]
Enzyme redesign: computational							
gramicidine S synthetase A phenylalanine adenylation domain	Substrate specificity from Phe to Leu, Arg, Lys, Glu, or Asp	K* algorithm (mutagenesis with rotamer library, flexible backbone	Computational (SCMF entropy-based method)	Site-directed mutagenesis	<10	600-fold specificity shift for Phe→Leu due to changes in K_M-values. Designs for charged amino	[26]

Target	Project goal	Methodology			evaluated	Conclusions / Comments	Ref.
		Design (active site)	***Design (distal regions)***	***Experimental***	**library size**		
		and dynamic ligand substrate)				acid substrates yield more promiscuous enzymes.	
human guanine deaminase	Substrate specificity for ammelide/cytosine	RosettaDesign (variation of active site loop length & composition)	none	PCR overlap assembly, site-directed mutagenesis	<10	$>10^6$ specificity change with moderate catalytic efficiency	[27]
De novo design							
Diels-Alderase	Biocatalyst for intermolecular Diels-Alder reaction	QM/MM simulations, RosettaMatch and Design software	none	Site-directed mutagenesis	<100	Stereoselective Diels-Alderase whose functional performance matches catalytic antibodies.	[31••]
Nitric oxide reductase	Reconstitute active site of NOR in myoglobin	VMD software (molecular modeling)	none	Site-directed mutagenesis	<10	Functional model of NOR	[38]