Received: 19 August 2019 | Revised: 14 October 2019 | Accepted: 19 October 2019

DOI: 10.1002/aic.16847

AIChE JOURNAL

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.[1] 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.[2] 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 1:1.58:0.28:0.3:0.01 (experiments performed by Johannes Mulder)[3] and reported that some of the proteins are catalytic. It was much later in 1878, that German physiologist Wilhelm Kuhne 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.[4] During the same time, German chemist, Emil Fischer postulated the "lock and key" hypothesis[5] 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

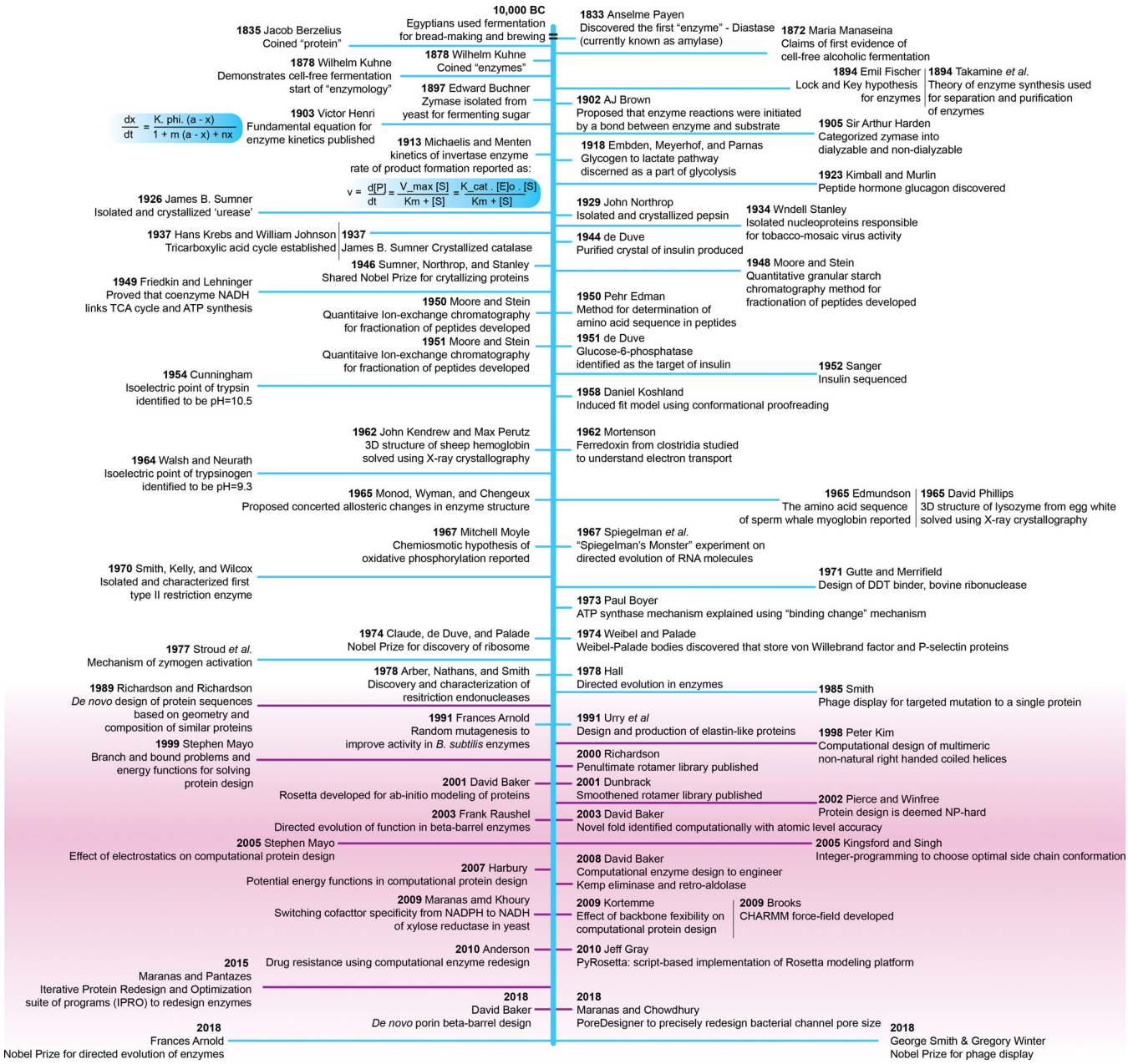

FIGURE 1 Seventy notable events in the history of enzyme engineering starting with the Egyptians using wild grain for bread-making and brewing to directed evolution and phage display techniques for which the Nobel Prize in chemistry was awarded in 2018. The computational milestones are indicated as purple lines [Color figure can be viewed at wileyonlinelibrary.com]

explained by the rigid enzyme model[6] till 1958 when Koshland laid out the "induced fit" theory[7] 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[8,9] of enzyme kinetics from his investigations on the invertase enzyme that hydrolyzes sucrose to glucose and fructose. 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.[10] About a decade later during 1930s, Sumner, Northrop, and Stanley independently crystallized urease,[11] pepsin,[12] and nucleoproteins responsible for tobacco-mosaic virus activity,[13] 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[14] 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[15,16] building on the work

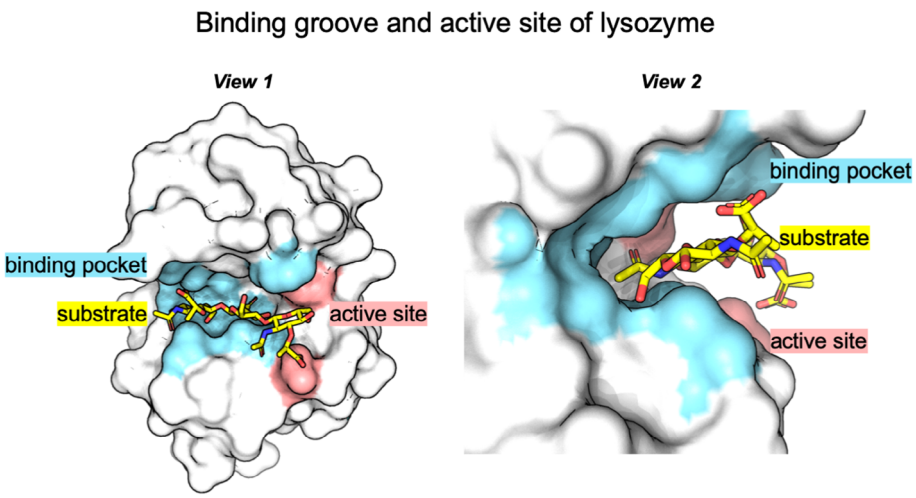

FIGURE 2 Two different views of the lysozyme binding site (marked in blue) and the active site residues highlighted in red. The peptidoglycan substrate is shown as yellow sticks [Color figure can be viewed at wileyonlinelibrary.com]

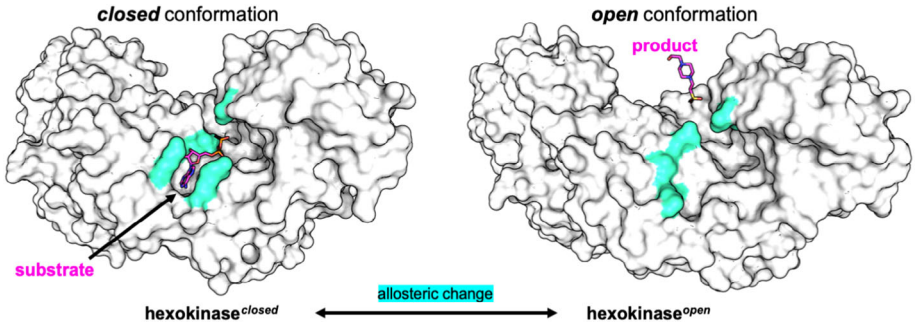

FIGURE 3 Conformational change in hexokinase during product release. The active site has been highlighted in bright green. The substrate and products have been marked as pink sticks. Accession IDs for closed and open hexokinase conformations are 2E2N, and 2E2Q [Color figure can be viewed at wileyonlinelibrary.com]

started by Charles Chibnall.[17] 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.[18] Subsequently, they followed up with a better and faster quantification approach[19] of amino acids in peptides. The next two decades saw isolation, purification, and characterization of various enzymes ranging from myoglobin from sperm whales[20] to high resolution lysozyme structures from egg white using X-ray crystallography.[21] 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[22] 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.[23,24] 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%.[25] 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.[26] Within a decade, Eigen and Gardiner proposed a cyclic "evolutionary machine"[27] 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 $>10^{10}$ designs and has been a cornerstone in the history of enzyme engineering.[28-31]

1.2 | 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.[32] 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 combinatorically 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[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[34] and Kunkel et al[35] were able to enhance the rate of mutation (from 10^{-10} to 10^{-4}) 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.[36] A modified epPCR was developed by Joyce et al that used a combination of *Taq* polymerase, 0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP, and 1 mM of dTTP, higher $MgCl_2$, and 0.5 mM $MnCl_2$ to reduce polymerase fidelity without affecting gene amplification and alleviated the strong bias toward A→G, and T→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,[37] and exploring novel carotenoid biosynthesis routes.[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 increase 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[39] containing degenerate codons for a targeted amino acid change is inserted[40] into a vector plasmid for expression of a desired enzyme variant. Parra et al[41] fed focused mutagenesis library of xylanase to epPCR to identify 12 more thermostable 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[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[43] also reported an optimized DNA shuffling workflow to control the point mutagenesis rate to as low as 0.05% by adding Mn^{2+} and Mg^{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[44] for estimating the occurrence of mutations in error-prone PCR after multiple generations[45] and the location of crossovers in directed evolution experiments.[46,47]

Recent trends in directed evolution has 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.[48-50] More recently, biochemists have aimed at dialing in novel functionalities in enzymatic proteins which are not seen in nature.[51,52] A decade old review by Toscano et al[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[54,55] and combinatorial search algorithms[56] devoted toward the efficient traversal of combinatorial space of residue alternatives. Software tools that integrate all these tasks include RosettaDesign,[57] Osprey,[58] Tinker,[59] TransCent,[60] and IPRO.[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,[62-65] 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.[66,67] In addition, efforts at introducing a novel enzymatic activity are also very difficult.[68] Nevertheless, there has been a lot of exciting, industrially relevant research focused on generation of stable humanized immunoproteins[69,70] with biopharmaceutical relevance and enzymes with enhanced turnover,[71] altered substrate-, and stereo-selectivity[72] 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[73] 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[74] developed statistical potentials using orientations of different amino acid side chains seen in experimentally resolved crystal structures to predict folded conformations for a given protein sequence. On the other hand, Lin et al[75] used evolutionary information from multiple sequence alignment of homologous proteins from closer organisms to develop knowledge-based statistic potentials. Here each protein was converted to several binary profiles, each containing information about different parameters (dihedrals, solvent accessibility, etc.) for each position (instead of actual amino acid sequences). An associated scoring system assessed how close a designed structure would be to existing structures from the alignment to have consistent folding. A similar statistical potential (TmFoldRec) for predicting folds in membrane-segments of transmembrane channels by learning from 124 crystallized transmembrane folds was published by Kozma et al.[76] Knowledge-based protein design tools provide the advantage of introducing additional descriptor terms (such as helix propensity and solvent exposure) without enhancing computing time significantly. Poole and Ranganathan[77] provide a comprehensive review of such similar knowledge-based potentials used for computational protein design. An integrated approach using a library (rotamer libraries[78]) of statistically preferred amino acid side-chain conformations in the phi–psi dihedral space and molecular-mechanics calculations to score a choice of a substituent amino acid rotamer forms the basis of most current day protein design software.[61,79]

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 helps 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,[80] CHARMM,[81] OPLS,[82,83] GROMOS,[84] and Rosetta.[85,86] Depending on whether each and every atom or only heavy atoms and polarizable hydrogens are represented within the force field, they are called "all atom" or "united atom" force fields. GROMOS is exclusively united atom force fields, Amber—(ff14SB[87] or ff15FB[88]), CHARMM and Rosetta all atom, while OPLS has both versions. Mackerel et al[81] 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[57,61,89] or coarse-grained[90,91] representations of proteins have been developed over the last two decades. Go and Taketomi[92] 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[93] was developed for predicting peptide conformations in lipid-bilayers. This was an extension to the lipid-exclusive Martini-force field.[91] 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

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 Colluza.[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] Maestro[95] Schrodinger Inc., PoreDesigner,[89] and IPRO[61] 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 al[96] and Pantazes et al[97] provides a comprehensive review of other algorithms for computational protein design.

2 | SUCCESSES

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 estimated[98](using a beta-lactamase as a proxy) that the fraction of all sequences that fold into viable enzymes with some minimal

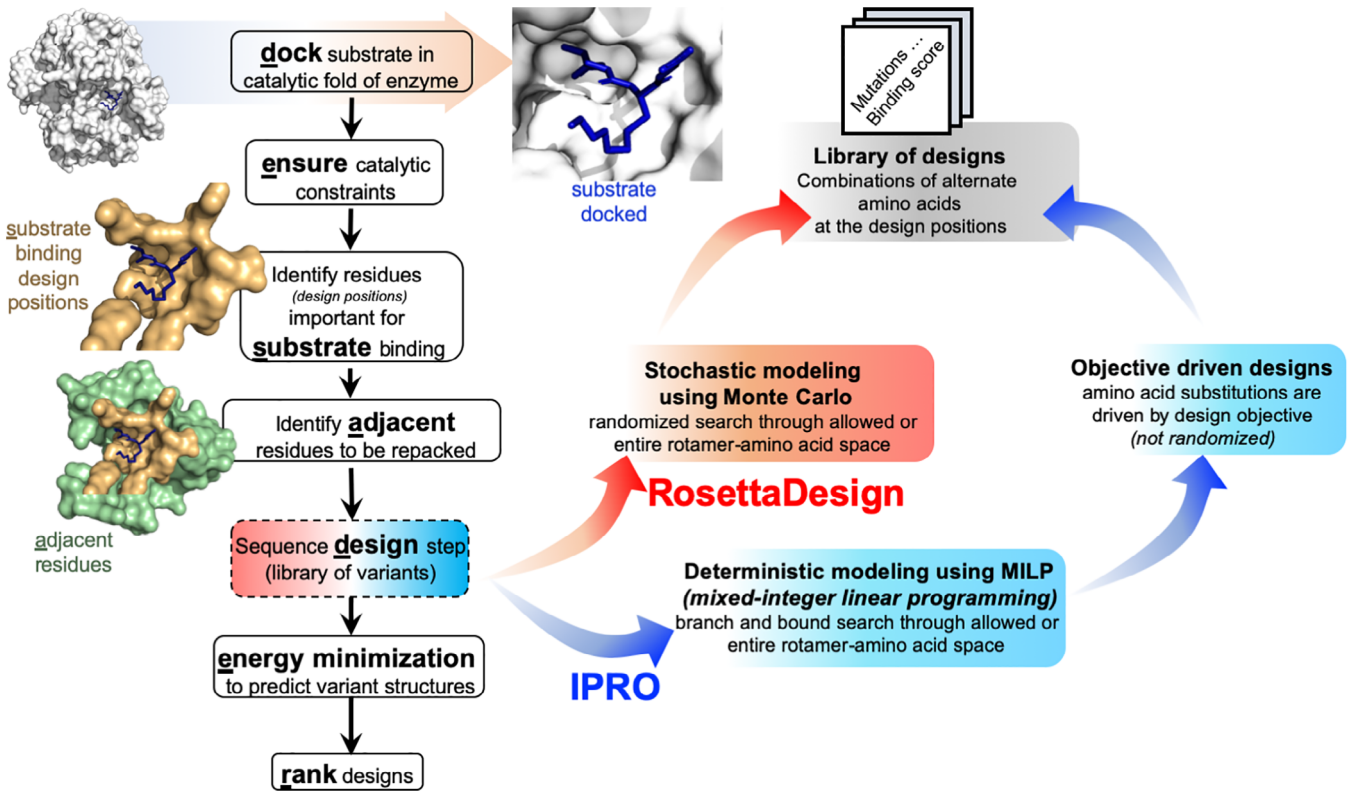

FIGURE 4 The seven-step DESADER schematic overview of enzyme redesign computational workflows of RosettaDesign and IPRO. RosettaDesign uses a stochastic Monte Carlo to create a library of enzyme variants. IPRO uses a deterministic mixed-integer linear optimization program to identify amino acid substitutions 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.[77] 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[99] used a two-step protocol where first, epPCR was employed to obtain a short library of enantioselective cyclohexane monooxygenases 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[100] 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 mmol product/(min mg enzyme) where wild type activity was 0.125 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[101] which, however, necessitate extensive sequence and structure information a priori. Ultimately, one of the most important insight gleaned from this study was that there is always an increase in activity with temperature until the enzyme denatures. This subsequently means simultaneous low-temperature activity and thermostability screening is sufficient to produce highly active variants viable across a wide temperature range.

Random mutagenesis explorations have been instrumental beneficial mutations which are beyond the scope of rational design strategies. Kim et al[102] demonstrated that random mutagenesis on *Agrobacterium 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[103] 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)[104] for in vitro mutagenesis and recombination of polynucleotide sequence. In contrast to optimized DNA-shuffling[43] 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 eightfold 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[105] 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[106] which find applications as cement

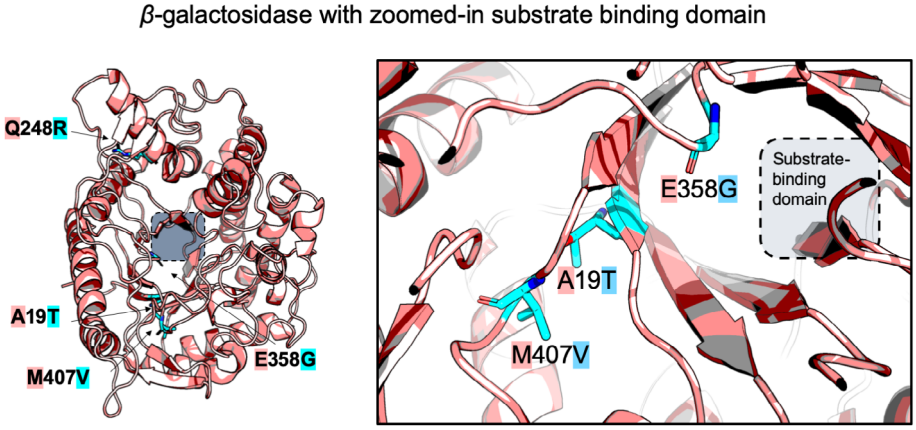

FIGURE 5 The four mutations that led to enhanced catalysis in beta-glucosidase have been marked as blue sticks and the protein is represented as light pink cartoon. All four mutations are too far from the binding pocket to interact with the substrate [Color figure can be viewed at wileyonlinelibrary.com]

material for knee or hip arthroplasty surgeries,[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 ~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 and experimental steps to generate stable de novo protein scaffolds, catalytic antibodies, and highly active enzyme redesigns. Several in silico tools have been able to glean design rules which have been used to tune the substrate and cofactor specificities of various enzymes along with unraveling novel, nonnatural catalytic modes. In 1997 the Mayo lab reported the first case of de novo redesign of streptococcal protein G β1 domain using a van der Waals potential to compute steric contributions, atomic solvation potential to favor burial of non-polar residues. The selection algorithm iteratively scanned and identified optimal sidechain conformation for a given backbone pose and accepted designs based on the sum of two pairwise interaction terms: (a) side chain and backbone, and (b) side chain and side chain. A statistically preferred set of 1.1×10^{62} side chain rotamers[165] was used, and a dead-end elimination theorem[166] was employed to constrain the search space to non-clashing ones and complete sequence design for a 50-residue window was achieved for every single design run. The design process only targeted nonpolar residues from the surface residues. The design exhibited 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[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,[168] Amber,[80] gromacs[169]). These force fields started factoring in near-accurate contributions of van der Waals, electrostatics, and solvation terms. Furthermore, Sumners and Schulten introduced molecular dynamics for studying temporal fold changes and stability of biomolecular complexes[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[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 eliminase reaction is the amino-induced elimination of benzisoxazole into relavant *o*-cyanophenolate ion.[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[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[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[174] reaction that involved breaking a carbon–carbon bond of a hydroxycarbonyl compound to form an aldehyde (or ketone) and another carbonyl moiety using acid–base catalysis initiated by a nucleophilic

TABLE 1 List of 50 computational enzyme design successes till date (grouped as per relevance)

Year	Study	Experimental contribution	References
2004	Design of "bait-and-switch" catalytic antibodies	Reactive immunization	Xu et al[108]
2000	Shape complementarity, binding-site dynamics, and transition state (TS) stabilization using 1E9 antibody by QM study of Diels–Alder catalysis		Chen et al[109]
1993	QM calculations on *endo* and *exo* stereoisomeric TSs of Diels–Alder cycloaddition	Diels–Alder cycloaddition reaction to obtain enantiomerically pure products	Gouverneur et al[110]
1998	QM calculations identify Asp^{H50} and Tyr^{L36} as catalytic residues and Asn^{L91} for TS stability in 13G5 catalyzed *exo* Diels-Alder between *N*-butadienyl carbamate and *N*,*N*-dimethylacrylamide	Crystal structure of *exo* Diels–Alderase inhibitor complex solved at 1.95 Å	Heine et al[111]
2003	MD relaxation of 13G5 antibody around rigid TS revealed role of three water molecules in orienting catalytic base Asp^{H50}	Absolute enantiomeric selectivities of 13G5 and 4D5 antibodies established	Cannizzaro et al[112]
2002	Shape complementarity of TS and catalytic triad Trp-Phe-Ser identified in 10F11 antibody for retro Diels-Alder reaction	Crystal structures of Fab 10F11 and 9D9 antibodies in complex with substrate analogs solved at 1.8 and 2.3 Å	Hugot et al[113]
1990	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		Leach et al[114]
2002	Proof of nonspecific TS binding offered by antibodies guided by solvophobic effects, unlike enzymes in Diels-Alder reactions	Kinetic constants for TS-antibody binding calculated for 1E9, 39A11, 13G5, 4D5, 22C8, and 7D4 antibodies	Kim et al[115]
1995 1996	Proof of Kemp eliminase activity being related to TS geometry and polarity of the solvent in 34E4 antibody	Catalytic antibody found for ring opening in Kemp elimination	Casey et al[116] Kemp et al[117]
1988	QM calculations on aldol reactions exploring relative stability of "chair" and "twist boat" TS structures		Li and Houk[118]
1995 1997 1998	TS studies on ab38C2, ab84G3, and ab33F12 aldolase antibodies to catalyze aldol and retro-aldol reactions akin to class I aldolases using ϵ-amino group of catalytic Lys^{H93}	Activities of aldolase antibodies measured to be comparable to natural aldolases	Wagner et al[119] Barbas et al[120] Hoffmann et al[121]
2003	QM study on polar residues at binding pockets of aldolase Abs in C–C bond-formation step		Arnó and Domingo[122]
1975 1995	Aprotic polar solvents desolvate carboxylate reactant by stabilizing TS through dispersion interactions using Monte-Carlo free energy perturbation (FEP) calculations	Kinetic constants for this reaction are experimentally estimated	Kemp et al[123] Zipse et al[124]
2003 1991	QM, MD, and FEP calculations on 21D8 antibody for decarboxylation-catalyzed ring-opening reaction	21D8 catalyzes decarboxylation of 5-nitro-3-carboxybenisoxazole by 61,000-fold than in water.	Ujaque et al[125] Lewis et al[126]
1993 1995	QM study on *endo*-tet TS for cyclization of trans-epoxy alcohols show S_N1 behavior and Asp^{H95}-His^{L89} catalytic residues	X-ray structure of antibody Fab5C8 crystallized	Na et al[127] Gruber et al[128]
1994 1999	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.	Water-mediated hydrogen-bonding network at the active site is key for catalysis seen from X-ray crystal of 43C9	Roberts et al[129] Thayer et al[130]
2003	QM, MD, and FEP calculations on 43C9 reveals alternate mechanism using direct hydride attack		Chong et al[131]
1995 1994	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.	Catalytic rate analysis performed after crystallizing IF7 antibody.	Wiest et al[132] Haynes et al[133]

(Continues)

TABLE 1 (Continued)

Year	Study	Experimental contribution	References
2010	Reconstructed evolutionary adaptive path (REAP) analysis at active site of *Thermus aquaticus* DNA polymerase to accept unnatural NTPs	Whole-gene synthesis (library size = 93) reveals predicted single amino acid changes efficiently catalyze unnatural NTPs	Chen et al[134]
2010	Improved enantioselectivity via 3DM analysis at four specific active sites on *Pseudomonas fluorescens* esterase	Site-saturation mutagenesis (library size = ~500) yielded ~200-fold improvement in activity and ~20-fold higher enantioselectivity	Jochens et al[135]
2008	Hot-spot selection to improve pH/protease stability of *S. capsulata* prolyl endopeptidase based on multiple sequence alignment and ML on peptide library	Whole-gene synthesis (library size = 91) revealed 200-fold higher protease resistance and 20% higher activity	Ehren et al[136]
2009	MD simulations to identify mutational hotspots in access tunnels to active site of *Rhodococcus rhodochrous* haloalkane dehalogenase	Site-directed and site-saturation mutagenesis (library size = 2,500) showed 32-fold higher activity	Pavlova et al[137]
2009	SCHEMA structure guided recombination of peptide fragment from three CBH II cellulase for increased thermostability	Whole gene synthesis (library size = 48) showed 15°C higher thermostability	Heinzelman et al[138]
2010	MOE molecular modeling analysis for altered substrate specificity, solvent tolerance and thermostability on *Arthobacter sp*. transaminase	Site-saturation and random mutagenesis (library size = 36,000).	Savile et al[139]
2009	K* algorithm and SCMF entropy-based protocol using rotamer library and flexible ligand docking to switch specificity from Phe to Leu/Arg/Lys/Glu/Asp on gramicidine S synthetase A Phe-adenylation domain	Site-directed mutagenesis (library size = 10) showed 600-fold specificity shift from Phe to Leu by changes in k_M values	Chen et al[66]
2019	IPRO used to explore promiscuity of A domain of Ser-specific NRPS from *E. coli*	Site-directed mutagenesis (library size = 160) identified 152 new Ser-specific domains	Throckmorton et al[140]
2009	RosettaDesign to vary active-site and loop-length composition for human guanine deaminase to switch specificity for ammelide/cytosine	Site-directed mutagenesis and PCR assembly (library size = 10) showed >10^6 specificity change	Murphy et al[141]
2010	QM/ MM simulations using RosettaMatch on Diels–Alderase	Site-directed mutagenesis (library size = 100) showed activity similar to catalytic antibodies	Siegel et al[142]
2009	VMD modeling to reconstitute active site of nitric oxide reductase (NOR) in myoglobin	Site-directed mutagenesis yielded functional NOR	Yeung et al[143]
2009	Hotspot wizard server to create mutability maps based on sequence-structure information from existing protein databases	Haloalkane dehalogenase (DhaA) engineering from *Rhodococcus rhodochrous*	Pavelka et al[144] Pavlova et al[137]
2007	Engineering proteinase K using machine learning and synthetic genes	24 amino acid substitutions in 59 variants were tested for hydrolase activity on tetrapeptides at 68°C	Liao et al[145]
2010	Functional benefits of distal mutations through induced allostery for enantioselective Baeyer–Villiger monooxygenase using MD simulations (also discerned active site geometry changes)	Directed evolution experiments (library size = 400) revealed one double mutant that induced allostery	Wu et al[146]
2003	Pairwise alignment of *N*-acetyluraminate lyase (NAL) and dihydropicolate synthase (DHDPS) revealed Leu-Arg mismatch at active site	An L142R mutation in NAL abolished NAL activity and improved DHDPS activity by eightfold	Joerger et al[147]
2005	Four mutations to active site of keto-L-gluconate phosphate synthase identified to enhance promiscuity to arabinose-hex3-ulose six phosphate synthase (HPS)	170-fold higher HPS activity is recorded.	Yew et al[148]
2000	De novo design of helical bundle scaffolds for metal-chelation	Dinuclear metal-binding activity recording using His-triad catalytic motif	Hill et al[149]
2018	PoreDesigner to redesign beta-barrel scaffold from *E. coli* OmpF to access any user-defined sub-nm pore size	Stopped-flow light scattering experiments reveal narrowest design perform like aquaporin	Chowdhury et al[150]

(Continues)

TABLE 1 (Continued)

Year	Study	Experimental contribution	References
2003	RosettaDesign used to design nine globular proteins	Circular dichroism experiments confirmed 8/9 of these proteins to be folded akin to native, and 6/9 showed up to 7 kcal/mol stabilities than wild type	Dantas et al[151]
2003	Top7 alpha/beta for accessing novel folds by iterative search through sequence design and structural folds	93 residue alpha/beta fold protein crystallized and matched structure prediction with RMSD = 1.2 Å.	Kuhlman et al[152]
2008	Computational design of periplasmic binding proteins through conformer sampling and continuous minimization revealed the importance of accurate capture of partial charges and electrostatic potentials	Evaluation of kinetic constants experimentally revealed the design to outperform the native K_d values (17microM vs. 210 nM-native)	Boas and Harbury[153]
2010	De novo alpha-helical bundle designed to bind heme-like large cofactors	UV/visible and circular dichroism, size exclusion chromatography and analytic centrifugation indicate active enzyme but low activities	Fry et al[154]
2010	MD simulations with all-atom Amber force fields were used to assess the integrity of a Kemp eliminase identifying caveats that static simulations are agnostic to		Kiss et al[155]
2015	Computational protocol for zeolites with detailed description of active site interactions		Sauer and Freund[156]
2010 2010	Influence of structural fluctuations on active-site preorganization in RA22 using molecular dynamics revealed an alternate conformation of substrate relative to His233 allows nucleophilic attack by Lys159 where Asp53 (original catalytic residue) is solvated and hence noncatalytic	A separate experimental endeavor discerned that majorly catalysis is done by Lys159 due to the favorable interaction with the naphthyl group of the substrate	Ruscio et al[157] Lasilla et al[158]
2010	Empirical valence bond calculations using FEP umbrella sampling on Kemp eliminase (KE) designs (KE07, KE70, KE59)	Difficulty in improving the KE activity is due to improper partial charge characterization	Frushicheva et al[159]
2010	Eight mutations identified on KE07 with the objective of improving activity further	2.6-fold lower K_M and 76-fold higher k_{cat} value yielding 200-fold higher activity	Khersonsky et al[160]
2011	Computational sequence optimization for increased activity of KE70	Nine rounds of random mutagenesis along with computational predictions yielded 12-fold lower K_M and 53-fold higher k_{cat}	Khersonsky et al[161]
2012	Fold-stabilizing mutations were predicted to enhance activity of KE59.	16 rounds of directed evolution yielded >2000-fold increase in activity	Khersonsky et al[162]
2012	Iterative approach to ensure every design cycle necessitates active enzyme redesigns and MD screening of mutants before experiments	Kinetic characterization of Kemp eliminases HG-3, HG-2 show higher activity of HG-3	Privett et al[163]
2017	Computational redesign of Acyl-ACP thioesterase with improved selectivity toward medium-chain-length fatty acids	27 variants with enhanced C8-production titers were constructed and best mutant was crystallized (5TID)	Grisewood et al[67]
2017	Highly active C8-Acyl-ACP using synthetic selection and computational modeling	1.7 g/L C8-titers with >90% specificity toward C8 and 15-fold increase in k_{cat} over WT	Lozada et al[164]

Note: The experimental aspect of each of these endeavors have been noted as well thus indicating that majority of these successes are due to synergistic effort of simulations and experiments.

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,[175] 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).[63] 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[176] for grafting a binding site from one protein into another protein scaffold, rational design to obtain 200-fold higher D-hydantoinase activity in *Bacillus stearothermophilus* using just two amino acid changes,[177] OptZyme[178] 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[61] for fully-automated protein redesign, and altering substrate specificity of thioesterase enzyme from long-chain fatty acyl ACP to medium-chain ones.[50]

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[179] came up with a supervised learning of enantioselective enzyme sequences and activity of n individual point mutations to predict the activity of all combinations (2^n) 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[180] 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[181] 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 be aimed at creating synthetic pathways that take advantage of the new enzymes (e.g., Schwander et al[182] and Siegel et al[183]) along with several updated genome-scale networks of eukaryotes[184,185] and pathway redesign tools.[186,187] 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

REFERENCES

1. Ponomarenko EA, Poverennaya EV, Ilgisonis EV, et al. The size of the human proteome: the width and depth. *Int J Anal Chem*. 2016; 2016:1-6. https://doi.org/10.1155/2016/7436849.
2. Alba-Lois L, Segal-Kischinevzky C. Yeast fermentation and the making of beer and wine. *Nat Educ*. 2010;455. https://doi.org/10.1038/455699a.
3. F Farhani. Protein Purification Using Zeolite Adsorbent [dissertation]. UMP; 2009. http://umpir.ump.edu.my/id/eprint/916/

4. Eduard B. Alkoholische Gärung ohne Hefezellen. *Berichte der Dtsch Chem Gesellschaft*. 2018;30(1):1110-1113. https://doi.org/10.1002/cber.189703001215.
5. Fischer E. Ueber die Glucoside der Alkohole. *Berichte der Dtsch Chem Gesellschaft*. 1893;26(3):2400-2412. https://doi.org/10.1002/cber.18930260327.
6. Koshland DE. The key–lock theory and the induced fit theory. *Angew Chemie Int Ed English*. 1995;33:2375-2378. https://doi.org/10.1002/anie.199423751.
7. Koshland DE. Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci USA*. 1958;44(2):98-104. http://www.ncbi.nlm.nih.gov/pubmed/16590179%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC335371.
8. Henri V. Théorie générale de l'action de quelques diastases. *C R Hebd Seances Acad Sci*. 1902;135:916-919.
9. Henri V. *ois générales de l'action des diastases*. Paris, France: Hermann; 1903.
10. Michaelis L, Menten ML. Die Kinetik der Invertinwerkung (the kinetics of invertase activity). *Biochem Z*. 1913. https://path.upmc.edu/divisions/chp/PDF/Michaelis-Menten_Kinetik.pdf.
11. Sumner J. The isolation and crystallization of the enzyme urease. *J Biol Chem*. 1926;69:435.
12. Northrop JH. Crystalline pepsin. *Science (80-)*. 1929;69:580. https://doi.org/10.1126/science.69.1796.580.
13. Stanley WM. Isolation of a crystalline protein possessing the properties of tobacco-mosaic virus. *Science (80-)*. 1935;81:644-645. https://doi.org/10.1126/science.81.2113.644.
14. Edman P, Högfeldt E, Sillén LG, Kinell P-O. Method for determination of the amino acid sequence in peptides. *Acta Chem Scand*. 1950;4:283-293. https://doi.org/10.3891/acta.chem.scand.04-0283.
15. Sanger F, Tuppy H. The amino-acid sequence in the phenylalanyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates. *Biochem J*. 1951;49:463-481. https://doi.org/10.1042/bj0490463.
16. Sanger F, Thompson EOP. The amino-acid sequence in the glycyl chain of insulin. I. the identification of lower peptides from partial hydrolysates. *Biochem J*. 1953;53:353-366.
17. Chibnall AC. Bakerian lecture - amino-acid analysis and the structure of proteins. *Proc R Soc London Ser B - Biol Sci*. 1942;131:136-160. https://doi.org/10.1098/rspb.1942.0021.
18. Stein W, Moore S. Chromatography of amino acids on starch columns. Separation of phenylalanine, leucine, isoleucine, methionine, tyrosine and valine. *J Biol Chem*. 1948;176:337-365.
19. Moore S, Stein WH. Photometric ninhydrin method for use in the chromatography of amino acids. *J Biol Chem*. 1948;176:367-388.
20. Edmundson AB. Amino-acid sequence of sperm whale myoglobin. *Nature*. 1965;205:883-887. https://doi.org/10.1038/205883a0.
21. Phillips DA. The hen egg-white lysozyme molecule. *Proc Natl Acad Sci*. 1967;57(3):483-495.
22. Lerner SA, Wu TT, Lin ECC. Evolution of a catabolic pathway in bacteria. *Science (80-)*. 1964;146:1313-1315. https://doi.org/10.1126/science.146.3649.1313.
23. Mills DR, Peterson RL, Spiegelman S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc Natl Acad Sci USA*. 1967;58:217-224.
24. Levisohn R, Spiegelman S. Further extracellular Darwinian experiments with replicating RNA molecules: diverse variants isolated under different selective conditions. *Proc Natl Acad Sci USA*. 2006;63(3):805-11. https://doi.org/10.1073/pnas.63.3.805.
25. Francis JC, Hansche PE. Directed evolution of metabolic pathways in microbial populations. I. Modification of the acid phosphatase pH optimum in *S. cerevisiae*. *Genetics*. 1972;70(1):59-73.
26. Hall BG. Number of mutations required to evolve a new lactase function in *Escherichia coli*. *J Bacteriol*. 1977;129(1):540-543.
27. Eigen M, Gardiner W. Evolutionary molecular engineering based on RNA replication. *Pure Appl Chem*. 1984;56:967-978. https://doi.org/10.1351/pac198456080967.
28. Robertson DL, Joyce GF. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*. 1990;344:467-468. https://doi.org/10.1038/344467a0.
29. Cadwell RC, Joyce GF. Mutagenic PCR. *Genome Res*. 1994;3:S136-S140. https://doi.org/10.1101/gr.3.6.S136.
30. Cadwell RC, Joyce GF. Randomization of genes by PCR mutagenesis. *Genome Res*. 1992;2:28-33. https://doi.org/10.1101/gr.2.1.28.
31. Beaudry AA, Joyce GF. Directed evolution of an RNA enzyme. *Science (80-)*. 1992;257:635-641. https://doi.org/10.1126/science.1496376.
32. Farinas ET, Bulter T, Arnold FH. Directed enzyme evolution. *Curr Opin Biotechnol*. 2001;12:545-551. https://doi.org/10.1016/S0958-1669(01)00261-0.
33. Leung DW, Chen E, Goeddel DV. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique*. 1989;1:11-15.
34. Zaccolo M, Williams DM, Brown DM, Gherardi E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J Mol Biol*. 1996;255:589-603. https://doi.org/10.1006/jmbi.1996.0049.
35. Tindall KR, Kunkel TA. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry*. 1988;27:6008-6013. https://doi.org/10.1021/bi00416a027.
36. Gupta RD, Tawfik DS. Directed enzyme evolution via small and effective neutral drift libraries. *Nat Methods*. 2008;5:939-942. https://doi.org/10.1038/nmeth.1262.
37. Landwehr M, Hochrein L, Otey CR, Kasrayan A, Bäckvall JE, Arnold FH. Enantioselective α-hydroxylation of 2-arylacetic acid derivatives and buspirone catalyzed by engineered cytochrome P450 BM-3. *J Am Chem Soc*. 2006;128:6058-6059. https://doi.org/10.1021/ja061261x.
38. Umeno D, Tobias AV, Arnold FH. Diversifying carotenoid biosynthetic pathways by directed evolution. *Microbiol Mol Biol Rev*. 2005;69:51-78. https://doi.org/10.1128/mmbr.69.1.51-78.2005.
39. Wells JA, Vasser M, Powers DB. Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites. *Gene*. 1985;34:315-323. https://doi.org/10.1016/0378-1119(85)90140-4.
40. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009;6:343-345. https://doi.org/10.1038/nmeth.1318.
41. Acevedo JP, Reetz MT, Asenjo JA, Parra LP. One-step combined focused epPCR and saturation mutagenesis for thermostability evolution of a new cold-active xylanase. *Enzyme Microb Technol*. 2017;100:60-70. https://doi.org/10.1016/j.enzmictec.2017.02.005.
42. Coco WM, Levinson WE, Crist MJ, et al. DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat Biotechnol*. 2001;19:354-359. https://doi.org/10.1038/86744.
43. Zhao H, Arnold FH. Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res*. 1997;25:1307-1308. https://doi.org/10.1093/nar/25.6.1307.
44. Moore GL, Maranas CD. Modeling DNA mutation and recombination for directed evolution experiments. *J Theor Biol*. 2000;205:483-503. https://doi.org/10.1006/jtbi.2000.2082.
45. Pritchard L, Corne D, Kell D, Rowland J, Winson M. A general model of error-prone PCR. *J Theor Biol*. 2005;234:497-509. https://doi.org/10.1016/j.jtbi.2004.12.005.
46. Wangkumhang P, Chaichoompu K, Ngamphiw C, et al. WASP: a web-based allele-specific PCR assay designing tool for detecting SNPs and mutations. *BMC Genomics*. 2007;8:275. https://doi.org/10.1186/1471-2164-8-275.

47. Shugay M, Zaretsky AR, Shagin DA, et al. MAGERI: computational pipeline for molecular-barcoded targeted resequencing. *PLoS Comput Biol*. 2017;13(5):e1005480. https://doi.org/10.1371/journal.pcbi.1005480.
48. Arnold FH. Combinatorial and computational challenges for biocatalyst design. *Nature*. 2001;409:253-257. https://doi.org/10.1038/35051731.
49. Chartrain M, Salmon PM, Robinson DK, Buckland BC. Metabolic engineering and directed evolution for the production of pharmaceuticals. *Curr Opin Biotechnol*. 2000;11:209-214. https://doi.org/10.1016/S0958-1669(00)00081-1.
50. Hernández Lozada NJ, Lai RY, Simmons TR, et al. Highly active C8-acyl-ACP Thioesterase variant isolated by a synthetic selection strategy. *ACS Synth Biol*. 2018;7:2205-2215. https://doi.org/10.1021/acssynbio.8b00215.
51. Brustad EM, Arnold FH. Optimizing non-natural protein function with directed evolution. *Curr Opin Chem Biol*. 2011;15:201-210. https://doi.org/10.1016/j.cbpa.2010.11.020.
52. Zhao H. Directed evolution of novel protein functions. *Biotechnol Bioeng*. 2007;98:313-317. https://doi.org/10.1002/bit.21455.
53. Toscano MD, Woycechowsky KJ, Hilvert D. Minimalist active-site redesign: teaching old enzymes new tricks. *Angew Chemie - Int Ed*. 2007;46:3212-3236. https://doi.org/10.1002/anie.200604205.
54. Moal IH, Moretti R, Baker D, Fernández-Recio J. Scoring functions for protein-protein interactions. *Curr Opin Struct Biol*. 2013;23:862-867. https://doi.org/10.1016/j.sbi.2013.06.017.
55. Huang SY, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys*. 2010;12:12899-12908. https://doi.org/10.1039/c0cp00151a.
56. Istrail S, Lam F. Combinatorial algorithms for protein folding in lattice models: a survey of mathematical results. *Commun Inf Syst*. 2009;9:303-346. https://doi.org/10.4310/cis.2009.v9.n4.a2.
57. Liu Y, Kuhlman B. RosettaDesign server for protein design. *Nucleic Acids Res*. 2006;34:W235-W238. https://doi.org/10.1093/nar/gkl163.
58. Gainza P, Roberts KE, Georgiev I, et al. Osprey: protein design with ensembles, flexibility, and provable algorithms. *Methods Enzymol*. 2013;523:87-107. https://doi.org/10.1016/B978-0-12-394292-0.00005-9.
59. Rackers JA, Wang Z, Lu C, et al. Tinker 8: software tools for molecular design. *J Chem Theory Comput*. 2018;14:5273-5289. https://doi.org/10.1021/acs.jctc.8b00529.
60. Fischer A, Enkler N, Neudert G, Bocola M, Sterner R, Merkl R. TransCent: computational enzyme design by transferring active sites and considering constraints relevant for catalysis. *BMC Bioinformatics*. 2009;10:54. https://doi.org/10.1186/1471-2105-10-54.
61. Pantazes RJ, Grisewood MJ, Li T, Gifford NP, Maranas CD. The iterative protein redesign and optimization (IPRO) suite of programs. *J Comput Chem*. 2015;36(4):251-263. https://doi.org/10.1002/jcc.23796.
62. 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.
63. Khoury GA, Fazelinia H, Chin JW, Pantazes RJ, Cirino PC, Maranas CD. Computational design of *Candida boidinii* xylose reductase for altered cofactor specificity. *Protein Sci*. 2009;18:2125-2138. https://doi.org/10.1002/pro.227.
64. 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). *FEBS J*. 2015;282:2339-2351. https://doi.org/10.1111/febs.13282.
65. García-Guevara F, Bravo I, Martínez-Anaya C, Segovia L. Cofactor specificity switch in Shikimate dehydrogenase by rational design and consensus engineering. *Protein Eng des Sel*. 2016;30:533-541. https://doi.org/10.1093/protein/gzx031.
66. Chen CY, Georgiev I, Anderson AC, Donald BR. Computational structure-based redesign of enzyme activity. *Proc Natl Acad Sci USA*. 2009;106:3764-3769. https://doi.org/10.1073/pnas.0900266106.
67. Grisewood MJ, Hernández-Lozada NJ, Thoden JB, et al. Computational redesign of acyl-ACP Thioesterase with improved selectivity toward medium-chain-length fatty acids. *ACS Catal*. 2017;7(6):3837-3849. https://doi.org/10.1021/acscatal.7b00408.
68. Khare SD, Fleishman SJ. Emerging themes in the computational design of novel enzymes and protein-protein interfaces. *FEBS Lett*. 2013;587:1147-1154. https://doi.org/10.1016/j.febslet.2012.12.009.
69. Chowdhury R, Allan MF, Maranas CD. OptMAVEn-2.0: De novo Design of Variable Antibody Regions against Targeted Antigen Epitopes. *Antibodies*. 2018;7(3):23.
70. Tiller KE, Chowdhury R, Li T, et al. Facile affinity maturation of antibody variable domains using natural diversity mutagenesis. *Front Immunol*. 2017;8:986. https://doi.org/10.3389/fimmu.2017.00986.
71. Kirk O, Borchert TV, Fuglsang CC. Industrial enzyme applications. *Curr Opin Biotechnol*. 2002;13:345-351. https://doi.org/10.1016/S0958-1669(02)00328-2.
72. Williams GJ, Domann S, Nelson A, Berry A. Modifying the stereochemistry of an enzyme-catalyzed reaction by directed evolution. *Proc Natl Acad Sci USA*. 2003;100:3143-3148. https://doi.org/10.1073/pnas.0635924100.
73. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol*. 2000;295:337-356. https://doi.org/10.1006/jmbi.1999.3371.
74. Buchete NV, Straub JE, Thirumalai D. Development of novel statistical potentials for protein fold recognition. *Curr Opin Struct Biol*. 2004;14:225-232. https://doi.org/10.1016/j.sbi.2004.03.002.
75. Dong QW, Wang XL, Lin L. Novel knowledge-based mean force potential at the profile level. *BMC Bioinformatics*. 2006;7(1):324. https://doi.org/10.1186/1471-2105-7-324.
76. Kozma D, Tusnády GE. TMFoldRec: a statistical potential-based transmembrane protein fold recognition tool. *BMC Bioinformatics*. 2015;16:201. https://doi.org/10.1186/s12859-015-0638-5.
77. Poole AM, Ranganathan R. Knowledge-based potentials in protein design. *Curr Opin Struct Biol*. 2006;16:508-513. https://doi.org/10.1016/j.sbi.2006.06.013.
78. Dunbrack RL. Rotamer libraries in the 21st century. *Curr Opin Struct Biol*. 2002;12:431-440. https://doi.org/10.1016/S0959-440X(02)00344-5.
79. Lauck F, Smith CA, Friedland GF, Humphris EL, Kortemme T. RosettaBackrub-a web server for flexible backbone protein structure modeling and design. *Nucleic Acids Res*. 2010;38(suppl_2):W569-75. https://doi.org/10.1093/nar/gkq369.
80. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general Amber force field. *J Comput Chem*. 2004;25:1157-1174. https://doi.org/10.1002/jcc.20035.
81. Mackerell AD. Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem*. 2004;25(13):1584-1604. https://doi.org/10.1002/jcc.20082.
82. Jorgensen WL, Tirado-Rives J. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and Crambin. *J Am Chem Soc*. 1988;110:1657-1666. https://doi.org/10.1021/ja00214a001.
83. Tieleman DP, MacCallum JL, Ash WL, Kandt C, Xu Z, Monticelli L. Membrane protein simulations with a united-atom lipid and all-atom protein model: lipid-protein interactions, side chain transfer free energies and model proteins. *J Phys Condens Matter*. 2006;18:S1221-S1234. https://doi.org/10.1088/0953-8984/18/28/S07.
84. Scott WRP, Hünenberger PH, Tironi IG, et al. The GROMOS biomolecular simulation program package. *J Phys Chem A*. 1999;103(19):3596-607. https://doi.org/10.1021/jp984217f.
85. Alford RF, Leaver-Fay A, Jeliazkov JR, et al. The Rosetta all-atom energy function for macromolecular modeling and design. *J Chem*

Theory Comput. 2017;13:3031-3048. https://doi.org/10.1021/acs.jctc.7b00125.
86. Rohl CA. Protein structure estimation from minimal restraints using Rosetta. *Methods Enzymol*. 2005;394:244-260. https://doi.org/10.1016/S0076-6879(05)94009-3.
87. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput*. 2015;11:3696-3713. https://doi.org/10.1021/acs.jctc.5b00255.
88. Wang LP, McKiernan KA, Gomes J, et al. Building a more predictive protein force field: a systematic and reproducible route to AMBER-FB15. *J Phys Chem B*. 2017;121:4023-4039. https://doi.org/10.1021/acs.jpcb.7b02320.
89. Chowdhury R, Ren T, Shankla M, et al. PoreDesigner for tuning solute selectivity in a robust and highly permeable outer membrane pore. *Nat Commun*. 2018;9:3661. https://doi.org/10.1038/s41467-018-06097-1.
90. Tozzini V. Coarse-grained models for proteins. *Curr Opin Struct Biol*. 2005;15:144-150. https://doi.org/10.1016/j.sbi.2005.02.005.
91. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH. The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B*. 2007;111:7812-7824. https://doi.org/10.1021/jp071097f.
92. Taketomi H, Ueda Y, Gō N. STUDIES ON PROTEIN FOLDING, UNFOLDING AND FLUCTUATIONS BY COMPUTER SIMULATION. *Int J Pept Protein Res*. 2009;7:445-459. https://doi.org/10.1111/j.1399-3011.1975.tb02465.x.
93. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI coarse-grained force field: extension to proteins. *J Chem Theory Comput*. 2008;4:819-834. https://doi.org/10.1021/ct700324x.
94. Coluzza I. Computational protein design: a review. *J Phys Condens Matter*. 2017;29:143001. https://doi.org/10.1088/1361-648X/aa5c76.
95. Laimer J, Hofer H, Fritz M, Wegenkittl S, Lackner P. MAESTRO - multi agent stability prediction upon point mutations. *BMC Bioinformatics*. 2015;16:116. https://doi.org/10.1186/s12859-015-0548-6.
96. Gainza P, Nisonoff HM, Donald BR. Algorithms for protein design. *Curr Opin Struct Biol*. 2016;39:16-26. https://doi.org/10.1016/j.sbi.2016.03.006.
97. Pantazes RJ, Grisewood MJ, Maranas CD. Recent advances in computational protein design. *Curr Opin Struct Biol*. 2011;21:467-472. https://doi.org/10.1016/j.sbi.2011.04.005.
98. Axe DD. Estimating the prevalence of protein sequences adopting functional enzyme folds. *J Mol Biol*. 2004;341:1295-1315. https://doi.org/10.1016/j.jmb.2004.06.058.
99. Reetz MT, Brunner B, Schneider T, Schulz F, Clouthier CM, Kayser MM. Directed evolution as a method to create enantioselective cyclohexanone monooxygenases for catalysis in Baeyer-Villiger reactions. *Angew Chemie - Int Ed*. 2004;43:4075-4078. https://doi.org/10.1002/anie.200460272.
100. Giver L, Gershenson A, Freskgard PO, Arnold FH. Directed evolution of a thermostable esterase. *Proc Natl Acad Sci USA*. 1998;95:12809-12813.
101. Stearman RS, Frankel AD, Freire E, Liu B, Pabo CO. Combining Thermostable mutations increases the stability of λ repressor. *Biochemistry*. 1988;27:7571-7574. https://doi.org/10.1021/bi00419a059.
102. Kim YW, Lee SS, Warren RAJ, Withers SG. Directed evolution of a glycosynthase from agrobacterium sp. increases its catalytic activity dramatically and expands its substrate repertoire. *J Biol Chem*. 2004;279:42787-42793. https://doi.org/10.1074/jbc.M406890200.
103. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997;18:2714-2723. https://doi.org/10.1002/elps.1150181505.
104. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol*. 1998;16:258-261. https://doi.org/10.1038/nbt0398-258.
105. Chen K, Arnold F. Tuning the activity of an enzyme for unusual environments: in dimethylformamide. *PNAS*. 1992;90:5618-5622. https://doi.org/10.1073/pnas.90.12.5618.
106. Warwel S, Steinke G, Klaas MRG. An efficient method for lipase-catalysed preparation of acrylic and methacrylic acid esters. *Biotechnol Tech*. 1996;10(4):283-286. https://doi.org/10.1007/BF00184030.
107. Speeckaert AL, Brothers JG, Wingert NC, Graham JH, Klena JC. Airborne exposure of methyl methacrylate during simulated Total hip Arthroplasty and fabrication of antibiotic beads. *J Arthroplasty*. 2015;30:1464-1469. https://doi.org/10.1016/j.arth.2015.02.036.
108. Xu Y, Yamamoto N, Janda KD. Catalytic antibodies: Hapten design strategies and screening methods. *Bioorganic Med Chem*. 2004;12(20):5247-68. https://doi.org/10.1016/j.bmc.2004.03.077.
109. Chen J, Deng Q, Wang R, Houk KN, Hilvert D. Shape complementarity, binding-site dynamics, and transition state stabilization: A theoretical study of Diels-Alder catalysis by antibody 1E9. *Angew Chem (Int Ed Eng)*. 2000;1(4):255-61.
110. Gouverneur VE, Houk KN, De Pascual-Teresa B, Beno B, Janda KD, Lerner RA. Control of the exo and endo pathways of the Diels-Alder reaction by antibody catalysis. *Science (80-)*. 1993;262(5131):204-8. https://doi.org/10.1126/science.8211138.
111. Heine A, Stura EA, Yli-Kauhaluoma JT, et al. An antibody exo Diels-Alderase inhibitor complex at 1.95 Å resolution. *Science (80-)*. 1998;279(5358):1934-40. https://doi.org/10.1126/science.279.5358.1934.
112. Cannizzaro CE, Ashley JA, Janda KD, Houk KN. Experimental determination of the absolute enantioselectivity of an antibody-catalyzed Diels-Alder reaction and theoretical explorations of the origins of stereoselectivity. *J Am Chem Soc*. 2003;125(9):2489-506. https://doi.org/10.1021/ja020879d.
113. Hugot M, Bensel N, Vogel M, et al. A structural basis for the activity of retro-Diels-Alder catalytic antibodies: Evidence for a catalytic aromatic residue. *Proc Natl Acad Sci USA*. 2002;99:9674-9678. https://doi.org/10.1073/pnas.142286599.
114. Leach AG, Houk KN, Reymond JL. Theoretical investigation of the origins of catalysis of a retro-Diels-Alder reaction by antibody 10F11. *J Org Chem*. 2004;69:3683-3692.
115. Kim SP, Leach AG, Houk KN. The origins of noncovalent catalysis of intermolecular Diels-Alder reactions by cyclodextrins, self-assembling capsules, antibodies, and RNAses. *J Org Chem*. 2002;67:4250-4260. https://doi.org/10.1021/jo011180d.
116. Casey ML, Kemp DS, Paul KG, CDD. The physical organic chemistry of benzisoxazoles. I. The mechanism of the base-catalyzed decomposition of benzisoxazoles. *J Org Chem*. 1973;38(13):2294-2301. https://doi.org/10.1021/jo00953a006.
117. Kemp DS, Casey ML. Physical organic chemistry of benzisoxazoles. II. Linearity of the brønsted free energy relationship for the base-catalyzed decomposition of benzisoxazoles. *J Am Chem Soc*. 1973;95(20):6670-6680. https://doi.org/10.1021/ja00801a024.
118. Li Y, Houk KN. Transition structures for the allylboration reactions of formaldehyde by allylborane and allylboronic acid. *J Am Chem Soc*. 1989;111(4):1236-1240. https://doi.org/10.1021/ja00186a011.
119. Wagner J, Lerner RA, Barbas CF. Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes. *Science*. 1995;270(5243):1797-1800. https://doi.org/10.1126/science.270.5243.1797.
120. Barbas CF, Heine A, Zhong G, et al. Immune versus natural selection: Antibody aldolases with enzymic rates but broader scope. *Science*. 1997;278(5346):2085-2092. https://doi.org/10.1126/science.278.5346.2085.

121. Hoffmann T, Zhong G, List B, et al. Aldolase antibodies of remarkable scope. *J Am Chem Soc*. 1998;120(12):2768-2779 https://doi.org/10.1021/ja973676b.
122. Arnó M, Domingo LR. Theozyme for antibody aldolases. Characterization of the transition-state analogue. *Org Biomol Chem*. 2003;1(4):637-643. https://doi.org/10.1039/b209636f.
123. Kemp DS, Cox DD, PKG. The physical organic chemistry of benzisoxazoles. IV. The origins and catalytic nature of the solvent rate acceleration for the decarboxylation of 3-carboxybenzisoxazoles. *J Am Chem Soc*. 1975;97(25):7312-7318. https://doi.org/10.1021/ja00858a018.
124. Zipse H, Apaydin G, Houk KN. A quantum mechanical and statistical mechanical exploration of the thermal decarboxylation of Kemp's other acid (benzisoxazole-3-carboxylic acid). The influence of solvation on the transition state geometries and kinetic isotope effects of a reaction with an awesome solvent effect. *J Am Chem Soc*. 1995;117:8608-17. https://doi.org/10.1021/ja00138a017.
125. Ujaque G, Tantillo DJ, Hu Y, Houk KN, Hotta K, Hilvert D. Catalysis on the coastline: Theozyme, molecular dynamics, and free energy perturbation analysis of antibody 21D8 catalysis of the decarboxylation of 5-nitro-3-carboxybenzisoxazole. *J Comput Chem*. 2003;24(1):98-110. https://doi.org/10.1002/jcc.10151.
126. Lewis C, Krämer T, Robinson S, Hilvert D. Medium effects in antibody-catalyzed reactions. *Science*. 1991;253(5023):1019-1022. https://doi.org/10.1126/science.1887215.
127. Na J, Houk KN, Shevlin CG, Janda KD, Lerner RA. The energetic advantage of 5-exo versus 6-endo epoxide openings: A preference overwhelmed by antibody catalysis. *J Am Chem Soc*. 1993;115(18):8453-8454. https://doi.org/10.1021/ja00071a067.
128. Gruber K, Zhou B, Houk KN, Lerner RA, Shevlin CG, Wilson IA. Structural basis for antibody catalysis of a disfavored ring closure reaction. *Biochemistry*. 1999;38(22):7062-7074. https://doi.org/10.1021/bi990210s.
129. Roberts VA, Stewart J, Benkovic SJ, Getzoff ED. Catalytic antibody model and mutagenesis implicate arginine in transition-state stabilization. *J Mol Biol*. 1994;235(3):1098-1116. https://doi.org/10.1006/jmbi.1994.1060.
130. Thayer MM, Olender EH, Arvai AS, et al. Structural basis for amide hydrolysis catalyzed by the 43C9 antibody. *J Mol Biol*. 1999;291(2):329-345. https://doi.org/10.1006/jmbi.1999.2960.
131. Chong LT, Bandyopadhyay P, Scanlan TS, Kuntz ID, Kollman PA. Direct hydroxide attack is a plausible mechanism for amidase antibody 43C9. *J Comput Chem*. 2003;24(12):1371-1377. https://doi.org/10.1002/jcc.10293.
132. Wiest O, Black KA, Houk KN. Density functional theory isotope effects and activation energies for the Cope and Claisen rearrangements. *J Am Chem Soc*. 1994;116:10336-10337. https://doi.org/10.1021/ja00101a078.
133. Haynes MR, Stura EA, Hilvert D, Wilson IA. Routes to catalysis: Structure of a catalytic antibody and comparison with its natural counterpart. *Science*. 1994;263(5147):646-652. https://doi.org/10.1126/science.8303271.
134. Chen F, Gauchera EA, Leal NA, et al. Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proc Natl Acad Sci USA*. 2010;107(5):1948-1953. https://doi.org/10.1073/pnas.0908463107.
135. Jochens H, Bornscheuer UT. Natural diversity to guide focused directed evolution. *ChemBioChem*. 2010;11(13):1861-1866. https://doi.org/10.1002/cbic.201000284.
136. Ehren J, Govindarajan S, Morón B, Minshull J, Khosla C. Protein engineering of improved prolyl endopeptidases for celiac sprue therapy. *Protein Eng Des Sel*. 2008;21(12):699-707. https://doi.org/10.1093/protein/gzn050.
137. Pavlova M, Klvana M, Prokop Z, et al. Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat Chem Biol*. 2009;5:727-733. https://doi.org/10.1038/nchembio.205.
138. Heinzelman P, Snow CD, Wu I, et al. A family of thermostable fungal cellulases created by structure-guided recombination. *Proc Natl Acad Sci USA*. 2009;106(14):5610-5615. https://doi.org/10.1073/pnas.0901417106.
139. Savile CK, Janey JM, Mundorff EC, et al. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science*. 2010;329(5989):305-309. https://doi.org/10.1126/science.1188934.
140. Throckmorton K, Vinnik V, Chowdhury R, et al. Directed evolution reveals the functional sequence space of an adenylation domain specificity code. *ACS Chem Biol*. 2019;14(9):2044-2054. https://doi.org/10.1021/acschembio.9b00532.
141. Murphy PM, Bolduc JM, Gallaher JL, Stoddard BL, Baker D. Alteration of enzyme specificity by computational loop remodeling and design. *Proc Natl Acad Sci USA*. 2009;106(23):9215-9220. https://doi.org/10.1073/pnas.0811070106.
142. Siegel JB, Zanghellini A, Lovick HM, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction. *Science*. 2010;329(5989):309-313. https://doi.org/10.1126/science.1190239.
143. Yeung N, Lin YW, Gao YG, et al. Rational design of a structural and functional nitric oxide reductase. *Nature*. 2009;462:1079-1082. https://doi.org/10.1038/nature08620.
144. 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. https://doi.org/10.1093/nar/gkp410.
145. Liao J, Warmuth MK, Govindarajan S, et al. Engineering proteinase K using machine learning and synthetic genes. *BMC Biotechnol*. 2007;7:16. https://doi.org/10.1186/1472-6,750-7-16.
146. Wu S, Acevedo JP, Reetz MT. Induced allostery in the directed evolution of an enantioselective Baeyer-Villiger monooxygenase. *Proc Natl Acad Sci USA*. 2010;107(7):2775-2780. https://doi.org/10.1073/pnas.0911656107.
147. Joerger AC, Mayer S, Fersht AR. Mimicking natural evolution in vitro: An N-acetylneuraminate lyase mutant with an increased dihydrodipicolinate synthase activity. *Proc Natl Acad Sci USA*. 2003;100(10):5694-5699. https://doi.org/10.1073/pnas.0531477100.
148. Yew WS, Wise EL, Rayment I, Gerlt JA. Evolution of enzymatic activities in the orotidine 5′-monophosphate decarboxylase suprafamily: Mechanistic evidence for a proton relay system in the active site of 3-keto-L-gulonate 6-phosphate decarboxylase. *Biochemistry*. 2004;43(21):6427-6437. https://doi.org/10.1021/bi049741t.
149. Hill RB, Raleigh DP, Lombardi A, Degrado WF. De novo design of helical bundles as models for understanding protein folding and function. *Acc Chem Res*. 2000;33(11):745-754. https://doi.org/10.1021/ar970004h.
150. Chowdhury R, Ren T, Shankla M, et al. PoreDesigner for tuning solute selectivity in a robust and highly permeable outer membrane pore. *Nat Commun*. 2018;9(1):3661. https://doi.org/10.1038/s41467-018-06097-1.
151. Dantas G, Kuhlman B, Callender D, Wong M, Baker D. A large scale test of computational protein design: Folding and stability of nine completely redesigned globular proteins. *J Mol Biol*. 2003;332(2):449-460. https://doi.org/10.1016/S0022-2836(03)00888-X.
152. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science*. 2003;302(5649):1364-1368. https://doi.org/10.1126/science.1089427.
153. Boas FE, Harbury PB. Design of Protein-Ligand Binding Based on the Molecular-Mechanics Energy Model. *J Mol Biol*. 2008;380(2):415-424. https://doi.org/10.1016/j.jmb.2008.04.001.

154. Christopher Fry H, Lehmann A, Saven JG, Degrado WF, Therien MJ. Computational design and elaboration of a de novo heterotetrameric a-helical protein that selectively binds an emissive abiological (porphinato)zinc chromophore. *J Am Chem Soc*. 2010;132(11):3997-4005. https://doi.org/10.1021/ja907407m.
155. Kiss G, Röthlisberger D, Baker D, Houk KN. Evaluation and ranking of enzyme designs. *Protein Sci*. 2010;19(9):1760-1773. https://doi.org/10.1002/pro.462.
156. Sauer J, FHJ. Models in catalysis. *Catal Lett*. 2015;145(1):109-125. https://doi.org/10.1007/s10562-014-1,387-1.
157. 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(39):14111-14115. https://doi.org/10.1021/ja905396s.
158. Lassila JK, Baker D, Herschlag D. Origins of catalysis by computationally designed retroaldolase enzymes. *Proc Natl Acad Sci USA*. 2010;107(11):4937-4942. https://doi.org/10.1073/pnas.0913638107.
159. Frushicheva MP, Cao J, Chu ZT, Warshel A. Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proc Natl Acad Sci USA*. 2010;107(39):16869-16874. https://doi.org/10.1073/pnas.1010381107.
160. Khersonsky O, Röthlisberger D, Dym O, et al. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the ke07 series. *J Mol Biol*. 2010;396(4):1025-1042. https://doi.org/10.1016/j.jmb.2009.12.031.
161. Khersonsky O, Röthlisberger D, Wollacott AM, et al. Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *J Mol Biol*. 2011;407(3):391-412. https://doi.org/10.1016/j.jmb.2011.01.041.
162. Khersonsky O, Kiss G, Röthlisberger D, et al. Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc Natl Acad Sci USA*. 2012;109(26):10358-10363. https://doi.org/10.1073/pnas.1121063109.
163. Privett HK, Kiss G, Lee TM, et al. Iterative approach to computational enzyme design. *Proc Natl Acad Sci USA*. 2012;109(10):3790-3795. https://doi.org/10.1073/pnas.1118082108.
164. Hernández Lozada NJ, Lai RY, Simmons TR, et al. Highly active C8-acyl-ACP thioesterase variant isolated by a synthetic selection strategy. *ACS Synth Biol*. 2018;7(9):2205-2215. https://doi.org/10.1021/acssynbio.8b00215.
165. Ponder JW, Richards FM. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol*. 1987;193:775-791. https://doi.org/10.1016/0022-2836(87)90358-5.
166. Desmet J, De Maeyer M, Hazes B, Lasters I. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*. 1992;356:539-542. https://doi.org/10.1038/356539a0.
167. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403-410. https://doi.org/10.1016/S0022-2836(05)80360-2.
168. Brooks BR, Brooks CL, Mackerell AD, et al. CHARMM: the biomolecular simulation program. *J Comput Chem*. 2009;30(10):1545-1614. https://doi.org/10.1002/jcc.21287.
169. Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model*. 2001;7:306-317. https://doi.org/10.1007/S008940100045.
170. Mesirov JP, Schulten K, Sumners DW. *Mathematical Applications to Biomolecular Structure and Dynamics, IMA Volumes in Mathematics and its Applications*. Springer-Verlag; Spain: 1996.
171. Röthlisberger D, Khersonsky O, Wollacott AM, et al. Kemp elimination catalysts by computational enzyme design. *Nature*. 2008;453:190-195. https://doi.org/10.1038/nature06879.
172. D'Anna F, La Marca S, Noto R. Kemp elimination: a probe reaction to study ionic liquids properties. *J Org Chem*. 2008;73:3397-3403. https://doi.org/10.1021/jo702662z.
173. Kingsley LJ, Lill MA. Substrate tunnels in enzymes: structure-function relationships and computational methodology. *Proteins Struct Funct Bioinform*. 2015;83:599-611. https://doi.org/10.1002/prot.24772.
174. Jiang L, Althoff EA, Clemente FR, et al. De novo computational design of retro-aldol enzymes. *Science (80-)*. 2008;319:1387-1391. https://doi.org/10.1126/science.1152692.
175. Richard JP, Zhai X, Malabanan MM. Reflections on the catalytic power of a TIM-barrel. *Bioorg Chem*. 2014;57:206-212. https://doi.org/10.1016/j.bioorg.2014.07.001.
176. Fazelinia H, Cirino PC, Maranas CD. OptGraft: a computational procedure for transferring a binding site onto an existing protein scaffold. *Protein Sci*. 2009;18(1):180-195. https://doi.org/10.1002/pro.2.
177. Lee SC, Chang YJ, Shin DM, et al. Designing the substrate specificity of d-hydantoinase using a rational approach. *Enzyme Microb Technol*. 2009;44(3):170-175. https://doi.org/10.1016/j.enzmictec.2008.10.020.
178. Grisewood MJ, Gifford NP, Pantazes RJ, et al. OptZyme: computational enzyme redesign using transition state analogues. *PLoS One*. 2013;8:e75358. https://doi.org/10.1371/journal.pone.0075358.
179. Cadet F, Fontaine N, Li G, et al. A machine learning approach for reliable prediction of amino acid interactions and its application in the directed evolution of enantioselective enzymes. *Sci Rep*. 2018;8:16757. https://doi.org/10.1038/s41598-018-35033-y.
180. Khatib F, Cooper S, Tyka MD, et al. Algorithm discovery by protein folding game players. *Proc Natl Acad Sci USA*. 2011;108:18949-18953. https://doi.org/10.1073/pnas.1115898108.
181. Wang J, Cao H, Zhang JZH, Qi Y. Computational protein design with deep learning neural networks. *Sci Rep*. 2018;8:1-9. https://doi.org/10.1038/s41598-018-24760-x.
182. Schwander T, Von Borzyskowski LS, Burgener S, Cortina NS, Erb TJ. A synthetic pathway for the fixation of carbon dioxide in vitro. *Science (80-)*. 2016;354:900-904. https://doi.org/10.1126/science.aah5237.
183. Siegel JB, Smith AL, Poust S, et al. Computational protein design enables a novel one-carbon assimilation pathway. *Proc Natl Acad Sci USA*. 2015;112(12):3704-3709. https://doi.org/10.1073/pnas.1500545112.
184. Thiele I, Swainston N, Fleming RMT, et al. A community-driven global reconstruction of human metabolism. *Nat Biotechnol*. 2013;31:419-425. https://doi.org/10.1038/nbt.2488.
185. Chowdhury R, Chowdhury A, Maranas CD. Using gene essentiality and synthetic lethality information to correct yeast and CHO cell genome-scale models. *Metabolites*. 2015;5(4):536-570. https://doi.org/10.3390/metabo5040536.
186. Burgard AP, Pharkya P, Maranas CD. OptKnock: a Bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*. 2003;84:647-657. https://doi.org/10.1002/bit.10803.
187. Ranganathan S, Suthers PF, Maranas CD. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol*. 2010;6:e1000744. https://doi.org/10.1371/journal.pcbi.1000744.

How to cite this article: Chowdhury R, Maranas CD. From directed evolution to computational enzyme engineering—A review. *AIChE J*. 2020;66:e16847. https://doi.org/10.1002/aic.16847