TRIBUTE TO FOUNDERS: FRANCES ARNOLD.
BIOMOLECULAR ENGINEERING, BIOENGINEERING,
BIOCHEMICALS, BIOFUELS, AND FOOD

From directed evolution to computational enzyme engineering—A review

Ratul Chowdhury | Costas D. Maranas

Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania

Correspondence
Costas D. Maranas, Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802.
Email: costas@psu.edu

Funding information
Division of Chemical, Bioengineering, Environmental, and Transport Systems, Grant/Award Number: CBET-1703274

Abstract
Nature relies on a wide range of enzymes with specific biocatalytic roles to carry out much of the chemistry needed to sustain life. Enzymes catalyze the interconversion of a vast array of molecules with high specificity—from molecular nitrogen fixation to the synthesis of highly specialized hormones and quorum-sensing molecules. Ever increasing emphasis on renewable sources for energy and waste minimization has turned enzymes into key industrial workhorses for targeted chemical conversions. Modern enzymology is central to not only food and beverage manufacturing processes but also finds relevance in countless consumer product formulations such as proteolytic enzymes in detergents, amylases for excess bleach removal from textiles, proteases in meat tenderization, and lactoperoxidases in dairy products. Herein, we present an overview of enzyme science and engineering milestones and the emergence of directed evolution of enzymes for which the 2018 Nobel Prize in Chemistry was awarded to Dr. Frances Arnold.

KEYWORDS
biocatalysis, computational protein design, directed evolution, enzyme engineering, mutagenesis

1 | BACKGROUND

1.1 | Background of enzymes, enzymology, and enzyme design

Biological systems are masterful chemists that build complex molecules and systems from simple precursor compounds. At the heart of this complex machinery are enzymes that account for ~4% of proteins. The use of biocatalysts by humanity which emerged as an accidental by-product of gathering wild grain dates at least back to the ancient Egyptians (circa 10,000 BC) who used fermentation for bread-making and brewing purposes. However, it was not until the 19th century that fermentation was recognized as carried out by living cells. In 1835, Swedish chemist Jacob Berzelius used the term "proteins" to describe similar molecules extracted from egg-whites, blood, serum, fibrin, and wheat gluten which all had atomic ratios of C:H:N:O:S to be approximately 1:1.58:0.28:0.30:0.01 (experiments performed by Johannes Mulder) and reported that some of the proteins are catalytic. It was much later in 1878, that German physiologist Wilhelm Kühne coined the term "enzymes." Figure 1 shows a chronological compilation of 70 key events in the history of enzyme engineering starting from the 1830s up to 2018 when the Nobel Prize in chemistry was co-awarded for the "directed evolution of enzymes" and "phage display." Edward Buchner in 1897 isolated an enzyme complex which he called zymase from cell-free yeast extracts and successfully demonstrated that it can catalyze the breakdown sugars in alcoholic fermentation. During the same time, German chemist, Emil Fischer postulated the "lock and key" hypothesis for enzyme activity where the substrate (key) was thought to rigidly fit into the complementary groove of the enzyme (lock). Figure 2 shows the lysozyme binding pocket with a peptidoglycan substrate occupying the binding pocket. However, most of enzymatic catalysis could not be adequately
explained by the rigid enzyme model till 1958 when Koshland laid out the "induced fit" theory for enzyme substrate action. The three principal tenets of "induced fit" theory as explained in the article were:

(a) the precise orientation of catalytic groups is required for enzyme action,
(b) the substrate causes an appreciable change in the three-dimensional relationship of the amino acids at the active site, and
(c) the changes in the protein structure caused by the substrate brings the catalytic groups into the proper alignment, whereas a non-substrate does not." Figure 3 shows the change in hexokinase structure (from closed to open) necessary for product release. In 1903, French chemist Victor Henri derived a functional form of enzyme kinetics based on geometry and composition of similar proteins. However, a simplified and more celebrated version of the equation that equated the rate of the reaction with the rate at which the concentration of various species involved in the reaction was formalized by Michaelis and Menten in 1913. About a decade later during 1930s, Sumner, Northrop, and Stanley independently crystallized urease, pepsin, and nucleoproteins responsible for tobacco-mosaic virus activity, respectively, for which they shared the Nobel Prize in chemistry in 1946. These structural studies were soon complemented with methods for discerning the sequence of short peptide chains in 1950 by Pehr Edman and in 1952 Frederick Sanger reported the complete amino acid sequence of polypeptide chains A and B of bovine insulin building on the work of polypeptides.
started by Charles Chibnall. Concurrently, William Stein and Stanford Moore collaborated on developing an analytical procedure to determine the amino acid content of any protein at The Rockefeller Institute of Medical Research. Stein and Moore used potato starch in a column for fractionation of proteins from peptides along with simultaneous counting of the amino acids. Subsequently, they followed up with a better and faster quantification approach of amino acids in peptides. The next two decades saw isolation, purification, and characterization of various enzymes ranging from myoglobin from sperm whales to high resolution lysozyme structures from egg white using X-ray crystallography. It was in 1978 when directed evolution revolutionized the search for better enzymes.

Mutation followed by natural selection was established by Darwin's *On the Origin of Species* in 1859 as the organizing principle in biology. However, for thousands of years before humans unknowingly exploited this process in selective breeding and domestication. It was only in the late 1970s that evolution was brought inside the laboratory with the specific objective of discovering microbial phenotypes for better utilization of desired carbon substrates. Lerner et al designed a xylitol utilization phenotype of *Aerobacter aerogenes* in 1964. In 1967 Spiegelman et al performed in vitro reconstitution of RNA templates with pure RNA replicase to study the effect of selective pressures for several generations in the famous "Spiegelman's monster" experiments. Inspired by these, Francis and Hansche performed "directed evolution" in yeast and achieved 30% higher orthophosphate activity with a single mutation but with a growth rate trade-off of 83%. This was soon followed by a more comprehensive demonstration of directed evolution by Barry Hall where up to four mutations in the β-galactosidase coding region in *Escherichia coli* cultured with lactose as the sole carbon source yielded phenotypes spanning a wide range of growth rates. Within a decade, Eigen and Gardiner proposed a cyclic "evolutionary machine" comprised of genetic mutations, amplification and selection to produce stable mutant proteins in vitro. The subsequent development of error–prone polymerase chain reaction (PCR) for random mutagenesis enabled generation of large-scale mutant libraries with >1010 designs and has been a cornerstone in the history of enzyme engineering.

### Methods for directed evolution

Enzymes (and proteins in general) are modular biopolymers composed of 20 canonical amino acid monomers as encoded by their cognate nucleotide sequences (genes). They have the potential to evolve through changes in their amino acid sequence. This evolvability has been exploited to explore the combinatorial sequence space for catalyzing reactions with improved specificity, regioselectivity, and stereoselectivity. Thus, directed evolution of enzymes and binding proteins is a synthetic procedure relying on molecular insights, which emulates the natural evolution process in the laboratory at an expedited rate. The procedure commits to intended variation of protein sequences with prescribed randomness of amino acid choices. This is
further coupled to engineered screening and selection strategies. In other words, directed evolution involves iterative identification of a starting protein, diversification of its coding gene sequence, expression, and subsequent functional screening until an acceptable level of enzymatic activity, binding affinity, or specificity is accomplished.

Sampling the entire combinatorially explosive mutational landscape for any protein is impossible as complete randomization of a mere pentapeptide would yield ~10^{13} unique amino acid sequences. Gene diversification approaches are thus designed to perform an optimal sparse sampling of the multidimensional sequence space, with the objective of ascending in the landscape of desired phenotype by accruing beneficial mutations. Several gene diversification methods for directed evolution have been proposed over the last two decades. These strategies typically integrate random mutagenesis, focused mutagenesis, and homologous recombination.

1.3 Random mutagenesis

Random mutagenesis starts by obtaining a library of point mutants from a single parent sequence and transforming the library into a strain to express the variant proteins. A high-throughput screen for the desired phenotype then identifies the successful candidates. Error-prone PCR (epPCR), first described by Goeddel et al.\textsuperscript{33} utilized the low-fidelity of DNA polymerases to make point mutations during amplification of the gene that codes for the protein of interest. Gheraldi et al.\textsuperscript{34} and Kunkel et al.\textsuperscript{35} were able to enhance the rate of mutation (from 10^{-10} to 10^{-3}) by adding mutagenic dNTP analogues or increasing magnesium concentrations in the epPCR setup. Additional screening for properly expressed proteins by fusing the target gene with a green-fluorescent protein reporter was soon demonstrated by Tawfik et al.\textsuperscript{36} A modified epPCR was developed by Joyce et al.\textsuperscript{37} that used a combination of Taq polymerase, 0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP, and 1 mM of dTTP, higher MgCl\textsubscript{2}, and 0.5 mM MnCl\textsubscript{2} to reduce polymerase fidelity without affecting gene amplification and alleviated the strong bias toward A\textright-G, and T\textright-C transitions as faced by Goeddel and co-workers. Arnold and co-workers have documented several successes using random mutagenesis including introducing activity toward a wide range of native-like substrates in cytochrome P450\textsubscript{57} and exploring novel carotenoid biosynthesis routes.\textsuperscript{38}

1.4 Focused mutagenesis

The probability of identifying active redesigns which emerge from synergism of simultaneous point mutations (which are themselves marginally useful) is very low using random mutagenesis as the number of possible unique sequences increases exponentially with the number of randomized sites. To this end, focused mutagenesis uses phylogenetic analyses of homologous proteins to identify specific amino acid substitutions that are likely to improve substrate binding or catalysis. A mutagenic oligonucleotide cassette\textsuperscript{39} containing degenerate codons for a targeted amino acid change is inserted\textsuperscript{40} into a vector plasmid for expression of a desired enzyme variant. Parra et al.\textsuperscript{41} fed focused mutagenesis library of xylanase to epPCR to identify 12 more thermotable variants with the best mutant showing a 4.3°C increase in melting temperature.

1.5 Homologous recombination

An alternate strategy to access beneficial combinations of mutations is achieved using homologous recombination. This is a mimic of the natural process of biological evolution. One of the early approaches, DNA shuffling, involved a DNase-mediated fragmentation of a target gene, followed by random re-stitching using a PCR setup. Monticello et al.\textsuperscript{42} replaced the random priming of DNA fragments by a sophisticated random chimeragenesis technique (RACHITT). They were able to achieve several folds of higher recombination than any other method in a dibenzothiophene monooxygenase gene. The expressed proteins not only exhibited higher than wild-type activity, but also showed 20-fold higher affinity for several hydrophobic nonnatural substrates. Arnold et al.\textsuperscript{43} also reported an optimized DNA shuffling workflow to control the point mutagenesis rate to as low as 0.05% by adding Mn\textsuperscript{2+} and Mg\textsuperscript{2+} ions during DNase I digestion of the gene and appropriate choice of DNA polymerase to effect high-fidelity recombination. A number of modeling frameworks were developed\textsuperscript{44} for estimating the occurrence of mutations in error-prone PCR after multiple generations\textsuperscript{45} and the location of crossovers in directed evolution experiments.\textsuperscript{46,47}

Recent trends in directed evolution have seen attempts and successes at improving proteins with the biological proviso of still being relevant to the metabolic pathways they belong to, thus creating novel whole cell chemical factories for synthesis of value-added chemicals.\textsuperscript{48-50} More recently, biochemists have aimed at dialing in novel functionalities in enzymatic proteins which are not seen in nature.\textsuperscript{51,52} A decade old review by Toscano et al.\textsuperscript{53} on active site redesign strategies provide considerable insight toward function-driven enzyme redesign.

1.6 Methods for computational protein design

Computational methods provide the means to screen in silico many enzyme redesign alternatives thus focusing the number of variants to be tested experimentally. Existing approaches generally use biophysics-inspired or statistical fitness functions to screen design alternatives in terms of conservation (or enhancement) of desired interactions and absence of aberrant ones. There is an ever expanding literature of scoring function\textsuperscript{54,55} and combinatorial search algorithms\textsuperscript{56} devoted toward the efficient traversal of combinatorial space of residue alternatives. Software tools that integrate all these tasks include RosettaDesign,\textsuperscript{57} Osprey,\textsuperscript{58} Tinker,\textsuperscript{59} TransCent,\textsuperscript{60} and IPRO.\textsuperscript{61} The difficulty and success rate in computational design depends on how ambitious the enzyme redesign goal is. For example, attempts to switch cofactor or substrate
specificity of a well characterized enzyme have been met with many successes, however, efforts to improve the catalytic affinity of a native enzyme toward its preferred substrate is much more difficult with only a few success stories. In addition, efforts at introducing a novel enzymatic activity are also very difficult. Nevertheless, there has been a lot of exciting, industrially relevant research focused on generation of stable humanized immunoproteins with biopharmaceutical relevance and enzymes with enhanced turnover, altered substrate-, and stereo-selectivity in the past one and a half decade.

1.7 Statistical protein design approaches

Existing protein structures already contain a vast amount of information that correlate amino acid sequence to structure. A database-driven energy function reliant on the frequency of certain structural arrangement of amino acid backbones and side chains have been used to create a “knowledge-based potential.” DrugScore was used to predict and score ligand conformations at the active site of an enzyme using entropic contributions and implicit solvation upon learning from 159 experimentally resolved enzyme-ligand complexes. However, the lack of hydrogen atoms failed to capture the effect of protonation states, and also undermined electrostatic contributions to a great extent. Buchete et al developed statistical potentials using orientations of dihedral potentials without enhancing computing time. Several independently developed force fields have been developed till date—such as, Amber, CHARMM, OPLS, GROMOS, CHARMM and Rosetta all atom, while OPLS has both versions. Mackerel et al provides a detailed discussion on the development of empirical force fields.

1.9 Biophysical protein design tools

Biophysical protein design tools include computing enthalpic energy contributions of covalently bonded amino acids along the polypeptide backbone of a protein, and pairwise non-covalent interactions (van der Waals, electrostatics, and solvent effects) between atoms in proximity to each other. These force-field based energy scores are used in iterative or random-substitution computational workflows to make design choices toward identifying stable enzyme variants with improved ligand affinity, altered cofactor specificities, and other biochemical objectives. Several tools using either full atomistic or coarse-grained representations of proteins have been developed over the last two decades. Go and Taketomi employed non-transferable potentials tailored to the native structure of a protein by evaluating the partial contributions of long-range and short-range forces at play throughout the molecule. Any variant to the native protein (referred to as “Go-models”) would attain its lowest energy score when the corresponding inter-residue root mean square deviation with the native structure is minimum. Even though Go-proteins cannot explore novel folds, they have had high success rate in identifying functional variants that fold as only an extremely restricted set of positions permit substitutions to similar-to-native side chain properties (charge and size). The protein-module of Martini coarse-grained force-field was developed for predicting peptide conformations in lipid-bilayers. This was an extension to the lipid-exclusive Martini-force field. Using dioleylphosphatidylcholine bilayer and a series of pentapeptides as a model system, the potential of mean force for each amino acid was evaluated as a function of its distance from the center of the lipid region of the bilayer. These values were used as precedents to estimate the geometry of any new transmembrane protein.

1.8 Force fields for computational protein design

Force fields are used to compute interaction and overall stability energy scores of protein-ligand complexes or individual proteins. These energy terms (or scores) represent side chain and backbone geometries, protonation states, and effect of solvents and only enthalpic contributions are factored (not protein entropy). Force-field calculations help to assess enzyme substrate affinities and modeling of side chains. The most popular force field parameters (bond spring constant, bond angles, dihedrals, improper dihedrals, partial charges) are computed using ab initio quantum mechanical and molecular mechanics calculations. Knowledge-based force field like Rosetta uses extra potential energy terms obtained after refitting of statistical and experimental knowledge-based data. Unlike statistical knowledge-based potentials, these empirical force fields are capable of capturing actual forces between atoms (electrostatics, van der Waals, and solvent contribution). Several independently developed force fields have been developed till date—such as, Amber, CHARMM, OPLS, GROMOS, CHARMM and Rosetta all atom, while OPLS has both versions. Mackerel et al provides a detailed discussion on the development of empirical force fields.
whose overall geometry is dependent on the interactions with the surrounding lipid molecules. For detailed account of other coarse-grained models, we suggest the review by Ivan Colluzuerta.94 Full atomistic simulation packages on the hand, are capable of handling fully resolved all-atom structures of entire proteins and have precise description of bonded and nonbonded parameters and consequently involve longer compute times. RosettaDesign,57 Maestro95 Schrodinger Inc., PoreDesigner,99 and IPRO61 are examples of such full atom protein design packages. These packages have two essential compute modules: (a) rotamer chooser, and (b) force-field dependent evaluation of redesigned protein. During protein design RosettaDesign and Maestro both create large randomized libraries of protein variants with minimum deviation from native structure or a scored property (hydrophobicity, binding to an interacting partner, etc.), followed by evaluating enthalpic energy scores for each design using their respective empirical force-field energy functions. These energy scores are used to subsequently rank the designs depending on the design objective (such as interaction with a ligand). IPRO and PoreDesigner on the other hand iteratively uses a mixed-integer linear program to identify unique combinations of amino acid substitutions which satisfy the design objective. These choices are driven by CHARMM force-field based energy scores accounting for bonded and nonbonded energy terms. IPRO is an iterative protein redesign and optimization tool which emulates focused mutagenesis to identify stable enzyme variants that accomplish intended binding or unbinding of a substrate (or improve binding with one simultaneously eliminating with another). Figure 4 provides a general seven-step schematic overview of RosettaDesign and IPRO execution modules. DESADER acronym represents the seven general steps of: Dock substrate, Ensure catalytic constraints, Substrate binding residue identification, Adjacent residue repacking, Designing sequence, Energy minimization, and Ranking of designs. PoreDesigner relies on similar principles and predicts designs that enable users to precisely tune the pore size of any channel protein, thus offering interaction or size-based separations of aqueous solute mixtures. It has been experimentally validated to be able successfully redesign a bacterial porin to narrow pore sizes that performed perfect desalination using a membrane assembly. Donald et al96 and Pantazes et al97 provides a comprehensive review of other algorithms for computational protein design.

2 | SuCCeses

2.1 Successes in the “directed evolution” of enzymes

Enzyme design is a difficult challenge as only an infinitesimally small fraction of possible amino acid sequences adopts a functional fold. It has been estimated98(using a beta-lactamase as a proxy) that the fraction of all sequences that fold into viable enzymes with some minimal

![Figure 4](https://example.com/figure4.png)

**Figure 4** The seven-step DESADER schematic overview of enzyme redesign computational workflows of RosettaDesign and IPRO. RosettaDesign uses a stochastic Monte Carlo to create a library of enzyme variants. IPRO uses a deterministic mixed-integer linear optimization program to identify amino acid substitutions which are driven by the biochemical design objective (such as, maximize or minimize CHARMM-based interaction energy score with the ligand). [Color figure can be viewed at wileyonlinelibrary.com]
activity is as low as 1 in 10. This implies that random mutations predominantly tend to almost always adversely affect protein function. Thus, directed evolution capitalizes on the range of weak promiscuous activities of enzymes which can be quickly driven toward a desired catalytic activity only after pin-pointing few key mutations. Random mutagenesis, focused mutagenesis, and homologous recombination protocols along with efficient expression and screening of variant proteins have yielded several successes in redesigning enzymes for improved catalysis, altered substrate and cofactor specificities, and stability. Reetz et al. used a two-step protocol where first, epPCR was employed to obtain a short library of enantioselective cyclohexane monoxygenases with R or S selectivity that showed at least 95% activity compared to wild type, followed by random mutagenesis with subsequent screening for activity yielded eight with turnovers of the desired enantiomer ranging from two- to ninefold improvement over the wild type. Sequencing these mutants revealed only one to three amino acid changes in these eight mutants.

A separate endeavor by Arnold and co-workers pushed the activity–stability trade-off by using random mutagenesis, recombination, and screening of mesophilic Bacillus subtilis p-nitrobenzyl esterase and designed seven thermostable variants with melting temperatures ($T_m$) higher than the wild type by 5–14°C. Out of these seven, three best mutants were identified to show activities higher than wild type. These best mutant ($T_m = 66.5°C$; specific activity $= 0.16 \text{ mmol product)/(min mg enzyme)}$ where wild type activity was $0.125 \text{ mmol product)/(min mg enzyme)}$) was screened from 1,500 possible variants and exhibited stability at par with thermophilic enzymes. This was comparable with results from site directed mutagenesis which are beyond the scope of rational design strategies. Kim et al. demonstrated that random mutagenesis on bacterial sp. beta-glucosidase and screening using in vitro endocellulase-coupled assay yielded two highly active mutants with two (A19T, E358G) and four mutations (A19T, E358G, Q248R, M407V) with activities 7- and 27-folds higher than wild type, respectively. What sets this work apart from other similar studies is that—all these mutations were at least 9 Å away from the substrate and could not directly interact with the substrate to affect the turnover. This suggested that these mutations bring about conformation changes to the active site thereby providing a congenial groove for the substrate to sit and potentially react. We used the Agrobacterium sp. beta-galactosidase sequence (NCBI accession: WP_006316672.1), generated the best mutant sequence, and homology modeled it using Swiss MODEL to show that these mutations are distal from the substrate-binding domain (see Figure 5).

In another effort Zhao et al developed a staggered extension process (StEP) for in vitro mutagenesis and recombination of polynucleotide sequence. In contrast to optimized DNA-shuffling where DNase I digests a set of parent genes into an array of DNA fragments which are thermocycled into complete genes using DNA polymerase, StEP generates full-length recombination cassettes relying on a template-based extension using DNA polymerase. They tested the recombination efficiency between two thermostable subtilisin E genes which code for protease. Adenine to guanine changes in bases 1107 from gene 1, and 995 from gene 2 led to amino changes N181D, and N218S in the final protein. Single variants of N181D and N218S exhibited threefold and twofold longer, and the double mutant eight-fold longer half-lives than wild type at 65°C and were even stable at 75°C. Out of the 368 clones that were screened, 84% were active and showed wildtype-like catalytic activity. Out of the active ones, 21% exhibited thermostability like the double mutant, 61% were like the single mutant, and 18% were as thermostable as the wild type. In contrast to most random mutagenesis studies where an enzyme is engineered with the objective of finding a fitter variant with altered stereo-specificity, thermostability, or higher activity than wild type, Chen et al engineered a serine protease from B. subtilis to function in a highly nonnatural environment with high concentrations of polar organic solvent, dimethylformamide (DMF). Proteases and lipases are known to be promising catalysts for organic synthesis of acrylic and methacrylic esters which find applications as cement

**FIGURE 5** The four mutations that led to enhanced catalysis in beta-glucosidase have been marked as blue sticks and the protein is represented as light pink cartoon. All four mutations are too far from the binding pocket to interact with the substrate [Color figure can be viewed at wileyonlinelibrary.com]
material for knee or hip arthroplasty surgeries\textsuperscript{107} and PVC modifiers in plastic industry. In this work, mutagenesis and screening was performed with the objective of identifying amino acid substitutions that recover the lost catalytic activity of serine protease in organic media. Through three rounds of sequential screening, 10 amino acid changes were pinpointed within the binding groove of the substrate, in loop regions that offered sequence variability without affecting the tertiary folds of the reactive pocket—and restored catalysis. Seven out of the ten mutations were seen in other protease homologs from other organisms. To investigate the effects of each and every mutation, 10 single variants were generated and checked for enhancement of catalytic activity ($K_m$ and $k_{cat}$). Results indicated G131D mutation alone enhanced substrate affinity with 20% DMF by $\sim$90% (reduction of $K_m$ from 12.2 to 1.4 mM). Furthermore, N181S, T255A, E156G, S182G, and S188P mutations were also identified to reduce $K_m$ and all mutants with last three mutations improved catalysis even in aqueous media. Overall, we noticed that mutations where the substituted amino acids are hydrophobic led to lower $K_m$ in organic media and thus stabilized the substrate, whereas charged residues (D60N, Q103R, and G131D) improved $k_{cat}$ and thus product formation. This is also corroborated from the observation that the latter set of mutations also enhanced turnover in aqueous medium (0% DMF).

### 2.2 Successes in computational enzyme redesign

Nearly all engineered enzymes that are used today emerged from structure-based protein-engineering efforts of the 1980s. The successes have been notable, but the results came slowly till the advent of directed evolution in 1990s that led to major breakthroughs. However, most amino acid changes accumulated during evolution have marginal or no effect on the desired catalysis, making it a "needle in a haystack" problem to pinpoint key positions. To this end, computational methods have shown promise in sampling thousands of amino acid combinations and conformations with assessment of their impact on protein stability. Table 1 shows 50 key publications in protein design that uses computational enzyme design to create an overall active site framework. This work proves the potential of a synergistic workflow between computational enzyme design to create an overall active site framework, and molecular evolution to explore novel enzyme-mediated reactions. The Baker lab followed up with another breakthrough with designing enzymes for an energetically more demanding retro-aldol\textsuperscript{174} reaction that involved breaking a carbon–carbon bond of a hydroxy-carbonyl compound to form an aldehyde (or ketone) and another carbonyl moiety using acid–base catalysis initiated by a nucleophilic striking geometrical resemblance with zing-finger protein Zif268 even though the sequence similarity (39%) and identity (21%) were low, with most conserved residues located in buried and ordered regions of the protein indicating this to be a novel sequence. NCBI p-BLAST\textsuperscript{167} revealed this sequence to have similar alignment score (<39% identity) with any random amino acid sequence of similar length. This work paved the path for competing in silico methods to handle immense combinatorial search required for computational protein design, and inspired the development of various molecular-mechanics based force fields (CHARMM,\textsuperscript{168} Amber,\textsuperscript{80} gromacs\textsuperscript{169}). These force fields started factoring in near-accurate contributions of van der Waals, electrostatics, and solvation terms. Furthermore, Summers and Schulten introduced molecular dynamics for studying temporal fold changes and stability of biomolecular complexes\textsuperscript{170} (such as enzyme-ligand) to determine their macroscopic thermodynamic properties following the ergodic hypothesis.

Soon after, David Baker and colleagues used a novel computational enzyme design methodology\textsuperscript{171} to facilitate the Kemp elimination reaction—which has a high activation energy barrier and for which no naturally occurring enzyme existed. Eight in silico designs were generated containing one of the two proposed catalytic motifs. Directed evolution on these designs produced $>200$-fold increase in $k_{cat}/K_m$ values. The Kemp elimination reaction is the amino-induced elimination of benzisoxazole into relevant o-cyanoenolate ion.\textsuperscript{172} The reaction requires a base-mediated proton abstraction from a carbon with subsequent dispersion of the resulting negative charge or stabilization of the partial negative charge on the phenolic oxygen. To this end, the authors designed two alternative ideal catalytic bases—(a) Asp-His dyad, or (b) single aspartate or glutamate. Quantum-mechanical calculations on the backbone of the desired binding pocket was used to choose an optimal combination of amino acids that served the dual objective of stabilizing the substrate and positioning the catalytic base at the appropriate distance from the substrate. RosettaMatch\textsuperscript{57} was used to screen about $10^5$ binding pocket designs by finding the most stable side chain conformations of the pocket residues from each design, given their backbone conformation. The designs were scored based on binding free energies between the enzyme and the transition state. Forty-nine top designs were synthesized in vitro and eight showed catalytic activity. After seven rounds of mutagenesis and screening, one of the designs showed a $k_{cat}/K_m$ value of 2,600 (M s$^{-1}$) which was a result of fine-tuning the pocket residues to accommodate the substrate better. A recent work by Kingsley et al\textsuperscript{173} defines the binding pocket as "substrate tunnels" and the authors demonstrate that turnover can be severely impacted by altering the pocket residues even if the catalytic motif is unperturbed. This work proves the potential of a synergistic workflow between computational enzyme design to create an overall active site framework, and molecular evolution to explore novel enzyme-mediated reactions. The Baker lab followed up with another breakthrough with designing enzymes for an energetically more demanding retro-aldol reaction that involved breaking a carbon–carbon bond of a hydroxy-carbonyl compound to form an aldehyde (or ketone) and another carbonyl moiety using acid–base catalysis initiated by a nucleophilic...
<table>
<thead>
<tr>
<th>Year</th>
<th>Study</th>
<th>Experimental contribution</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>Design of “bait-and-switch” catalytic antibodies</td>
<td>Reactive immunization</td>
<td>Xu et al (^{108})</td>
</tr>
<tr>
<td>2000</td>
<td>Shape complementarity, binding-site dynamics, and transition state (TS) stabilization using 1E9 antibody by QM study of Diels–Alder catalysis</td>
<td></td>
<td>Chen et al (^{109})</td>
</tr>
<tr>
<td>1993</td>
<td>QM calculations on endo and exo stereoisomeric TSs of Diels–Alder cycloaddition</td>
<td>Diels–Alder cycloaddition reaction to obtain enantiomerically pure products</td>
<td>Gouverneur et al (^{110})</td>
</tr>
<tr>
<td>1998</td>
<td>QM calculations identify Asp(^{H50}) and Tyr(^{36}) as catalytic residues and Asn(^{L91}) for TS stability in 13G5 catalyzed exo Diels-Alder between N-butadienyl carbamate and N,N-dimethylacrylamide</td>
<td>Crystal structure of exo Diels–Alderase inhibitor complex solved at 1.95 Å</td>
<td>Heine et al (^{111})</td>
</tr>
<tr>
<td>2003</td>
<td>MD relaxation of 13G5 antibody around rigid TS revealed role of three water molecules in orienting catalytic base Asp(^{H50})</td>
<td>Absolute enantiomeric selectivities of 13G5 and 4D5 antibodies established</td>
<td>Cannizzaro et al (^{112})</td>
</tr>
<tr>
<td>2002</td>
<td>Shape complementarity of TS and catalytic triad Trp-Phe-Ser identified in 10F11 antibody for retro Diels-Alder reaction</td>
<td>Crystal structures of Fab 10F11 and 9D9 antibodies in complex with substrate analogs solved at 1.8 and 2.3 Å</td>
<td>Hugot et al (^{113})</td>
</tr>
<tr>
<td>1990</td>
<td>DFT calculations used to discern active site interactions (Trp(^{H104})-Phe(^{H101})) by (\pi)-(\pi) stacking to stabilize TS for 10F11 antibody during retro Diels–Alder reaction</td>
<td></td>
<td>Leach et al (^{114})</td>
</tr>
<tr>
<td>2002</td>
<td>Proof of nonspecific TS binding offered by antibodies guided by solvophobic effects, unlike enzymes in Diels–Alder reactions</td>
<td>Kinetic constants for TS-antibody binding calculated for 1E9, 39A11, 13G5, 4D5, 22C8, and 7D4 antibodies</td>
<td>Kim et al (^{115})</td>
</tr>
<tr>
<td>1995</td>
<td>Proof of Kemp eliminase activity being related to TS geometry and polarity of the solvent in 34E4 antibody</td>
<td>Catalytic antibody found for ring opening in Kemp elimination</td>
<td>Casey et al (^{116}), Kemp et al (^{117})</td>
</tr>
<tr>
<td>1988</td>
<td>QM calculations on aldol reactions exploring relative stability of “chair” and “twist boat” TS structures</td>
<td></td>
<td>Li and Houk (^{118})</td>
</tr>
<tr>
<td>1995</td>
<td>TS studies on ab38C2, ab84G3, and ab33F12 aldolase antibodies to catalyze aldol and retro-aldol reactions akin to class I aldolases using (\epsilon)-amino group of catalytic Lys(^{993})</td>
<td>Activities of aldolase antibodies measured to be comparable to natural aldolases</td>
<td>Wagner et al (^{119}), Barbas et al (^{120}), Hoffmann et al (^{121})</td>
</tr>
<tr>
<td>2003</td>
<td>QM study on polar residues at binding pockets of aldolase Abs in C–C bond-formation step</td>
<td></td>
<td>Arnó and Domingo (^{122})</td>
</tr>
<tr>
<td>1975</td>
<td>Aprotic polar solvents desolvate carboxylate reactant by stabilizing TS through dispersion interactions using Monte-Carlo free energy perturbation (FEP) calculations</td>
<td>Kinetic constants for this reaction are experimentally estimated</td>
<td>Kemp et al (^{123}), Zipse et al (^{124})</td>
</tr>
<tr>
<td>2003</td>
<td>QM, MD, and FEP calculations on 21D8 antibody for decarboxylation-catalyzed ring-opening reaction</td>
<td>21D8 catalyzes decarboxylation of 5-nitro-3-carboxybenzoxazole by 61,000-fold than in water.</td>
<td>Ujaque et al (^{125}), Lewis et al (^{126})</td>
</tr>
<tr>
<td>1993</td>
<td>QM study on endo-tet TS for cyclization of trans-epoxy alcohols show S(_{n1}) behavior and Asp(^{H05}), His(^{L95}) catalytic residues</td>
<td>X-ray structure of antibody Fab5C8 crystallized</td>
<td>Na et al (^{127}), Gruber et al (^{128})</td>
</tr>
<tr>
<td>1994</td>
<td>Homology model of 43C9 antibody variable region revealed Arg(^{L96}) to be the oxyanion hole and His(^{L91}) the catalytic nucleophile for hydrolysis of aromatic amides.</td>
<td>Water-mediated hydrogen-bonding network at the active site is key for catalysis seen from X-ray crystal of 43C9</td>
<td>Roberts et al (^{129}), Thayer et al (^{130})</td>
</tr>
<tr>
<td>2003</td>
<td>QM, MD, and FEP calculations on 43C9 reveals alternate mechanism using direct hydride attack</td>
<td></td>
<td>Chong et al (^{131})</td>
</tr>
<tr>
<td>1995</td>
<td>QM calculations to mimic active site of chorismite mutase antibodies IF7 and IIF1-2E11 (for Claisen rearrangement) revealed H-bond donors at the active site.</td>
<td>Catalytic rate analysis performed after crystallizing IF7 antibody.</td>
<td>Wiest et al (^{132}), Haynes et al (^{133})</td>
</tr>
</tbody>
</table>

(Continues)
<table>
<thead>
<tr>
<th>Year</th>
<th>Study</th>
<th>Experimental contribution</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>Reconstructed evolutionary adaptive path (REAP) analysis at active site of <em>Thermus aquaticus</em> DNA polymerase to accept unnatural NTPs</td>
<td>Whole-gene synthesis (library size = 93) reveals predicted single amino acid changes efficiently catalyze unnatural NTPs</td>
<td>Chen et al\textsuperscript{134}</td>
</tr>
<tr>
<td>2010</td>
<td>Improved enantioselectivity via 3DM analysis at four specific active sites on <em>Pseudomonas fluorescens</em> esterase</td>
<td>Site-saturation mutagenesis (library size = 500) yielded 200-fold improvement in activity and 20-fold higher enantioselectivity</td>
<td>Jochens et al\textsuperscript{135}</td>
</tr>
<tr>
<td>2008</td>
<td>Hot-spot selection to improve pH/protease stability of <em>S. capsulata</em> prolyl endopeptidase based on multiple sequence alignment and MLM on peptide library</td>
<td>Whole-gene synthesis (library size = 91) revealed 200-fold higher protease resistance and 20% higher activity</td>
<td>Ehren et al\textsuperscript{136}</td>
</tr>
<tr>
<td>2009</td>
<td>MD simulations to identify mutational hotspots in access tunnels to active site of <em>Rhodococcus rhodochrous</em> haloalkane dehalogenase</td>
<td>Site-directed and site-saturation mutagenesis (library size = 2,500) showed 32-fold higher activity</td>
<td>Pavlova et al\textsuperscript{137}</td>
</tr>
<tr>
<td>2009</td>
<td>SCHEMA structure guided recombination of peptide fragment from three CBH II cellulase for increased thermostability</td>
<td>Whole gene synthesis (library size = 48) showed 15% higher thermostability</td>
<td>Heinzelman et al\textsuperscript{138}</td>
</tr>
<tr>
<td>2010</td>
<td>MOE molecular modeling analysis for altered substrate specificity, solvent tolerance and thermostability on <em>Arthrobacter sp.</em> transaminase</td>
<td>Site-saturation and random mutagenesis (library size = 36,000)</td>
<td>Savile et al\textsuperscript{139}</td>
</tr>
<tr>
<td>2009</td>
<td>K* algorithm and SCMF entropy-based protocol using rotamer library and flexible ligand docking to switch specificity from Phe to Leu/Arg/Lys/Asp on gramicidin S synthetase A Phe-adenylation domain</td>
<td>Site-directed mutagenesis (library size = 10) showed 600-fold specificity shift from Phe to Leu by changes in $k_M$ values</td>
<td>Chen et al\textsuperscript{146}</td>
</tr>
<tr>
<td>2019</td>
<td>IPRO used to explore promiscuity of A domain of Ser-specific NRPS from <em>E. coli</em></td>
<td>Site-directed mutagenesis (library size = 160) identified 152 new Ser-specific domains</td>
<td>Throckmorton et al\textsuperscript{140}</td>
</tr>
<tr>
<td>2009</td>
<td>RosettaDesign to vary active-site and loop-length composition for human guanine deaminase to switch specificity for ammelide/cytosine</td>
<td>Site-directed mutagenesis and PCR assembly (library size = 10) showed $&gt;10^6$ specificity change</td>
<td>Murphy et al\textsuperscript{141}</td>
</tr>
<tr>
<td>2010</td>
<td>QM/ MM simulations using RosettaMatch on Diels-Alderase</td>
<td>Site-directed mutagenesis (library size = 100) showed activity similar to catalytic antibodies</td>
<td>Siegel et al\textsuperscript{142}</td>
</tr>
<tr>
<td>2009</td>
<td>VMD modeling to reconstitute active site of nitric oxide reductase (NOR) in myoglobin</td>
<td>Site-directed mutagenesis yielded functional NOR</td>
<td>Yeung et al\textsuperscript{143}</td>
</tr>
<tr>
<td>2009</td>
<td>Hotspot wizard server to create mutability maps based on sequence-structure information from existing protein databases</td>
<td>Holoalkane dehalogenase (DhaA) engineering from <em>Rhodococcus rhodochrous</em></td>
<td>Pavelka et al\textsuperscript{144} Pavlova et al\textsuperscript{137}</td>
</tr>
<tr>
<td>2007</td>
<td>Engineering proteinase K using machine learning and synthetic genes</td>
<td>24 amino acid substitutions in 59 variants were tested for hydrolase activity on tetrapeptides at 68$^\circ$C</td>
<td>Liao et al\textsuperscript{145}</td>
</tr>
<tr>
<td>2010</td>
<td>Functional benefits of distal mutations through induced allosteric for enantioselective Baeyer-Villiger monoxygenase using MD simulations (also discerned active site geometry changes)</td>
<td>Directed evolution experiments (library size = 400) revealed one double mutant that induced allosteric activity for</td>
<td>Wu et al\textsuperscript{146}</td>
</tr>
<tr>
<td>2003</td>
<td>Pairwise alignment of <em>N</em>-acetyluraminate lyase (NAL) and dihydropicolate synthase (DHDPS) revealed Leu-Arg mismatch at active site</td>
<td>An L142R mutation in NAL abolished NAL activity and improved DHDPS activity by eightfold</td>
<td>Joerger et al\textsuperscript{147}</td>
</tr>
<tr>
<td>2005</td>
<td>Four mutations to active site of keto-L-gluconate phosphate synthase identified to enhance promiscuity to arabinose-hex3-ulos e six phosphate synthase (HPS)</td>
<td>170-fold higher HPS activity is recorded.</td>
<td>Yew et al\textsuperscript{148}</td>
</tr>
<tr>
<td>2000</td>
<td>De novo design of helical bundle scaffolds for metal-chelation</td>
<td>Dinuclear metal-binding activity recording using His-triad catalytic motif</td>
<td>Hill et al\textsuperscript{149}</td>
</tr>
<tr>
<td>2018</td>
<td>PoreDesigner to redesign beta-barrel scaffold from <em>E. coli</em> OmpF to access any user-defined sub-nm pore size</td>
<td>Stopped-flow light scattering experiments reveal narrowest design perform like aquaporin</td>
<td>Chowdhury et al\textsuperscript{150}</td>
</tr>
</tbody>
</table>
attack on the ketone. In thin enamine catalysis, the carbinolamine intermediate undergoes spontaneous dehydration to yield imine/iminium product. Subsequently, the enamine tautomerizes to another imine which undergoes a similar dehydration to release the product and frees the enzyme. The authors constructed several protein scaffolds that can simultaneously accommodate both the transition states for the two-step reaction and grafted four alternative quantum-mechanically optimized catalytic motifs that would initiate the acid-base catalysis. Altogether 32 designs showed weak catalytic activity with the most active designs containing a co-crystallized water molecule which served the dual role of stabilizing the intermediate and also as a proton acceptor. Even though the X-ray crystallographic structure of the active site showed great agreement (RMSD < 1 Å), the catalytic efficiencies were low (0.74 (M s)$^{-1}$, and turnover of 1 molecule of product every 2 hr). Interestingly, from both these studies, the most promising folds generated in silico that had high catalytic efficiencies.
were triose-phosphate-isomerase (TIM) type containing eight alpha and beta helices. TIM toroids are known to be very effective for enzymatic reactions, and thus this shows convergence between in vivo and in silico fold preference.

Most computational enzyme redesign approaches were aimed at improving substrate specificity and catalyzing nonnatural reactions. One of the earliest examples to switch cofactor specificity was in *Candida boidinii* xylose reductase (CbXR). The authors gleaned and incorporated key cofactor switching mutation information from previous studies and were able to successfully alter the cofactor preference from NADPH to NADH. Amino acid changes in the CbXR binding pocket were systematically chosen using a mixed-integer linear program with the objective of simultaneously improving binding to NADH while eliminating binding with NADPH—where the binding score was expressed as a sum of van der Waals, electrostatics, and solvation terms. After sampling nearly 8,000 possible CbXR variants, 10 were found to show enhanced affinity for NADH and 7 of the 10 designs showed significant xylitol production. Eight out of ten designs showed more than 90% abolition of NADPH dependent activity while the remaining two showed equal preference for NADPH and NADH. The best design exhibited a 27-fold improvement in NADH-dependent activity. Other successes from the same group include, OptGraft for grafting a binding site from one protein into another protein scaffold, rational design to obtain 200-fold higher D-hydantoinase activity in *Bacillus steathermophilus* using just two amino acid changes, OptZyme for redesigning enzymes by improving binding to a transition state analogue instead of the substrate as it correlates with greater turnover, IPRO Suite of programs for fully-automated protein redesign, and altering substrate specificity of thioesterase enzyme from long-chain fatty acyl ACP to medium-chain ones.

Although the articles discussed show that computational enzyme designs are feasible, the catalytic activities of artificial enzymes with novel folds show significantly lower catalytic activities barring the high activities seen in Kemp eliminase. However, computational designs that maintain the wild-type binding groove geometry have remained extremely successful in exploiting the promiscuity of enzymes to drive a desired reaction by minimal residue interventions. Thus, it remains an open question, if computational designs alone will be able to outperform natural enzymes. A synergy between computational predictions and directed evolution still remains the best bet to date.

### 3 | NEW APPROACHES AND FUTURE DIRECTIONS

Even though designing a protein remains a challenging task due to the large sequence space that requires sampling, the number of resolved crystal structures are increasing day by day. A number of algorithms that use these sequence and structure databases to learn various sequence to structure features are emerging. Needless to say, machine learning and deep-learning neural networks are emerging as key players in this domain. Cadet et al came up with a supervised learning of enantioselective enzyme sequences and activity of *n* individual point mutations to predict the activity of all combinations of these point mutations. The method involves numerically encoding the sequences (wild type and single mutants) and experimental activities, converting them to a signal using Fourier transform, and using a partial least square step to predict the activity of a mutant which is a combination of multiple point mutations fed in the learning step. The correlation coefficient between 28 mutants validated experimentally revealed a good agreement ($R^2 = .81$). A nonconventional crowd-sourced online competitive gaming protocol—Foldit—to use human intuitions as a lever for accessing novel catalytic folds or predicting folded polypeptide geometries. Factoring in contributions from binding pocket geometries, alternate catalytic motifs, and hydrophobicity of the pocket would be a step forward in using these algorithms more reliably. Popova et al have developed a deep reinforcement learning tool for drug discovery to identify molecules with desired properties such as: hydrophobicity, melting point, and inhibitory activity against specific enzymes. Instead of constructing novel small molecule libraries, if this workflow can be used for screening whether a ligand will show activity against a library of an enzyme and its mutants—this could emerge as a useful enzyme engineering tool. Protein design thus remains an active field of research for the search of a unified set of rules that can be used for tuning substrate and cofactor specificity and tailoring novel functionalities or redesigning them anew. It could be worth mentioning, that directed evolution and computational design have also been aimed at creating synthetic pathways that take advantage of the new enzymes (e.g., Schwander et al and Siegel et al) along with several updated genome-scale networks of eukaryotes and pathway redesign tools. The marriage of new algorithms and directed evolution approaches bears promise of generating efficient catalysts needed by the food, pharmaceutical, and renewable energy industries.

### ACKNOWLEDGMENTS

We thank Dr. Brian Pfleger for performing experiments and characterization of enzyme variants for our several oleochemical enzyme redesign efforts. The US National Science Foundation (NSF) grant CBET-1703274 supports this collaboration. We also acknowledge Center for Bioenergy Innovation of US Department of Energy for funding our plant enzyme redesign work.

### ORCID

Ratul Chowdhury [https://orcid.org/0000-0003-4522-6911](https://orcid.org/0000-0003-4522-6911)

### REFERENCES

CHOWDHURY AND MARANAS

53. Toscano MD, Woycechowsky KJ, Hilvert D. Minimalist active-site
54. Lehmann A, Saven JG. Computational design of four-helix bundle
55. Huang SY, Grinter SZ, Zou X. Scoring functions and their evaluation
56. Fischer A, Enkler N, Neudert G, Bocola M, Sterner R, Merkl R. Trans-
57. Liu Y, Kuhlman B. RosettaDesign server for protein design.
59. Chartrain M, Salmon PM, Robinson DK, Buckland BC. Metabolic
60. Arnold FH. Combinatorial and computational challenges for biocata-
61. Pantazes RJ, Grisewood MJ, Li T, Gifford NP, Maranas CD. The iter-
62. Khoury GA, Fazelinia H, Chin JW, Pantazes RJ, Cirino PC,
63. Khare SD, Fleishman SJ. Emerging themes in the computational
64. Cui D, Zhang L, Jiang S, et al. A computational strategy for altering
65. Chen CY, Georgiev I, Anderson AC, Donald BR. Computational
structure-based redesign of enzyme activity. Proc Natl Acad Sci USA.
tional redesign of acyl-ACP Thioesterase with improved selectivity
toward medium-chain-length fatty acids. ACS Catal. 2017;7(6):3837-
67. Williams GJ, Domann S, Nelson A, Berry A. Modifying the stereo-
chemistry of an enzyme-catalyzed reaction by directed evolution.
1073/pnas.0635924100.
68. Khare SD, Fleishman SJ. Emerging themes in the computational
69. Chowdhury R, Allan MF, Maranas CD. OptiMAVEN-2.0: De novo
Design of Variable Antibody Regions against Targeted Antigen Epi-
70. Tiller KE, Chowdhury R, Li T, et al. Facile affinity maturation of anti-
body variable domains using natural diversity mutagenesis. Front
71. Kirk O, Borchert TV, Fuglsang CC. Industrial enzyme applications.
S0959-4693(02)00328-2.
73. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function
to predict protein-ligand interactions. J Mol Biol. 2000;295:337-
74. Buchete NV, Straub JE, Thirumalai D. Development of novel statisti-
75. Khoury GA, Fazelinia H, Chin JW, Pantazes RJ, Cirino PC.
76. Scott WRP, Hünenberger PH, Tironi IG, et al. The GROMOS biomo-
77. Poole AM, Ranganathan R. Knowledge-based potentials in protein
https://doi.org/10.1016/j.sbi.2006.06.013.
78. Dunbrack RL. Rosamer imaging in the 21st century. Curr Opin Struct
20035-9.
79. Lauck F, Smith CA, Friedland GF, Humphris EL, Kortemme T.
80. Tiller KE, Chowdhury R, Li T, et al. Facile affinity maturation of anti-
body variable domains using natural diversity mutagenesis. Front
81. Mackerell AD. Empirical force fields for biological macromolecules:
google.com/10.1002/jcc.20035.
82. Jorgensen WL, Tirado-Rives J. The OPLS potential functions for pro-
84. Scott WRP, Hünenberger PH, Tironi IG, et al. The GROMOS biomo-
tional redesign of acyl-ACP Thioesterase with improved selectivity
toward medium-chain-length fatty acids. ACS Catal. 2017;7(6):3837-
86. Scott WRP, Hünenberger PH, Tironi IG, et al. The GROMOS biomo-
doi.org/10.1021/acs.jctc.8b00529.
89. Fischer A, Enkler N, Neudert G, Bocola M, Sterner R, Merkl R. Trans-
cent: computational enzyme design by transferring active sites and
considering constraints relevant for catalysis. BMC Bioinformatics.
90. Pantazes RJ, Grisewood MJ, Li T, Gifford NP, Maranas CD. The iter-
ative protein redesign and optimization (IPRO) suite of programs.
23796.
91. Lehmann A, Saven JG. Computational design of four-helix bundle
proteins that bind nonbiological cofactors. Biotechnol Prog. 2008;24:
74-79. https://doi.org/10.1021/bp070178q.
92. Khoury GA, Fazelinia H, Chin JW, Pantazes RJ, Cirino PC,
93. Khoury GA, Fazelinia H, Chin JW, Pantazes RJ, Cirino PC,
94. Tiller KE, Chowdhury R, Li T, et al. Facile affinity maturation of anti-
body variable domains using natural diversity mutagenesis. Front
doi.org/10.1021/acs.jctc.8b00529.
96. Fischer A, Enkler N, Neudert G, Bocola M, Sterner R, Merkl R. Trans-
cent: computational enzyme design by transferring active sites and
considering constraints relevant for catalysis. BMC Bioinformatics.
97. Pantazes RJ, Grisewood MJ, Li T, Gifford NP, Maranas CD. The iter-
ative protein redesign and optimization (IPRO) suite of programs.
23796.
98. Lehmann A, Saven JG. Computational design of four-helix bundle
proteins that bind nonbiological cofactors. Biotechnol Prog. 2008;24:
74-79. https://doi.org/10.1021/bp070178q.
99. Khoury GA, Fazelinia H, Chin JW, Pantazes RJ, Cirino PC,
100. Cui D, Zhang L, Jiang S, et al. A computational strategy for altering
an enzyme in its cofactor preference to NAD(H) and/or NADP(H).
specificity switch in Shikimate dehydrogenase by rational design and
https://doi.org/10.1093/protein/gzx031.


