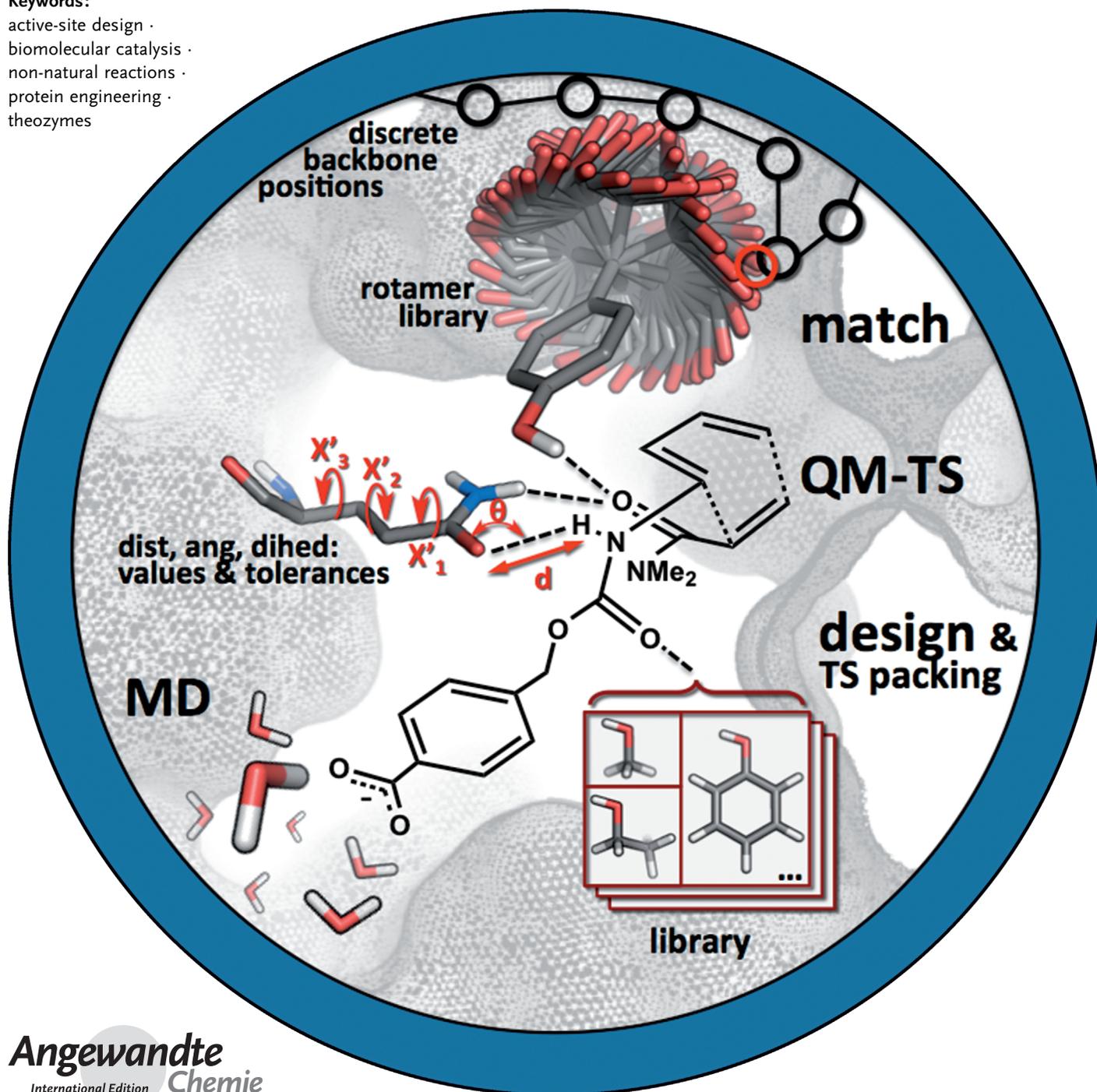


Computational Enzyme Design

Gert Kiss, Nihan Çelebi-Ölçüm, Rocco Moretti, David Baker, and K. N. Houk*

Keywords:

active-site design ·
biomolecular catalysis ·
non-natural reactions ·
protein engineering ·
theozymes



Recent developments in computational chemistry and biology have come together in the “inside-out” approach to enzyme engineering. Proteins have been designed to catalyze reactions not previously accelerated in nature. Some of these proteins fold and act as catalysts, but the success rate is still low. The achievements and limitations of the current technology are highlighted and contrasted to other protein engineering techniques. On its own, computational “inside-out” design can lead to the production of catalytically active and selective proteins, but their kinetic performances fall short of natural enzymes. When combined with directed evolution, molecular dynamics simulations, and crowd-sourced structure-prediction approaches, however, computational designs can be significantly improved in terms of binding, turnover, and thermal stability.

1. Introduction

Life depends on protein catalysts that control and accelerate reactions involved in metabolism. Over three billion years of evolution has led to enzymes that can catalyze chemical transformations that are too slow to be measured under normal conditions. The most proficient enzymes can accelerate reactions with turnover rates that occur as rapidly as the diffusion of reactants to the catalyst. Furthermore, enzymes enhance the catalytic rate of specific reactions and substrates. They frequently depend on cofactors or coenzymes, and are often sensitive to environmental conditions (pH value, temperature, and solvent), yet operate over a large spectrum thereof—making them ideal components for complex and tightly regulated metabolic pathways.

Humans have taken advantage of enzymatic processes for as long as naturally occurring fermentation has been controlled to preserve foods, to make bread and cheese, or to produce alcoholic beverages. However, the word “enzyme” from Ancient Greek “en zýmē” or “in dough/yeast” wasn’t coined until 1876, when German physiologist Wilhelm Kühne chose to simplify references to “elements that are responsible for fermentation processes” by giving them a name.^[1] At the time, the identity of enzymes was speculative, but Emil Fischer suggested in 1894 a “lock and key” model to explain the substrate specificity of enzymes. 32 years later, in 1926, James B. Sumner purified and crystallized urease, and showed that enzymes are proteins in their own right. In 1946, Linus Pauling speculated that enzymes are “closely complementary in structure to the activated complex for the reaction catalyzed”,^[2] a remarkable statement considering that at the time “no one [had] succeeded in determining the structure of any enzyme nor in finding out how the enzyme does its job”.^[3] The study of structure–function relationships at the atomic level continued to remain elusive for another two decades, until the first high-resolution crystal structure of an enzyme was solved and the field of structural biology emerged.^[4]

Since then, a surge of active research related to enzyme catalysis has continued to probe, adjust, and expand our understanding of these seemingly miraculous “nanomachines”. A variety of factors have been proposed to

explain the observed rate enhancements, and range from noncovalent transition-state (TS) stabilization (electrostatic, desolvation, restriction of motion, etc.) to covalent bonding (low energy barrier hydrogen bonds, formation of intermediates, metal–ion interactions, etc.). Researchers have come to embrace Pauling’s hypothesis as a general statement of what is responsible for the catalytic power of natural enzymes, but the power of preorganization and chemical catalysis is now recognized.^[5] The most proficient^[6] of these catalysts offer far more than an active site that is complementary to the TS; they enter into the reaction by altering the TS and thus change the free-energy profile from what it is in solution.^[7] This “covalent hypothesis” explains why the vast majority of enzymes can achieve TS binding constants that are orders of magnitude beyond what can be expected from noncovalent interactions.

Catalytic proficiency is formally the binding constant of the complex formed between the enzyme and the transition state, and was defined by Wolfenden as $K_{\text{tx}}^{-1} = (k_{\text{cat}}/K_{\text{M}})/k_{\text{uncat}}$.^[6] Remarkably, K_{tx}^{-1} spans 21 orders of magnitude (10^8 to 10^{29} M^{-1})^[6,8] for enzymes that have been studied to date,^[6,9–12] with an average K_{tx}^{-1} value of $10^{16.0 \pm 4.0} \text{ M}^{-1}$.^[13] This value corresponds to an average ΔG value for transition-state binding of 22 kcal mol^{-1} , but can range up to 38 kcal mol^{-1} , much higher than a noncovalent TS binding free energy of 15 kcal mol^{-1} .

From the Contents

1. Introduction	5701
2. Protein Engineering	5703
3. The Inside-out Approach to Computational Enzyme Design	5708
4. Computational Enzyme Design—Achievements	5711
5. Challenges in Enzyme Design	5719
6. Summary and Outlook	5721

[*] Dr. G. Kiss, Dr. N. Çelebi-Ölçüm, Prof. Dr. Dr. K. N. Houk
 Department of Chemistry and Biochemistry
 University of California, Los Angeles
 607 Charles E. Young Dr. East, Los Angeles CA, 90095 (USA)
 E-mail: houk@chem.ucla.edu

Dr. R. Moretti, Prof. Dr. D. Baker
 Department of Biochemistry and Howard Hughes Medical Institute
 University of Washington, Seattle, WA 98195 (USA)

Dr. G. Kiss
 Current address: Department of Chemistry, Stanford University,
 Stanford, CA 94305 (USA)

Dr. N. Çelebi-Ölçüm
 Current address: Yeditepe University, Department of Chemical
 Engineering, Istanbul (Turkey)

Chemists imagine the possibility of designing and synthesizing molecules with the attributes of enzymes (selective, proficient, “green”, operating in water under ambient conditions, nontoxic, and biodegradable). To do so, at least a subset of the above factors has to be considered, depending on the target reaction. Furthermore, it can be desirable to try and unite catalytic turnover with substrate-, stereo-, regio-, or chemoselectivity as well as a tolerance towards organic solvents, elevated temperatures, and chemical degradation. Many different approaches of this type have been reviewed extensively: these include bioinformatics approaches,^[14,15] natural evolution based engineering,^[16] host–guest and supra-molecular chemistry,^[13] directed evolution,^[17–21] catalytic antibodies,^[22–24] organocatalysis,^[25] rational structure-based protein engineering,^[26] and computational protein design.^[27]

In this Review we describe the computational “inside-out” approach to enzyme design, and the beginnings of what we strive to develop into a robust technology to make catalysts for synthesis, biotechnology, and therapeutics. The idea behind our computational design strategy is to utilize biochemical building blocks (amino acids, cofactors, co-

enzymes, etc.) to produce catalysts for nonbiological processes that can be made by microbiological techniques. The recent surge in computational power has spurred an increase in the development and testing of improved structure prediction and conformational search algorithms. Quantum mechanical methods lead to predictions of the arrangements of functional groups that maximize the binding and stabilization of the transition states of the desired reaction. If a protein can be designed that will fold into the necessary 3D geometry, catalytic conversion of non-natural chemicals into product(s) should be possible. To avoid having to predict the stability of new sequences from scratch, we incorporate the designed active site into stable protein folds. Furthermore, we adapt as much of the active-site components from natural precedent as is possible for the non-natural reaction or substrate.

We discuss the “inside-out” protocol, highlight approximations and bottlenecks, explore examples of successful design, examine achievements and challenges, and present cases in which variations and additions to the original design protocol were beneficial. We conclude that molecular dynamics (MD) simulations, post-design directed evolution, and



Gert Kiss studied chemistry in Heidelberg, Germany, carried out research in the lab of Noah W. Allen at UNCA, and then went on to pursue a PhD with K. N. Houk at UCLA. As a graduate student he was an NIH-CBI fellow, an LLNL Lawrence Scholar, and a recipient of the Stauffer Research Award. He is currently an NIH Simbios postdoctoral fellow with Vijay S. Pande at Stanford.



David Baker received his PhD with Randy Schekman at UC Berkeley and carried out postdoctoral studies in biophysics with David Agard at UCSF. He is currently a Professor of Biochemistry at the University of Washington and an Investigator at the Howard Hughes Medical Institute. He is a member of the National Academy of Sciences and the American Academy of Sciences.



Nihan Celebi-Ölçüm received a BS in chemistry at Boğaziçi University, Turkey, and an MS in computational and theoretical chemistry at Université Henri Poincaré in Nancy, France. She received her PhD with Viktoriya Aviyente at Boğaziçi University and then carried out postdoctoral research with Kendall Houk at UCLA. Recently she joined the faculty at Yeditepe University, Turkey.



K. N. Houk received his PhD with R. B. Woodward at Harvard, and then taught at Louisiana State University and the University of Pittsburgh before joining UCLA in 1986. He is now the Saul Winstein Chair in Organic Chemistry. He was Director of the Chemistry Division of the National Science Foundation from 1988–1990. He is a member of the National Academy of Sciences, the American Academy of Arts and Sciences, and the International Academy of Quantum Molecular Sciences.



Rocco Moretti received his BS in biochemistry at Worcester Polytechnic Institute, and his PhD in biochemistry at the University of Wisconsin with Aseem Ansari and Jon Thorson. He is currently a senior fellow in the research group of David Baker at the University of Washington.

crowd-sourced redesign can lead to improved efficiencies. We begin by discussing some of the protein engineering approaches that have paved the way.

2. Protein Engineering

2.1. Catalytic Antibodies

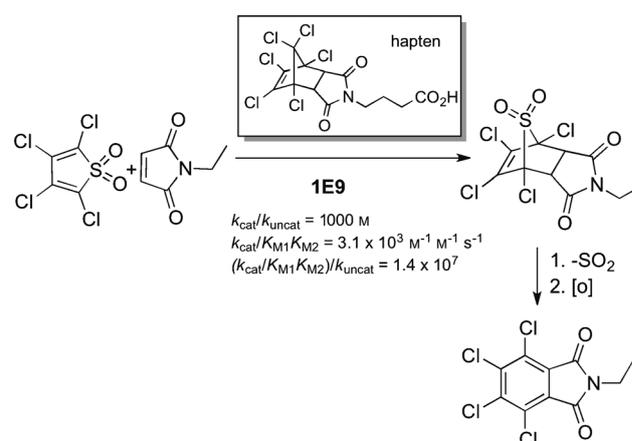
Following the pioneering work of the research groups of Lerner and Schultz in the mid-1980s,^[28,29] catalytic antibodies have been produced for a wide range of chemical transformations.^[24,30] The concept is based on Pauling's hypothesis that enzymes provide an environment complementary in structure and electronic distribution to that of the rate-limiting TS.^[2,31] When challenged with a hapten that resembles the key TS characteristics for a given reaction, antibodies are produced that can bind the hapten and thus also the TS it mimics. Transition-state binding equates to a lowered reaction barrier and thus to an increased turnover rate compared to the uncatalyzed reaction in solution.^[32] The production of catalytic antibodies takes advantage of the rapid rates of mutation and selection against a specific antigen that is a key characteristic of adaptive immune responses. The resulting binding interactions are specific and can be harnessed to catalyze non-natural reactions, and also to promote the conversion of non-natural substrates. Quantum mechanical computations are useful for the design of transition-state analogues (TSAs) that can serve as haptens for a given reaction.^[33] Janda and co-workers reviewed strategies and challenges in the development of new haptens.^[24]

Among the reactions that have been catalyzed by antibodies are Diels–Alder cycloadditions, acyl transfer reactions, oxy-Cope rearrangements, and cyclizations. Catalytic proficiencies range from $K_{\text{ix}}^{-1} = 10^{4.6} \text{ M}^{-1}$ to $K_{\text{ix}}^{-1} = 10^{8.6} \text{ M}^{-1}$, and so the transition states of the reactions they catalyze are bound more strongly than the substrates ($K_{\text{M}}^{-1} = 10^{3.5 \pm 1.0} \text{ M}^{-1}$).^[13,30] Nature's enzymes, on the other hand, exert massive K_{ix}^{-1} values, with an average range of 10^{12} to 10^{20} M^{-1} . Exceptional cases, such as ODCase and alkyl sulfatases, display K_{ix}^{-1} values of 10^{24} M^{-1} and 10^{29} M^{-1} , respectively.^[8,34] Naturally, the catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) of such enzymes is often limited only by the diffusion rate of the substrate and ranges from 10^4 to $10^9 \text{ M}^{-1} \text{ s}^{-1}$. In comparison, catalytic antibodies fall short of this limit by 4 orders of magnitude or more ($k_{\text{cat}}/K_{\text{M}} = 10^2$ – $10^5 \text{ M}^{-1} \text{ s}^{-1}$).^[13,24,34] This has been attributed to various factors, including product inhibition,^[35–37] lower binding constants,^[13] lack of covalent binding and catalysis,^[7] smaller buried surface area,^[13] differences in timescales of evolution,^[22] and inadequacies of the immunoglobulin fold.^[22] The comparatively low stability of the immunoglobulin fold and high cost of producing antibody catalysts further limit their applications in industrial settings. Nonetheless, there is increasing interest in their potential for therapeutic applications, such as neutralizing HIV-1,^[38,39] antibody-directed enzyme prodrug therapy (ADEPT), and the inactivation of addictive substances through the antibody-mediated breakdown of drug molecules.^[40]

While catalytic antibodies often suffer from product inhibition and can generally not be programmed for elaborate arrays of catalytic functionality, they have provided researchers with an important toolkit for the study of biocatalytic processes. In the following subsections, we highlight examples and discuss the role of computations in designing and understanding antibody catalysis.

2.1.1. Diels–Alder Reaction

The first catalytic antibody for a Diels–Alder reaction (1E9) was reported by Auditor and co-workers in 1989.^[41] It catalyzes the cycloaddition between tetrachlorothiophene and *N*-ethyl maleimide with a rate enhancement ($k_{\text{cat}}/k_{\text{uncat}}$) of 1000 M (Scheme 1).^[42] The crystal structure reveals a mostly

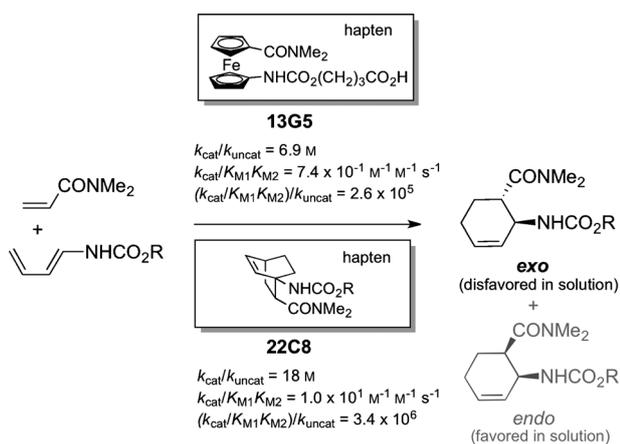


Scheme 1. Antibody-catalyzed cycloaddition of tetrachlorothiophene and *N*-ethyl maleimide.

hydrophobic binding pocket with a single polar residue (Asn^{H85}). Chen et al. studied 1E9 through a combination of QM calculations, docking studies, molecular dynamics simulations, and a linear interaction energy approach.^[43] The active site of 1E9 offers high shape complementarity to the TS geometry and offers electrostatic interactions that favor the TS over the reactants.

The Diels–Alder cycloaddition provides the opportunity to aim beyond catalytic rate accelerations and towards achieving stereochemical control. Gouverneur et al. demonstrated this in a spectacular way: The cycloaddition between *trans*-1-*N*-acylamino-1,3-butadiene and *N,N*-dimethylacrylamide affords a mixture of *endo* and *exo* stereoisomers under thermal conditions. QM calculations were used to elucidate the characteristics of stereoisomeric transition states and to design transition-state analogues for the *endo* and *exo* pathways.^[44]

Antibody 13G5 catalyzes the disfavored *exo*-Diels–Alder reaction between methyl *N*-butadienyl carbamate and *N,N*-dimethylacrylamide ($k_{\text{cat}} = 1.20 \times 10^{-3} \text{ min}^{-1}$, $k_{\text{cat}}/k_{\text{uncat}} = 6.9 \text{ M}$), and yields a single enantiomer in high enantiomeric excess (Scheme 2).^[45] The crystal structure and QM calculations showed that Asp^{H50} and Tyr^{L36} account for most of the catalytic effect of 13G5, while Asn^{L91} better stabilizes the

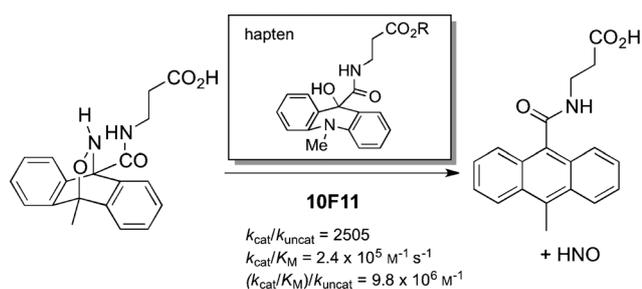


Scheme 2. Antibody-catalyzed disfavored *exo*-Diels–Alder reaction of methyl *N*-butadienyl carbamate and *N,N*-dimethylacrylamide.

ground state, and slightly retards the reaction.^[45,46] This finding suggests that Asn^{L91} provides a structural framework for the antibody to orient the substrates rather than having a catalytic effect.

The absolute configuration of the product was determined experimentally to be *exo*-(3*S*,4*S*). The specificity of the reaction was initially explored by docking the transition state into the crystal structure of antibody 13G5.^[45] MD relaxation of the antibody around the frozen TS revealed that the catalytic base (Asp^{H50}) can be coordinated by one and three water molecule(s) in the presence of the *exo*-(3*S*,4*S*) and *exo*-(3*R*,4*R*) TS, respectively,^[46] and the interaction of the catalytic Asp^{H50} with the carbamate NH group is significantly weakened in the *exo*-(3*R*,4*R*) pathway.

Antibody 10F11 catalyzes a retro-Diels–Alder reaction that liberates HNO with a $k_{\text{cat}}/k_{\text{uncat}}$ value of 2500 (Scheme 3).^[47] Inspection of the crystal structure of 10F11



Scheme 3. Antibody-catalyzed retro-Diels–Alder reaction.

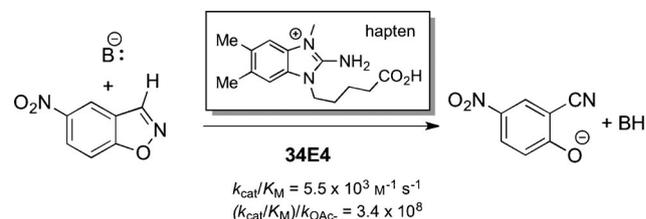
suggested good shape complementarity with the TS, and identified specific active-site residues (Trp, Phe, Ser) that were proposed to contribute to catalysis.^[48] Density functional theory (DFT) calculations were employed on models of the active site to study the interactions that can stabilize the transition state.^[49]

Kim et al. reviewed and compared the noncovalent catalysis of Diels–Alder reactions by cyclodextrins, self-assembling capsules, antibodies, and RNAses, and concluded

that—unlike enzyme catalysts—none of these hosts provide substantial specific binding of the transition states.^[50]

2.1.2. Kemp Elimination

The first catalytic antibody to catalyze the ring opening of 5-nitrobenzisoxazole (Kemp elimination) was reported by Hilvert and co-workers in 1996 (Scheme 4).^[51] 34E4 displays



Scheme 4. Antibody-catalyzed ring opening of 5-nitrobenzisoxazole.

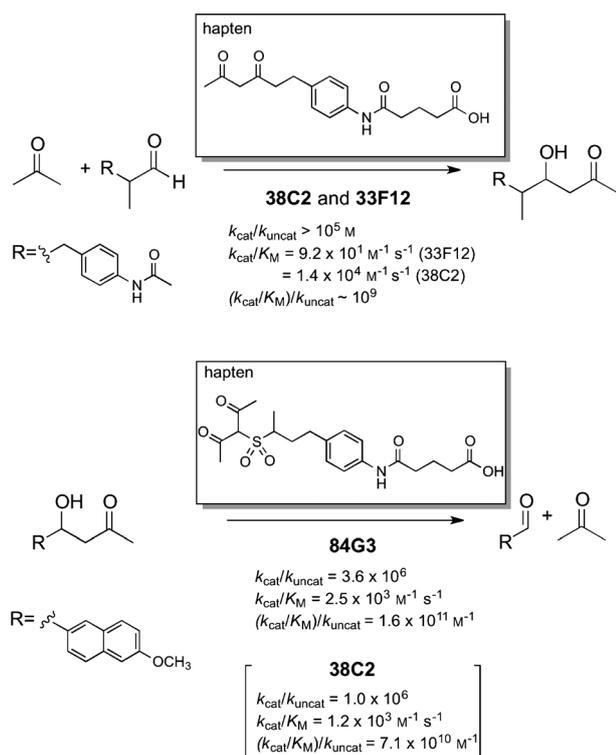
a $k_{\text{cat}}/K_{\text{M}}$ value of $5.5 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ and a $k_{\text{cat}}/k_{\text{uncat}}$ value of 2.1×10^4 compared to the uncatalyzed reaction ($k_{\text{uncat}} = 3.1 \times 10^{-5} \text{ s}^{-1}$), and a $(k_{\text{cat}}/K_{\text{M}})/k_{\text{OAc}^-}$ value of 3.4×10^8 compared to the rate of the acetate-promoted reaction in water ($k_{\text{OAc}^-} = 1.6 \times 10^{-5} \text{ M}^{-1} \text{ s}^{-1}$). Both experimental and computational investigations demonstrated that, similar to other proton-transfer reactions, the Kemp elimination is sensitive to the geometry in which the substrate and base are aligned.^[52,53] Furthermore, when a carboxylate functions as the catalytic base, Kemp elimination reactions are also highly sensitive to the polarity of the solvent.^[54,55] These features in combination with the simplicity of the reaction and the ease with which progress can be monitored (UV/Vis), has made the Kemp elimination a frequently studied model for base-catalyzed biochemical transformations.

2.1.3. Aldol/Retro-Aldol Reaction

QM calculations on aldol reactions date back to the 1980s,^[56] and offer early geometric descriptions of the transition state. Aldolase antibodies ab38C2, ab84G3, and ab33F12 were later raised against TS-analogous haptens and can catalyze aldol and retro-aldol reactions with activities comparable to natural aldolases, but with a broader substrate scope (Scheme 5).^[57–61] These antibodies resemble class I aldolases that utilize the ϵ -amino group of an active site lysine residue to form a Schiff base with the substrate. Polar residues outline the otherwise hydrophobic active site at distances that range between 5 and 7 Å from the ϵ -amino group of Lys^{H93}. In a QM study, Arnó and Domingo investigated the role of some of these residues as potential general acid catalysts in the C–C bond-formation step.^[62]

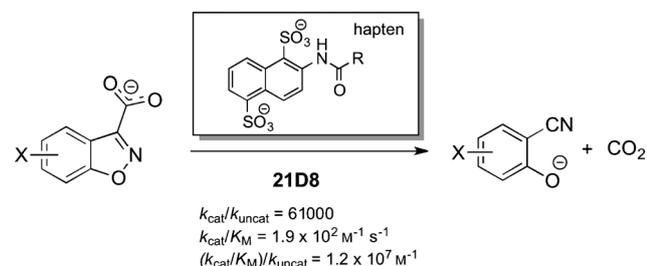
2.1.4. Decarboxylation-Catalyzed Ring-Opening Reaction

The rate of the decarboxylation reaction of 5-nitro-3-carboxybenzisoxazole varies by up to eight orders of magnitude depending on the solvent polarity.^[54,55] Aprotic polar solvents promote the reaction by desolvating the carboxylate



Scheme 5. Examples of antibody-catalyzed aldol and retro-aldol reactions.

reactant and by stabilizing the transition state through dispersion interactions.^[63–66] Antibody 21D8 catalyzes the decarboxylation of 5-nitro-3-carboxybenzoxazole^[67] by up to 61000-fold over the background reaction in water (Scheme 6).^[68]

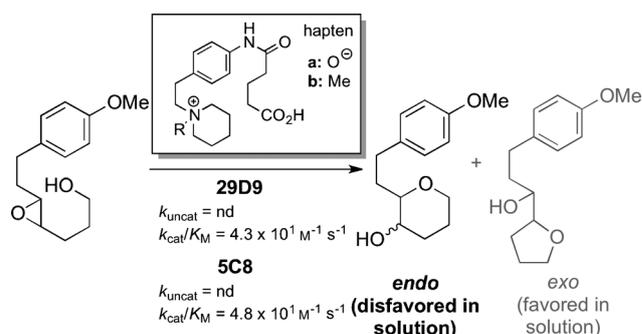


Scheme 6. Antibody-catalyzed decarboxylation reaction of 5-nitro-3-carboxybenzoxazole.

QM, MD, and free-energy perturbation (FEP) calculations were performed to explore the origins of catalysis further,^[69] and showed that partial solvation of the carboxylate group was detrimental to catalysis, but this was countered by favorable hydrogen-bonding interactions with the isoxazole oxygen atom.

2.1.5. Cyclization of *trans*-Epoxy Alcohols

Four different antibodies were raised with two different haptens to catalyze the disfavored *endo*-tet cyclization



Scheme 7. Antibody-catalyzed disfavored *endo*-tet cyclization reactions of *trans*-epoxy alcohols. nd = not determined.

reaction of *trans*-epoxy alcohols (Scheme 7).^[70,71] Quantum mechanical calculations show that the *endo*-tet transition state has S_N1 character and can be stabilized electrostatically by a carboxylate.^[72] The hypothesis was tested with QM models that consisted of motifs using concurrent general acid/base catalysis.^[73,74] Analogous calculations for 6-*exo* and 7-*endo* cyclizations predicted the preferential formation of the seven-membered product. This was verified experimentally,^[75] and is also in line with the X-ray structure eventually obtained for antibody Fab 5C8.^[71] The active site contains an Asp^{H95}–His^{L89} dyad that appears to be poised for acid/base catalysis.

2.1.6. Hydrolysis of Aromatic Amides and Esters

Antibody 43C9 catalyzes the hydrolysis of aromatic amides and esters with an unusually efficient $k_{\text{cat}}/k_{\text{uncat}}$ value of 2.5×10^5 .^[30,76–78] Getzoff and co-workers built a computational homology model of the antibody's variable region and proposed that Arg^{L96} functions as the oxyanion hole, while His^{L91} is the catalytic nucleophile.^[79] Subsequently, the X-ray structure of 43C9 was solved and supported the predictions, showing that a water-mediated hydrogen-bonding network in the active site is important for catalysis.^[80] Kollman and co-workers later performed QM calculations, MD simulations, and free-energy calculations on 43C9 and proposed a direct hydroxide attack as an alternative to the mechanism involving nucleophilic catalysis by His^{L91}.^[81]

2.1.7. Chorismate–Prephenate Rearrangement

The Claisen rearrangement of chorismate to prephenate is catalyzed by natural chorismate mutase enzymes,^[82,83] and by the catalytic antibodies IF7 and IIF1-2E11.^[84,85] Wiest and Houk employed QM calculations to mimic the active site of chorismate mutase and proposed that specific hydrogen-bond donors were responsible for the approximately 200-fold rate acceleration displayed by catalytic antibody IF7.^[86]

2.2. Directed Evolution

Directed or laboratory evolution has become one of the more mature forms of protein engineering and has found its way into modern industrial-scale applications.^[87] It is a power-

ful and commonly used approach to enzyme engineering that relies on iterative cycles of mutagenesis and selection.^[17,18] Examples of its application include improved thermostability,^[88] tolerance to organic solvents,^[89] strengthened protein–protein interactions,^[90] altered substrate promiscuity/specificity,^[91,92] enhanced enzymatic activity,^[93] and inversion of enantioselectivity.^[94,95] In the directed evolution of catalytic function, a starting gene is mutagenized to create a library of variants, which is screened for enzymes with an improvement of the sought-after property (stability, substrate specificity, activity, etc.). Typically the improvements in any one round are small, and the process is repeated many times.

Strategies for the construction of libraries include random whole-gene error-prone PCR or random mutagenesis (Figure 1 a), site saturation or targeted mutagenesis (CASTing,

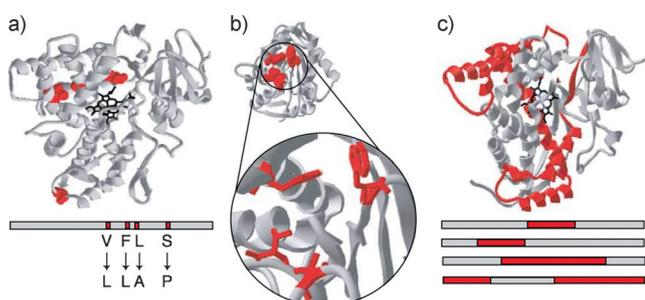


Figure 1. Three strategies for creating protein libraries by directed evolution. a) Random mutagenesis across the full sequence. b) Targeted mutagenesis that is focused on a specific site. c) Protein sequence recombination for the replacement of entire segments. Reprinted from Ref. [19], with permission.

ISM)^[95] (Figure 1 b), and the generation of chimeras through sequence recombination (Figure 1 c).^[19] A key strength of random mutagenesis is that no structural or mechanistic information about the enzyme is required and that beneficial mutations can be uncovered at unexpected positions distant from the active site.^[96]

Site saturation or targeted strategies, on the other hand, focus on certain areas of an enzyme (i.e. the active site) and require prior structural or biochemical knowledge about the protein. Reducing the randomizable sequence space increases the probability with which multiple beneficial mutations can be uncovered within the active site.^[97] The approach is of value when dramatic alterations to an enzyme's function are sought or when improved function depends on a combination of active-site variations.

Beneficial mutations within a library can be identified, for example, through statistical analysis of protein sequence–activity relationships (ProSAR),^[98] then combined and incorporated by gene shuffling.^[99] Molecular and functional diversity can be further expanded with neutral drift libraries, in which mutations are accumulated that are orthogonal to the function and stability of the enzyme.^[100,101]

A key challenge for directed evolution is the identification of individual variants that display the desired improvements out of a large set of randomized protein sequences.^[20] Selection-based *in vitro* techniques, such as mRNA display^[102]

and emulsion-based microfluidic FADS (fluorescence-activated droplet sorter),^[103] exhibit substantial throughput. Screening-based techniques that measure substrate or product concentrations are the most versatile, but are also more limited in their throughput.^[104] Once a genotype–phenotype link is established, directed evolution can work with all biologically produced proteins, including those that contain non-natural amino acids or non-natural prosthetic modifications.^[105]

Some experiments have involved completely naive starting points, but directed evolution works best for enzymes that display some level of activity towards the desired reaction or towards a highly similar one.^[19] While the success of directed evolution programs depends on a clear, uphill path from the starting point to a highly active variant,^[105] most protein sequences do not display the desired initial activity. This challenge can be overcome somewhat through neutral drift libraries and by gradually changing the selective pressure from the existing function to the desired one.^[20,21]

Many attempts have been made to engineer and redesign proteins and enzymes over the past few decades. Those that met with success, employed variations of directed evolution ranging from random mutagenesis to semirational or focused library-generating strategies and sophisticated statistical selection such as ProSAR in specific cases. Two recent examples are the asymmetric synthesis of chiral amines for the industrial production of the type-2 diabetes drug sitagliptin (Januvia)^[106] and the oxidative desymmetrization of the prochiral amine for the production of the hepatitis C drug Boceprevir.^[107] In other cases, computational approaches resulted in significant advances in understanding the mechanism by which directed evolution can change the enantioselectivity of an enzyme.^[108,109] In the past 5 years alone, over 60 articles were published that reported on enhancing the thermostability, substrate and cofactor specificity, enantioselectivity, and reaction rate of natural enzymes. Many of these were engineered for applications in asymmetric organic synthesis, and include transaminases, enoate reductases, esterases, monoamine oxidases, dehalogenases, and aldolases, as well as cytochrome P450s (oxidations and epoxidations) and Baeyer–Villiger monooxygenases. The topic has recently been the subject of several excellent reviews.^[87,110–113]

2.3. Natural Evolution and Enzyme Redesign

Nature has experimented with ways to generate new catalytic functions for billions of years. The study of these strategies can provide us with insights that can be used to make educated mutations to native active sites with the goal of eliciting new functions. Enzymes that belong to mechanistically diverse superfamilies are valuable starting points for such redesign efforts, particularly when they are structurally conserved among one another. Members of such superfamilies are frequently also promiscuous and one enzyme often catalyzes a number of chemical transformations, albeit at much lower rates than its physiological reaction.^[16,114]

One successful redesign approach has been to enhance promiscuous functions based on sequence and structure

alignments. Here, the redesign is based on a template enzyme with innate activity for the target reaction. Information on a naturally existing enzyme that is known to promote the target reaction is then applied for the redesign of the template enzyme. Fersht and co-workers, for example, compared the sequence of *N*-acetylneuraminidase (NAL) to that of the homologous dihydrodipicolate synthase (DHDPS) and identified a Leu-Arg mismatch in the active site.^[115] The Leu142Arg mutant was made (along with a number of stabilizing mutations for the new Arg). This switched the activity of NAL from its native retro-aldol cleavage (*N*-acetylneuraminidase to pyruvate and *N*-acetyl-D-mannosamine) to that of DHDPS (condensation of pyruvate with L-aspartate- β -semialdehyde). The native retro-aldol activity of NAL was abolished, while the rate of the innate DHDPS of the NAL was increased eightfold. Similarly, a Leu to Arg mutation switched the physiological activity of 4-oxalocrotonate tautomerase (4-OT) to that of *trans*-3-chloroacrylate dehalogenase (CaaD).^[116] Structural studies revealed only minor geometric changes. the k_{cat} value of the CaaD activity of the 4-OT was increased 9-fold; $k_{\text{cat}}/K_{\text{M}}$ increased 50-fold. A somewhat more ambitious study introduced four mutations into the active site of keto-L-gulonate 6-phosphate decarboxylase (KGPDC) to increase the rate of its promiscuous activity for the D-arabinose-hex-3-ulose 6-phosphate synthase (HPS) 170-fold.^[117,118]

Although similar to the above, the redesign of an enzyme towards a reaction for which it does not possess any promiscuous activity is a grander challenge. Sequence and structure alignments with members of the same superfamily here too form the basis for redesign. Gerlt and co-workers combined a rational mutation with directed evolution for the redesign of L-Ala-D,L-Glu epimerase (AEE) and muconate lactonizing enzyme (MLE), respectively.^[119] The efforts were aimed at introducing OSBS (*o*-succinyl benzoate synthase) activity into AEE and MLE, neither of which shows promiscuity towards the OSBS reaction. The feat was achieved by altering the substrate specificity: a single mutation (Asp-Gly and Glu-Gly) allows AEE and MLE to accept the OSBS substrate, which readily reacts with the unchanged catalytic residues to give *o*-succinyl benzoate. Ohta and co-workers went a step further and generated α -aryl propionate racemase activity in the homologous aryl malonate decarboxylase (AMD) by introducing a catalytic acid/base into the active site (Gly74Cys).^[120] In a separate study, the enantioselectivity of that same decarboxylase was inverted by using a double mutant (Gly74Cys, Cys188Ser).^[121] The mutant gives (*R*)- α -thienyl propionate in a yield of 60% and an enantiomeric excess of 84% *ee*, but also displays an approximately 600-fold lower activity than the wild-type AMD. Random mutagenesis improved the k_{cat} value 10-fold and decreased the gap to the wild-type AMD to a 60-fold drop in activity.^[122] Dunaway-Mariano and co-workers impressively showed that function can be transplanted within the crotonase superfamily by replacing a His-Asp dyad with a Glu-Glu acid/base pair.^[123] Two glutamates were introduced into the 4-CBA-CoA dehalogenase active site and six additional mutations were necessary to give a fully soluble and stable protein. With a k_{cat} value of 0.064 s⁻¹, the octamutant activity is far below

that of the wild-type crotonase ($k_{\text{cat}} = 1000 \text{ s}^{-1}$), but the exercise shows that “an entirely new catalytic pathway can be created at the expense of the pre-existing pathway through a limited number of amino acid substitutions”.

The redesign of enzyme superfamily members is useful in deciphering the strategies and principles that guide the natural evolution of catalytic biomolecules. The chemical versatility that is accessible to the protein engineer by this route, however, is limited to the generally narrow range of reaction types within a superfamily. The crotonase superfamily (CS) is a notable exception in which “nature has varied common structural features to evolve catalysts for a remarkably diverse set of reactions”^[124] spanning all six classes of reaction defined in the Enzyme Commission (EC) classification scheme.

2.4. Rational and De Novo Protein Design

2.4.1. Design and Prediction of Protein Folds

More drastic engineering approaches that are based on a variety of computational techniques have led to the redesign of entire proteins. Early work in the field focused on the redesign of helical bundles,^[125] and employed strategies that aim at generating specific hydrophobic/hydrophilic patterns—a primary determinant for the orientation and register of helical bundles. The approach gave some sense of control over the formation of a fold without necessitating the prediction of specific side-chain orientations.^[126–128] Further work extended the computational design approach to protein structures with less-regular geometries.^[129] The general applicability of computational protocols, such as that of RosettaDesign, was tested by re-engineering a diverse set of nine small globular proteins.^[130] The computational design of proteins of complex topology is assisted by techniques such as dead-end elimination and Monte Carlo sampling that can attempt to pack side chains in their minimal energy positions. The scope of computer-based engineering is not limited to the redesign of existing topologies. Kuhlman et al., for example, iterated between sequence design and structure prediction to access novel protein folds, and produced the Top7 α/β topology.^[131] In contrast to previous design procedures that treated the backbone as rigid and require a vast conformational space to be sampled, the design of Top7 was possible in part because of a flexible backbone minimization step in the iterative protocol.

2.4.2. Protein-Protein Interactions

Computational approaches have also been employed for the design of protein-protein interactions. Huang et al. achieved micromolar binding affinities by using a design minimization approach in which the best amino acid identities and rotamers were predicted for the protein-protein interface.^[132] Improved affinities were obtained when the naturally occurring protein-protein interfaces were used as a guide.^[90] Key residues that are thought to account for the bulk of the binding affinity are chosen as “hot spots” and placed in locations likely to maximize binding. The rest of the interface

is then “filled in” to maximize packing around these key interactions and yielded a binding affinity of 130 nM. The somewhat more challenging task of designing a single binder to a fixed, biologically relevant partner gave rise to a computational design with a binding affinity of 200 nM.^[133] Directed evolution further improved the binding constants to 180 pM and 4 nM. Analysis of the mutations suggest that the computational designs could be improved by accounting for backbone flexibility, as well as improved electrostatic and solvation models.^[133] More recently, DeGrado and co-workers utilized their computational design approach CHAMP (computed helical antimembrane protein method)^[134] to produce a helical β peptide that targets a transmembrane helix of the integrin $\alpha_{\text{IIb}}\beta_3$.^[135] The DeGrado research group further showcased the utility of computational design approaches by generating helical protein assemblies along carbon nanotubes.^[136]

2.4.3. DNA Binders

The design of DNA binders is another interesting direction of computational protein design. One technique is to combine preexisting DNA binding modules by redesigning the intermodule interfaces, thereby reducing the problem to a design of protein–protein interactions.^[137] More targeted changes were made in the computational redesign of homing endonucleases that can recognize a single base pair difference.^[138,139] The design for recognition of multiple base pair changes has also been demonstrated.^[140] The simultaneous introduction of multiple adjacent base pair changes proved more successful than a stepwise combination of mutations from individual base pair changes. The design of such sequence specificity changes has been used in the case of the homing endonuclease I-AniI to probe the role of DNA sequence in binding and catalysis.^[141]

2.4.4. Protein–Ligand Interactions

Early work on protein–small-molecule binding appeared promising, with reports of binders for metal,^[142,143] lactate,^[144] serotonin,^[144] TNT,^[144] and nerve agents.^[145] However, doubts arose when the periplasmic binding proteins designed for lactate, serotonin, TNT, and nerve agents did not show ligand binding when assayed by isothermal calorimetry (ITC) or NMR spectroscopy.^[146] It is thought that the initial reports of success may have arisen from the reliance on an indirect environmentally sensitive fluorescence-based readout. While not a solved problem, some progress has been made on protein–small-molecule binding. Boas and Harbury applied computational design to periplasmic binding proteins and found that native site recapitulation required high-resolution rotamer sampling, continuous minimization, and accurate electrostatic calculations.^[147,148] DeGrado and co-workers were able to create an α -helical bundle which was able to bind a heme-like cofactor.^[148] Recent attempts at redesigning a dipeptide binder, on the other hand, were unsuccessful, presumably because of inadequately accounting for the binding site flexibility.^[149]

2.4.5. Catalytic Peptides and Proteins

Early examples for the de novo design of chemical functions include, but are not limited to, the following studies:

Johnsson et al. designed a metal-free oxaloacetate decarboxylase (oxaldie) that operates through an imine mechanism by incorporating a reactive amine onto an amphiphilic α -helix.^[150] Designed oxaldies catalyze the decarboxylation of oxaloacetate with a $k_{\text{cat}}/K_{\text{M}}$ value of $0.63 \text{ M}^{-1} \text{ s}^{-1}$. The rate of imine formation is found to be three to four orders of magnitude larger with oxaldie than with simple amine catalysts and comparable to catalytic antibodies (10^3 – 10^6).^[151]

Sasaki and Kaiser designed “helichrome”, an artificial hemeprotein, in which four amphiphilic α helices were covalently tethered to one face of the porphyrin ring to create a hydrophobic pocket for substrate binding. The Fe^{III} complex of helichrome showed hydrolase activity and converted aniline into *p*-aminophenol in the presence of NADPH with a $k_{\text{cat}}/K_{\text{M}}$ value of $1.67 \text{ M}^{-1} \text{ s}^{-1}$.^[152]

Broo et al. designed a hairpin helix–loop–helix motif that dimerizes to form four-helix bundles and that utilizes histidine residues to catalyze the acyl-transfer reaction of activated esters.^[153] Rossi et al. inserted two and four copies of the artificial triazacyclononane amino acid into three distinct helix–loop–helix peptides. They generated Zn^{II} binding sites capable of catalyzing the transesterification of an RNA model substrate up to 380-fold.^[154]

Dutton and co-workers used a tryptophan and a tyrosine radical maquette, $\alpha_3\text{W}^1$ and $\alpha_3\text{Y}^1$,^[155] as models of radical enzymes to study how side-chain radicals are generated, controlled, and directed towards catalysis.^[156] In more recent work, Pecoraro and co-workers utilized $\alpha_3\text{D}$ as a scaffold for the placement of three cysteine residues that are capable of binding the heavy metal ions Cd^{II} , Hg^{II} , and Pb^{II} .^[157]

DeGrado and co-workers described the catalysis of an O_2 -dependent phenol oxidase reaction by de novo diiron model proteins based on the four-chain heterotetrameric helical bundle DF_{tet} .^[158] The most active variant catalyzes the oxidation of 4-nitrophenyl acetate with a 1000-fold rate enhancement.

Bolon and Mayo used a “compute and build” strategy to incorporate hydrolase activity onto a catalytically inert *E. coli* thioredoxin scaffold. The resulting PZD2 utilizes a nucleophilic histidine to promote the hydrolysis of *p*-nitrophenyl acetate 180-fold.^[159]

3. The Inside-out Approach to Computational Enzyme Design

In recent years, computational algorithms have become increasingly reliable for identifying amino acid sequences compatible with a target tertiary structure. Efforts towards solving the inverse protein folding problem^[160–163] reached a milestone with the design and successful experimental proof of the structure of the 93-residue α/β protein Top7.^[131] This showed that, for an arbitrary fold, it is possible to use computational methods to predict sequences that would produce that stable fold. While a great deal remains to be

done in this area, another great challenge is to create functional proteins that can promote non-natural chemical reactions.

A collaborative effort between the research groups of Baker and the Houk has led to the development of an “inside-out” protocol towards this goal (Figure 2). At the core of the

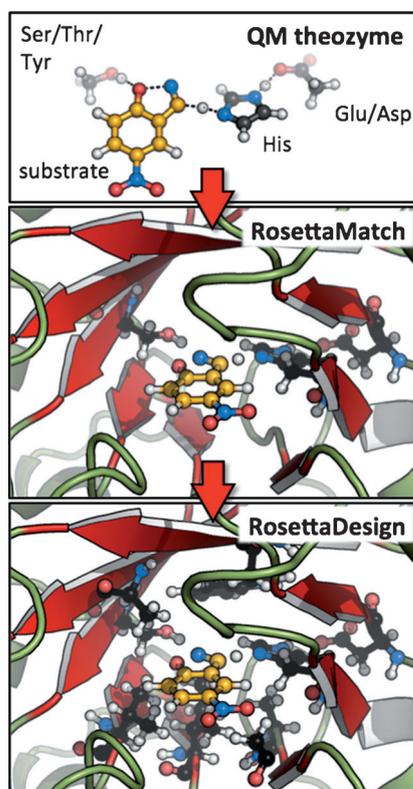


Figure 2. Key steps in the computational inside-out design protocol (shown here for the Kemp elimination): from QM theozyme, to match, to design.

computational design protocol is a theoretical active site (theozyme, Figure 2, top panel) with the appropriate functionality for catalysis. Here, quantum mechanical (QM) calculations are employed to determine the catalytic units that will be most effective at stabilizing the transition state (TS) in a precise geometrical arrangement. Protein scaffolds are selected from the PDB (<http://www.rcsb.org>)^[164] and are used as templates into which the QM transition-state geometry is grafted (RosettaMatch,^[165] Figure 2, center panel). Amino acid residues surrounding the QM theozyme are mutated and optimized to ensure good packing and fold stability, and to complement the geometric and electronic features of the TS (RosettaDesign, Figure 2, bottom panel).

3.1. Theozymes

In the first step of the inside-out design protocol, QM calculations are carried out to generate three-dimensional arrangements of functional groups that are optimal for stabilizing the TS of the targeted reaction.^[166] A theozyme

(short for theoretical enzyme) is typically constructed from an array of amino acid side chains and backbone amides, but incorporation of unnatural amino acids and cofactors can further expand the chemical space. For a given reaction, a number of distinct theozyme motifs are usually generated, each of which varies in the composition of its functional groups. The energy profile of each motif is computed and the magnitude of catalysis is assessed. The theozyme motifs are further diversified geometrically by producing an ensemble of conformations without disrupting the catalytic interactions.

3.2. Incorporating Theozymes into Protein Scaffolds

RosettaMatch has been used to search the native active sites of existing protein structures for backbone positions that can accommodate the three-dimensional side-chain arrangement in a theozyme. The program “matches” the theozyme motif into the pocket by sequentially attaching each side chain of the theozyme to the backbone of the protein scaffold. Side-chain rotamers are generated for every position in the scaffold active site to which the functional groups of the theozyme are mapped. An ideal match is then one in which the exact three-dimensional geometry of the theozyme can be realized. Deviations from the optimum geometry by just a few tenths of an Angstrom and single-digit angles can lead to energetic penalties of up to 5 kcal mol⁻¹, which translates to four orders of magnitude in terms of the reaction rate (k_{cat}). In practice, an ideal match has not yet been obtained for any of the designed enzymes; a circumstance that can be attributed to the discrete nature of both the protein backbone and the primary matching algorithm as well as to the computational cost associated with the mapping out of conformational space. Matching then quickly becomes a bottleneck in the computational design protocol, particularly when a theozyme invokes three or more catalytic residues. Hence, an exact search typically does not give a single match and it becomes necessary to assign tolerance values to catalytic distances, angles, and dihedrals. The resulting matches are generally distorted from the theozyme geometry and necessitate some form of geometric filtering and ranking according to their theozyme-likeness. A useful utility for this purpose is EDGE (enzyme design geometry evaluation), which uses geometric hashing to compare theozyme atoms with a target structure and ranks matches based on the summation of their deviations.

SABER (selection of active/binding sites for enzyme redesign), a program developed by Houk and co-workers, offers an alternative to RosettaMatch: instead of placing theozymes into predefined active sites, SABER searches the Protein Data Bank (PDB) for proteins with the appropriate catalytic functionality already in place. When a suitable active site is found, only those amino acid residues need to be mutated that are required to accommodate the new substrate in its transition-state geometry. This stands in contrast to the RosettaMatch-based approach, where both the new catalytic functionality and the new substrate must be accommodated, generally requiring a larger number of mutations than the SABER-based approach.

3.3. Active-Site Design

After the theozyme has been attached to a scaffold protein, either by RosettaMatch or by SABER, the RosettaDesign module is used to restrain catalytic residues and to generate an optimal sequence/structure for the remainder of the active site. Rotamer sampling by Monte Carlo simulated annealing is used to optimize the identity and position of active-site residues, both in terms of their interactions with the theozyme and also with each other. To further refine the active site, this rotamer sampling is performed for multiple rounds, interspersed with minimization of the side chains, backbone, substrate conformation, and rigid body position. Throughout the process, the theozyme geometry is enforced through restraints. To ensure that the resultant sequence is intrinsically compatible with the theozyme, rather than being externally forced, a last cycle of repacking and minimization without the geometric restraints is commonly run.^[167] Ideally, these steps lead to the introduction of amino acid residues that add interactions to stabilize the positions of the key catalytic residues, tune their pK_a values, and optimize transition-state binding. In practice, each match that enters the active-site design stage contains a theozyme that is already significantly distorted compared to the ideal QM TS geometry. RosettaDesign is then tasked with generating the best possible stabilization for a geometry that in itself is non-ideal. While RosettaDesign attempts to constrain the design to the ideal theozyme, as specified by the geometric restraints, normally even the highest ranked final designs differ quite considerably from the original theozyme geometry. Figure 3 illustrates this point with four Kemp elimination designs that are superimposed onto the catalytic heavy atoms of their theozyme. The individual side chains cluster together in their general three-dimensional arrangements (Figure 3 a), but lack the precise positioning that naturally evolved enzymes display within a catalytic class (Figure 3 b).^[168]

3.4. Filtering, Ranking, and Evaluating Computational Designs

Prior to the experimental workup, final designs are assessed towards their capability to stabilize the key catalytic residues. They are ranked on the basis of empirical criteria such as Rosetta energy, ligand-binding scores, hydrogen bonding, active-site geometry, and packing scores. Comparison with the original scaffold protein plays an important role, as the native context forms a reference for what a well-folded protein looks like. Thus far, assessing the quality of final designs has relied heavily on the chemical intuition of the human designer for assessing how “enzyme-like” prospective designs are and for capturing properties that are currently not accounted for by the Rosetta scoring framework. Nature’s catalytic units are generally supported by frameworks of hydrogen bonds, steric packing, π - π stacking, limited dynamics, and limited water accessibility. At present, some of this is implicitly accounted for through various energy scores that penalize poor interactions and reward good ones throughout the design and repacking process. The explicit provision of supporting interactions for the catalytic unit can be viewed as

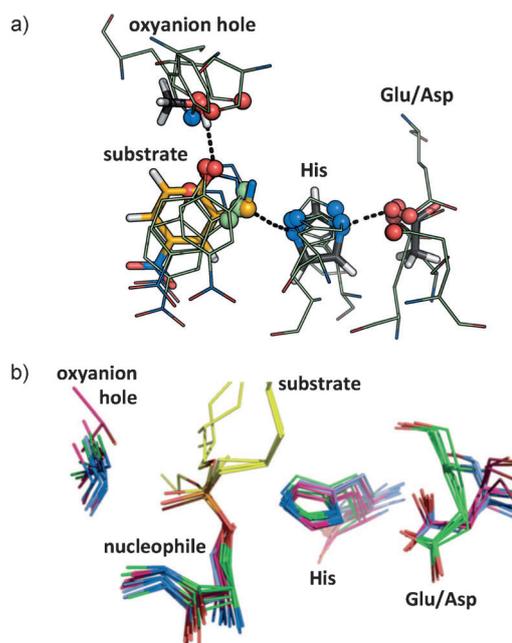


Figure 3. Geometric overlay of catalytic atoms. a) Theozyme (black/orange) over four final Rosetta designs in the TIM barrel fold (light green). The catalytic heavy atoms are highlighted as spheres. RMSD values of KE designs with the His–Glu/Asp dyad compared to theozyme: 1.2 Å for KE70 (second most active), 0.8 Å for KE38 (inactive), 0.8 Å for KE54 (inactive), 0.5 Å for KE66 (inactive). b) Catalytic triad from esterase crystal structures; RMSD = 0.45 Å within the same fold.^[168]

a second, and in a sense more challenging, stage in the design process, for which we are only now beginning to establish automated protocols. Increasingly, tools such as Foldit, EDGE, various in-house scripts, and more rigorous computational tests that probe the dynamics of the systems are being developed and refined with the goal of maximizing the success rate, particularly as more challenging reactions are pursued.

The design of Kemp eliminases and retro-aldolases, for example, was carried out with the first version of RosettaMatch and RosettaDesign. The scaffold set consisted of only 87 proteins, only a discrete matching algorithm was available, and the backbones of the proteins were treated as rigid. The assessment of designs was performed by manual inspection of the optimized final structures. The design of proteins towards catalysis of a bimolecular Diels–Alder cycloaddition was carried out with an updated version of Rosetta, using the discrete matching algorithm against a scaffold set of 227 proteins. Final designs were assessed both by manual inspection of the optimized geometries and by molecular dynamics simulations.

The current collection of Rosetta modules (Rosetta3) extends the scaffold set to the entire PDB, introduces a secondary, nondiscrete matching algorithm, which complements the primary one, and allows a small degree of backbone plasticity in response to a new active-site sequence. MD simulations were found to be valuable for assessing the structural integrity of a newly designed active site and for

exposing design flaws that are intractable from static evaluations.^[169]

The sequence of a final design generally differs by 10% or more from that of the wild-type template protein. Depending on the degree of the perturbation, the packing and hydrogen-bonding interactions within the modified protein are expected to be less ideal than those of the wild-type scaffold protein. Cycling through repacking and geometry optimization during the design process ensures that the overall conformation of the new protein is at a minimum of its potential energy landscape. However, neighboring minima may exist (corresponding to alternative conformations of side chains and loops) that could have become thermodynamically more favorable in the design process. The actual conformational state of a designed active site might thus differ significantly from that of the computational model—a possibility that can readily be investigated through molecular dynamics (MD) simulations. MD evaluations are now performed on a routine basis for finalized designs as a means to pinpoint structural weaknesses and to guide adjustments in the form of additional and/or alternative mutations.

3.5. Experimentation

Aside from the source of the genes (chemical synthesis versus cloning), experimental validation of computationally designed enzymes is much the same as activity measurements for any other enzyme.

3.5.1. Synthesis and Expression

In the case of the retro-aldolases,^[170] the Kemp eliminases,^[170,171] and the Diels–Alderases,^[172] the final optimized protein sequences were sent to a commercial gene synthesis company for typical codon optimization and cloning into a standard His-tagged *E. coli* expression vector. *E. coli* BL21(DE3) cells were then transformed with the plasmid, and the gene expressed under conventional IPTG or auto-induction conditions.^[173] Soluble protein can thus be obtained by conventional IMAC purification,^[174–176] along with gel filtration.

3.5.2. Enzyme Assays

One potential complication in assaying the activity of a designed enzyme is the low activity of most of the initial variants. Assays that can detect slightly above background levels of activity are thus preferred to identify these weak catalysts. Such assays are selected on the basis of the target reaction. The Kemp eliminases of Röthlisberger et al. and the retro-aldolases of Jiang et al. were designed for a reaction with a spectrophotometric shift, and continuous monitoring by UV/Vis spectroscopy over the course of over 10 min or 40 h, respectively, allowed for the detection of product formation. In contrast, the Diels–Alderases of Siegel et al. were designed against a reaction that was spectrophotometrically silent, so product formation was monitored by LC-MS, with time points taken over the course of several days. In this case, a chiral LC-

MS assay allowed for further characterization of the stereospecificity of the reaction, which showed that the catalyst was specific for the product configuration selected at the theozyme stage.

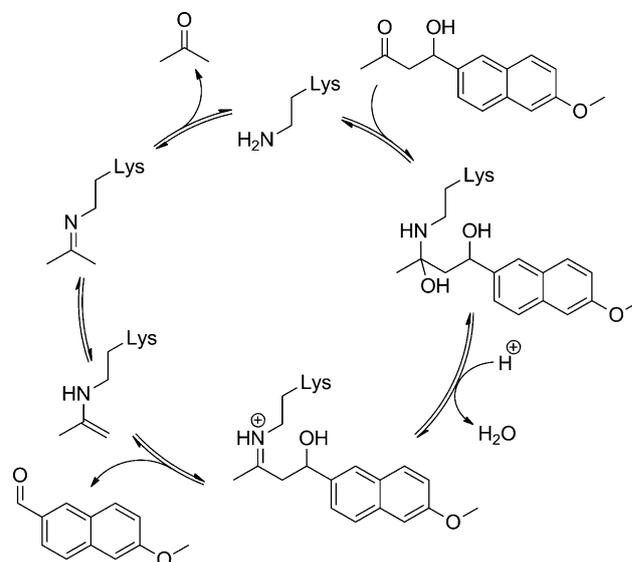
3.5.3. Directed Evolution

Typically, the initial successful designs have low activity. This low starting activity has been further improved through multiple rounds of directed evolution. A combination of random mutagenesis and targeted diversification has yielded improved activities. Further computational analysis also fed into this work, thereby allowing for the selection of potential mutations (including insertions) to incorporate during the rounds of selection. Specific examples are highlighted in Section 4.2.3.

4. Computational Enzyme Design—Achievements

4.1. Retro-Aldolases

The design of a novel retro-aldolase is the first example in which the computational inside-out approach was employed to construct a functional active site. The resulting retro-aldolases catalyze the C–C bond breaking in 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone.^[170] Analogous to the strategy used by type I aldolases, the reaction mechanism invokes a nucleophilic lysine and the formation of an iminium intermediate (Scheme 8).^[177]



Scheme 8. Steps in the amine-catalyzed retro-aldol reaction of 4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone.

The computational designs are based on four distinct theozymes (Figure 4). They feature a lysine as a Schiff base and a general acid/base (I: Lys/Asp dyad, II: Tyr, III: His/Asp dyad, IV: H₂O) for deprotonation of the β alcohol. The charged side chain (Lys-Asp-Lys) mediated proton transfer

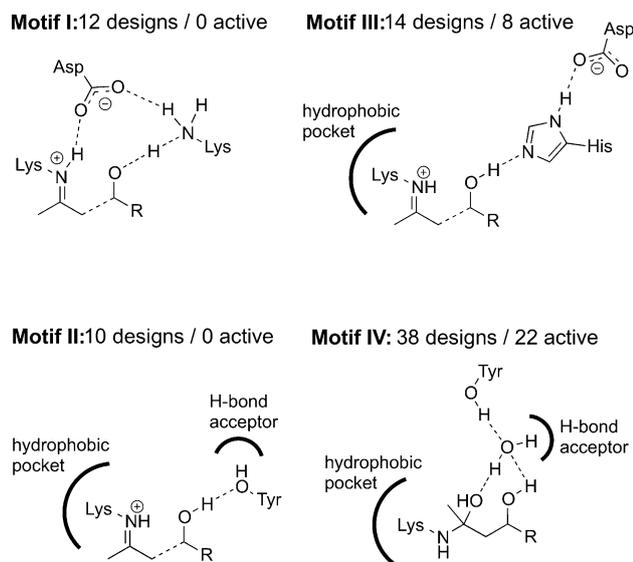


Figure 4. Retro-aldolase theozyme motifs.

scheme in motif I is analogous to that with D-2-deoxyribose-5-phosphate aldolase.^[178] Motifs II, III, and IV mimic the active sites of catalytic antibodies, in which a lysine is placed into a hydrophobic pocket to lower its pK_a value.

The geometries of the four active-site motifs were obtained from QM theozyme calculations, which were carried out for every step along the retro-aldol reaction path. The transition-state geometries were then combined to generate a composite active site that carries the geometric information of the complete reaction profile. The resulting consensus theozymes of the four motifs were further diversified by varying a) the internal degrees of freedom of the composite transition state, b) the orientation of the catalytic side chains with respect to the composite transition state within ranges consistent with catalysis, and c) the conformations of the catalytic chains. For each motif, a set of 10^{13} – 10^{18} unique active-site geometries was generated. The hashing algorithm within RosettaMatch^[165] was used to search for placements of these into the binding pockets of 71 protein scaffolds. Around 180 000 distinct solutions (matches) were found. Rosetta-Design was subsequently utilized to optimize the active-site sequence for optimal packing around the composite transition state and the catalytic lysine. A total of 72 designs in 10 different scaffolds were selected for experimental characterization. The final selection criteria were based on a) the predicted binding energy of the transition state, b) the extent to which the catalytic geometry was satisfied, c) the packing around the active lysine, and d) the consistency of side-chain conformation after side-chain repacking in the presence and absence of the composite transition state.

70 of the 72 proteins were soluble when expressed and purified from *Escherichia coli*, and a respectable 32 showed detectable retro-aldolase activity. Product formation was monitored with a fluorescence-based assay. The active designs span five different protein scaffolds (1mw4, 1f5j, 1thf, 1i4n, 1a53) from the triose phosphate isomerase (TIM) barrel and jelly-roll folds, and are based on the active-site theozyme

motifs III and IV. The designs in the relatively open jelly-roll scaffold show simple linear kinetics, whereas the TIM barrel designs with more enclosed active-site pockets displayed more complex kinetics—a potential indication of restricted substrate access and/or product release. Two apo structures (the S210A variant of RA22 and the M48K variant of RA61) were solved at 2.2 and 1.9 Å resolutions, respectively.^[170] The backbone geometries and side-chain orientations are in excellent agreement with those of the designs. Respectable rate enhancements of up to four orders of magnitude were achieved. However, even the best computational design falls two to three orders of magnitude short of the rate enhancement (k_{cat}/k_{uncat}) that is achieved by comparable catalytic antibodies.^[60,61] The catalytic efficiencies (k_{cat}/K_M) of the designs range between 0.02 and $0.74\text{ M}^{-1}\text{ s}^{-1}$ and are modest, particularly when compared to those of natural enzymes.

In an effort to shed light on the performance discrepancy of computationally designed enzymes relative to catalytic antibodies, Ruscio et al.^[179] studied the influence of structural fluctuations of the protein on the active-site preorganization of RA22 by using molecular dynamics. The authors found that an alternative orientation of the substrate with respect to His233 is optimal for the nucleophilic attack by Lys159. They further note that the His233–Asp53 dyad is disrupted due to the solvation of Asp53, which in turn provides conformational flexibility to His233, thus affecting its interaction with the substrate. The authors attributed the comparatively low activity of RA22 to these dynamic distortions in the deprotonation step of the reaction.

Lasilla et al. recently showed that the designed interactions of water with Tyr78 and Ser87 in RA61 do not contribute to catalysis.^[180] Activity is instead largely attributed to the nucleophilic character of the catalytic lysine ($pK_a = 6.8$ – 7.5) and to the favorable interaction energy between the enzyme and the naphthyl group of the substrate.

4.2. Kemp Eliminases

4.2.1. Computational Designs

The Kemp elimination (Figure 5a) is a well-studied ring-opening reaction that is initiated by deprotonation of the substrate. The reaction serves as a model for the biochemically relevant abstraction of a proton from carbon centers, although it does not have a natural counterpart. The reaction has become an attractive target for catalyst design, ranging from catalytic antibodies^[51] to “synzymes”.^[181] Most recently, DeGrado and co-workers employed a minimalist design approach to endow calmodulin with Kemp elimination activity.^[182]

The rate of the Kemp elimination depends strongly on the medium when a carboxylate functions as the general base, and rate accelerations of 10^7 can be achieved by simply placing acetate in a polar aprotic solvent such as acetonitrile compared to placing it in water.^[183] An additional acceleration of 10^6 can be achieved through precise positioning of the donor and acceptor for this reaction,^[53,183,184] thereby giving a theoretical limit for the rate enhancement of 10^{13} for the Kemp elimination.

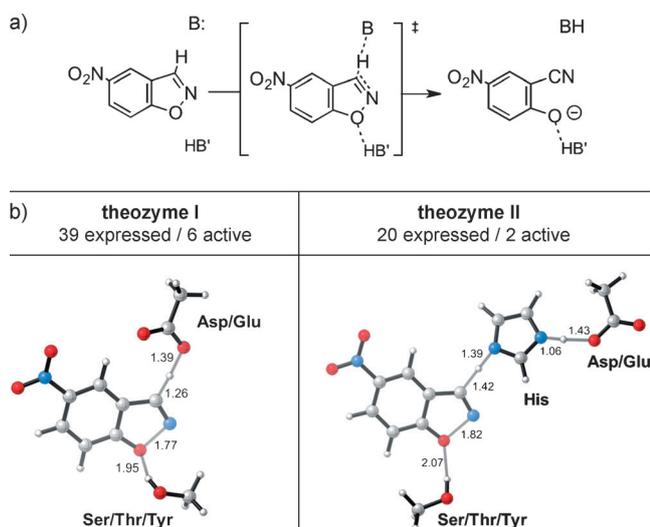


Figure 5. The Kemp elimination. a) Reaction scheme of 5-nitrobenz-isoxazole. b) The two theozymes that were employed.

In the first example of active-site design towards catalysis of an unnatural reaction, R othlisberger et al. used the inside-out protocol to produce eight active enzymes that promote the base-catalyzed ring opening of 5-nitrobenz-isoxazole (5-NBZ).^[171] Two distinct theozymes (Figure 5b) were employed in the process, and gave rise to catalysts with rate enhancements of up to 10^5 . A crystal structure was solved for KE07, an active design with $k_{\text{cat}}/K_M = 12.2 \text{ M}^{-1} \text{ s}^{-1}$, and superimposed well on the computational model.

The kinetic parameters of these eight computational Kemp elimination designs are comparable to those of catalytic antibodies. The rate enhancements ($k_{\text{cat}}/k_{\text{uncat}}$) range from 10^3 to 10^5 and compare well with the $k_{\text{cat}}/k_{\text{uncat}}$ value of 10^4 displayed by catalytic antibodies 34E4 and 35F10.^[51] In terms of substrate binding, on the other hand, the two catalytic antibodies outperform the eight computational designs up to 10-fold (antibody K_M of 0.6 to 0.1 mM compared to a range of 4.2 to 0.6 mM for the designs). Three of the computational designs were further enhanced by in vitro directed evolution. The k_{cat}/K_M value of KE07 was improved 200-fold,^[185] that of KE70 over 400-fold,^[186] and in the case of KE59 the k_{cat}/K_M value was increased over 2000-fold.^[187] The studies demonstrate how computational protein design can be used to generate enzymes with modest activities that can then be further optimized through directed evolution approaches.

4.2.2. Computational Analyses of De Novo Kemp Designs

4.2.2.1. PDDG/PM3 Monte Carlo Study

Alexandrova et al. described the analysis of the four active Kemp elimination designs KE07 (258 residues), KE10 (253), KE15 (258), and KE16 (258) by mapping out the reaction coordinate with a semiempirical PDDG/PM3 QM/MM Monte Carlo approach.^[188] The computational setup consisted of 200 residue cutaways of the four designs. The semiempirical QM part consisted of the 5-NBZ substrate and the catalytic base (Glu/Asp). Water molecules were not

included in the PM3 region. Side-chain motions were sampled while the protein backbones were held fixed. The attempt to gain insight into what governs the observed activities and to establish a correlation between the computed and experimental barriers was met with limited success. The computed barriers were plagued by large absolute error bars and by a trend that was opposite to what was found experimentally. It should be noted, however, that within the series of four designs that was chosen for this study, the free energies of activation (ΔG^\ddagger) span a range of merely $0.9 \text{ kcal mol}^{-1}$ —too narrow to be picked up by most modern QM methods.

4.2.2.2. DFT-Based Approaches

Density functional theory (DFT) calculations were employed to study six active and four inactive Kemp elimination designs with free energies of activation (ΔG^\ddagger) ranging from 18.3 to $20.6 \text{ kcal mol}^{-1}$ (actives) and to $\Delta G^\ddagger \geq 23.2 \text{ kcal mol}^{-1}$ (inactives).^[169] Three modeling approaches were explored, ranging from a minimalistic representation of the catalytic units (Figure 6, upper right), to QM on the full

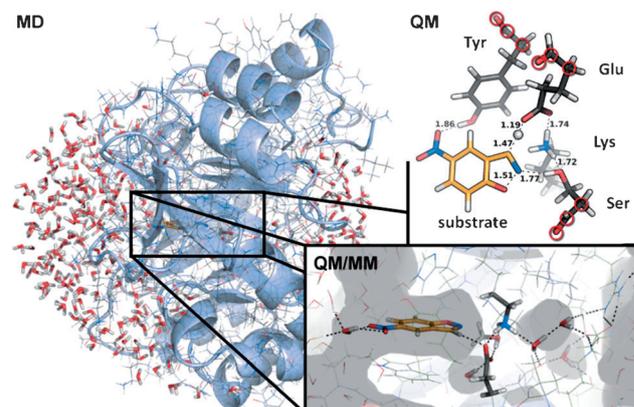


Figure 6. Modeling approaches for analysis of Kemp designs range from QM on the catalytic unit (top right; with circled backbone heavy atoms frozen) to full enzyme ONIOM QM/MM after 2 ns MD (QM layer shown as sticks at bottom right). Modified from Ref. [169].

active sites, and to computations on the entire protein systems after a short MD simulation (Figure 6, left) in which the active site and selected water molecules were treated with QM (Figure 6, bottom right).

Qualitatively, the full-protein MD-QM/MM approach compared best to experiment; the computed barriers for inactive designs were significantly higher than those of active designs. Aside from the qualitative agreement, however, the approach shows only a weak correlation with the experimentally determined energy barriers ($R^2 = 0.58$), thus indicating that significant contributions to catalysis also escape this computational model. A lesson from these studies is that computing energy barriers for base-catalyzed reactions such as the Kemp elimination, necessitates an explicit treatment of solvent molecules and other polar groups as part of the QM calculations

4.2.2.3. Empirical Valence Bond (EVB) Calculations

Warshel and co-workers used a two-layer EVB approach to evaluate Kemp designs. They applied free energy perturbation umbrella sampling (FEP/US) calculations on designs KE07 (and directed evolution variants),^[189] KE70, KE59, and HG-2.^[190] The EVB setup was calibrated to reproduce ab initio calculations of the reaction surface in a solvent cage and then applied to obtain free energies of activation (ΔG^\ddagger). For many systems, these are in exceptional agreement with experimentally determined values, yet for others (e.g. KE59, HG-2) the deviations are significant (Table 1).

Table 1: EVB activation free energies for computational Kemp eliminases.^[189,190]

System	PDB entry	Base	$\Delta G^\ddagger_{\text{exp}}$ [a]	$\Delta G^\ddagger_{\text{EVB}}$ [a]
HG-2 (S265T)	NA	Asp127	17.7	18.2
34E4 antibody	1vol	Glu ^{H50}	17.9	17.3
KE59	NA	Glu231	18.3 ^[b]	31.7
HG-2	3nyd	Asp127	18.5 ^[c]	24.3
KE70	3npu	His16-Asp44	18.5	19.3
1A53-2	3nyz	Glu178	20.0	20.7
KE07	2rxk	Glu231	20.1	19.5

[a] In kcal mol⁻¹. [b] Computed with $k_{\text{cat}} = 0.29 \text{ s}^{-1}$.^[171] [c] Computed from an extrapolated $k_{\text{cat}} = 0.22 \text{ s}^{-1}$.^[191] NA = not available.

4.2.2.4. Active-Site Dynamics from MD Simulations

Molecular dynamics simulations of 23 Kemp eliminases (14 active, 9 inactive), although of a rather qualitative nature, were more conclusive than previous computational studies. Analysis of the simulation data showed that the failed computational designs are unable to maintain essential active-site hydrogen bonds.^[169] This becomes particularly clear in the example of the inactive KE38 (Figure 7c). Compared to the catalytic His–Asn contact in the naturally evolved cathepsin K (Figure 7a) and the catalytic His–Asp dyad in the active KE70 (Figure 7b), there is no significant population in which the KE38 His–Glu dyad is intact, and His alone is too weak a base to deprotonate the substrate on its own.

Overall, a disassembly of the designed catalytic contacts was observed to occur through a combination of two factors: excessive solvent accessibility and alternative side-chain packing, both of which give rise to distinct distribution patterns. This observation is relevant to rational enzyme design in general, but particularly for the catalysis of reactions that depend on a carboxylate base, as they are usually sensitive to polar protic solvents such as water. Solvent molecules that come into direct contact with the carboxylate oxygen atoms can significantly reduce their base strength (up to 10^6 in terms of k_{cat}) and Figure 8a shows this trend for a cross-section of the dataset. On average, the active sites of functional designs are less hydrated than those of inactive designs (Figure 8b), but even the microenvironments of the most active designs are still far from those of naturally evolved acid/base catalysts such as cathepsin K (outermost right column in Figure 8b).

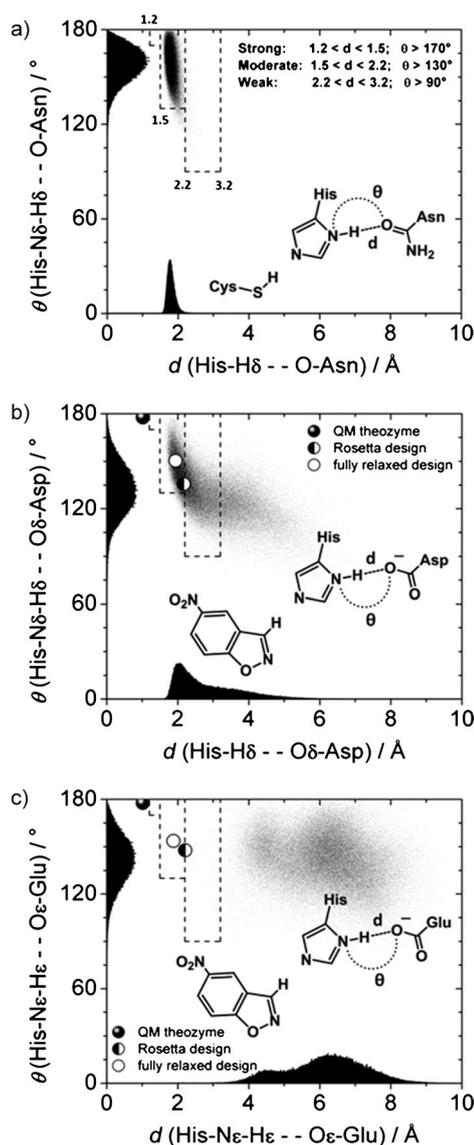


Figure 7. Angle (θ) versus distance (d) scatter plots of the catalytic contact. a) His–Asn contact of the naturally evolved cathepsin K catalytic triad; b) His–Asp contact of the active design KE70; c) His–Glu contact of the inactive design KE38. Data points are from 20 ns MD simulations. Three hydrogen bond categories^[192] are outlined with dotted lines. The individual distributions are projected onto the axes. The progression of the catalytic contact from QM theozyme, to final design, and the fully relaxed MD starting geometry is plotted with filled, half-filled, and empty circles, respectively. Modified from Ref. [169].

Taken together, active designs can be discerned from inactive ones when a multidimensional problem can be simplified to a two-dimensional model (Figure 9).

What transpires then from this study is that by querying the dynamics of a protein–substrate complex in the presence of explicitly represented solvent molecules, and by asking specific questions based on chemical intuition, one can gather a wealth of information about the system at hand and relate that to experimental observables. On this basis, it has become a useful approach to combine MD-based analyses with the

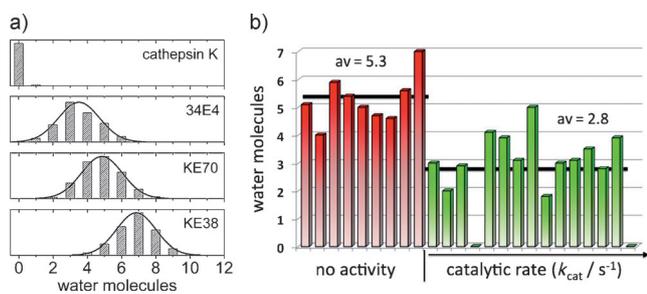


Figure 8. Water coordination distributions from MD simulations ($d < 3.2 \text{ \AA}$ to the catalytic carboxylate oxygen atoms). a) Asn182 in the naturally evolved cathepsin K ($k_{\text{cat}} = 4.2 \text{ s}^{-1}$), GluH50 in the catalytic antibody 34E4 ($k_{\text{cat}} = 0.7 \text{ s}^{-1}$), Asp44 in the active KE70 ($k_{\text{cat}} = 0.2 \text{ s}^{-1}$), and Glu170 in the inactive KE38. Reprinted from Ref. [169]. b) Maxima of the water coordination distributions of 23 distinct Kemp eliminases, including antibody 34E4 and cathepsin K.

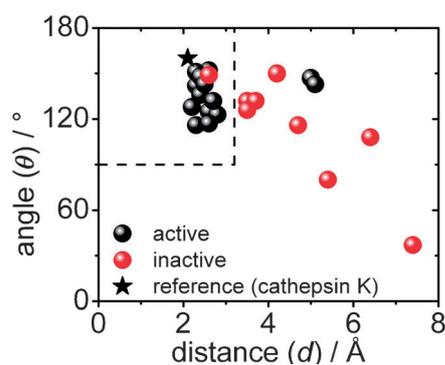


Figure 9. Angle (θ) versus distance (d) scatter plot of the catalytic hydrogen-bonding contacts of 23 distinct Kemp eliminases and cathepsin K. The data points correspond to the maxima of the angle–distance distributions. The two false-negatives correspond to two of the three least active Kemp eliminases. The dotted line marks the perimeter of weak hydrogen bonds.^[192] Atom pairs that are at larger distances or smaller angles are not considered to be hydrogen bonded. Reprinted from Ref.[169].

design and refinement of new enzymes and the interpretation of results from directed evolution experiments.

4.2.3. Directed Evolution of Kemp Eliminases KE07, KE70, and KE59

Tawfik and co-workers combined directed evolution methods with rational design and were able to further improve the catalytic activities of three computationally designed Kemp eliminases.

4.2.3.1. KE07

Seven rounds of random mutagenesis and selection resulted in up to eight mutations and a 200-fold increase in the $k_{\text{cat}}/K_{\text{M}}$ value compared to the computationally designed “wild-type”-KE07.^[185] The improvement resulted from a 2.6-fold lower K_{M} and a 76-fold higher k_{cat} value, which can largely be attributed to the Ile7Asp mutation adjacent to the

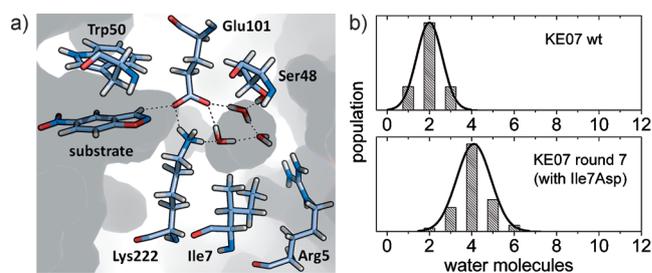


Figure 10. a) The active site of the computer-based “wild-type” KE07 design. Water molecules enter from the right and coordinate to the Glu–Lys salt bridge. b) The Ile7Asp mutation increases the average coordination number of water molecules around Glu101 from 2 in the “wild-type” KE07 to 4 (here in the example of the 7th round variant). Reprinted from Ref. [169].

active site. The Ile7Asp mutation weakens the partial salt bridge between the catalytic Glu101 and Lys222 (Figure 10a) in a dual fashion: Asp7 directly competes with Glu101 for Lys222 and also recruits additional water molecules that can directly interact with Glu101 (Figure 10b), effectively breaking up the salt bridge and tuning the pK_{a} value of the catalytic base.^[169,185]

4.2.3.2. KE70

A combination of computational optimization and nine rounds of random mutagenesis resulted in a >400-fold increase in the $k_{\text{cat}}/K_{\text{M}}$ value (up to 12-fold lower K_{M} and 53-fold higher k_{cat}).^[186] The improvement was attributed to tighter substrate binding, fine-tuned electrostatics (Figure 11a,b), and stabilization of the catalytic dyad in an orientation optimal for catalysis (Figure 11c,d). Progressive rounds of directed evolution cause the “D loop” to become less mobile (Figure 11c) and allow the catalytic dyad residues (His17 and Asp45) to form a stronger hydrogen bond (Figure 11d).

The active site of KE70 is based on theozyme II in Figure 5b. The interaction potential near the ideal distance (r_0) of the His17–Asp45 contact can be approximated to a harmonic function. Thus, the energetic penalty for deviations from r_0 is approximately proportional to $(\Delta r)^2$. Furthermore, by assuming a simple transition-state model and using the Eyring equation, $\ln(k_{\text{cat}})$ is proportional to the activation free energy. The linear relationship between $(\Delta r)^2$ and $\ln(k_{\text{cat}})$ (Figure 11d) then suggests that the increase in the k_{cat} value of the evolved variants results in a large part from the tightening of the hydrogen bond between His17 and Asp45, as the active site residues of the more evolved KE70 variants become more optimally placed and less mobile.

In contrast to KE07, beneficial mutations were not exclusive to the second and third shell, but also included first-shell residues.

4.2.3.3. KE59

The functional mutations that gave rise to a comparatively high initial activity of this computational Kemp design also caused it to be one of the least stable. In contrast to KE07 and

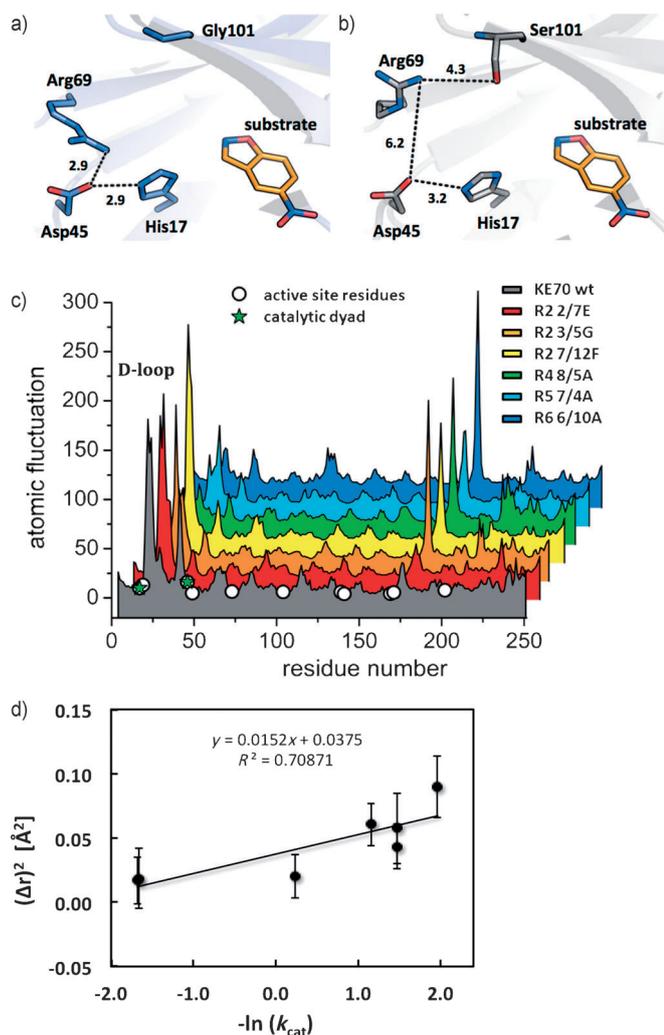


Figure 11. a) Crystal structure of the “wild-type” KE70 design. b) Crystal structure of the round 6 variant R6 6/10A. Gly101/Ser stabilizes Arg69 in an alternative conformation that does not interfere with the Asp45–His17 catalytic dyad. c) Atomic fluctuation profiles from MD simulations. The active site residues (circles) and the catalytic dyad (stars) are labeled. Peaks correspond to loops with elevated flexibility. d) The square of the deviation (Δr)² from the ideal hydrogen-bond distance (1.8 Å for this contact) versus $-\ln(k_{\text{cat}})$. Modified from Ref. [186].

KE70, a number of fold-stabilizing consensus mutations had to be introduced prior to the directed evolution. KE59 was then subjected to 16 rounds of directed evolution, which resulted in a >2000 -fold increase in the $k_{\text{cat}}/K_{\text{M}}$ value, mostly through a significantly increased k_{cat} value.^[187] The most proficient variant displayed a K_{M} value of 37 μM , a $k_{\text{cat}}/K_{\text{M}}$ value of $0.6 \times 10^6 \text{ s}^{-1}\text{M}^{-1}$, and a $k_{\text{cat}}/k_{\text{uncat}}$ value of approximately 10^7 —kinetic parameters that approach those of natural enzymes.

KE59 is also the only design in the series that accepts a variety of benzisoxazoles besides 5-NBZ. In fact, the largest optimization was achieved for 5,7-dichlorobenzisoxazole, a significantly less activated substrate. The large improvement in the k_{cat} value was attributed mainly to the ability of more advanced variants to more effectively exclude bulk solvent from interacting with the catalytic base (Figure 12a). Sub-

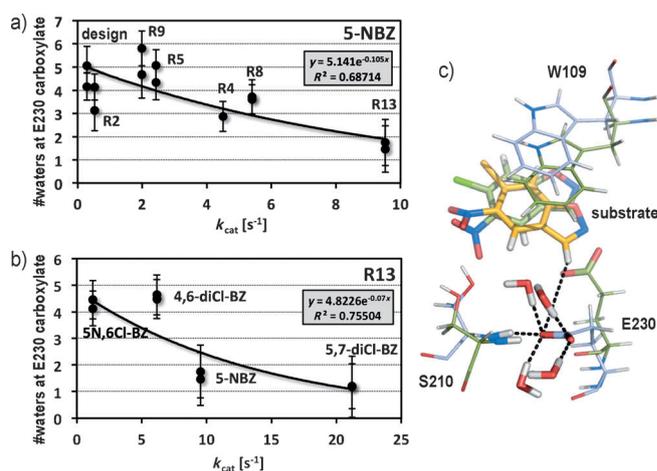


Figure 12. Number of water molecules within 3.2 Å of either of the Glu230 carboxylate oxygen atoms (from MD simulations), plotted a) over all evolved KE59 variants with available k_{cat} values for 5-nitrobenzisoxazole, and b) over all substrates with available k_{cat} values for variant R13. Error bars correspond to \pm the standard deviation of the MD-based distributions. c) The predominant conformation of Glu230 as observed in MD simulations with 5-nitro-6-chlorobenzisoxazole (blue, substrate in orange) versus the conformation observed in the crystal structures (green), shown here for the R13 3/11H variant. Modified from Ref. [187].

stituents at the 5- and 7-positions of the substrate were found to be well-suited for this purpose (Figure 12b). Conversely, Glu230 can adopt an alternative and catalytically suboptimal conformation in which it interacts with the backbone-NH group of Ser210 and an average of four water molecules (Figure 12c).

4.3. Diels–Alderase

Siegel et al. describe the inside-out computational design and experimental characterization of enzymes catalyzing a bimolecular Diels–Alder reaction (Figure 13a) with high stereoselectivity and substrate specificity (Figure 13b).^[172] No naturally occurring enzymes are known that can catalyze this cycloaddition. The catalytic motif was inspired by previous catalytic antibody studies, where an Asp, Asn, and Tyr formed the catalytic arrangement.^[44,46] The catalytic motif here consists of a Gln and two Tyr residues positioned such as to bind the bimolecular transition state leading to the 3*R*,4*S*-endo product. Two active proteins were produced: DA20 was designed into a β -propeller scaffold and DA42 into the KSI scaffold (Figure 14).

Computational evaluation methods (QM and MD) helped rationalize experimental observations and guided adjustments to early designs that resulted in improved kinetic parameters. A notable example is the development of DA_{20_10} from DA_{20_00}. Molecular dynamics simulations of DA_{20_00} ($k_{\text{cat}} = 0.1 \text{ h}^{-1}$) show that the catalytic Tyr121 can access a noncatalytic conformation in which it binds to the backbone carbonyl group of residue 271 (Figure 15, red). Increasing the steric bulk at position 272 was proposed to interfere with this interaction (Figure 15a, blue), allowing

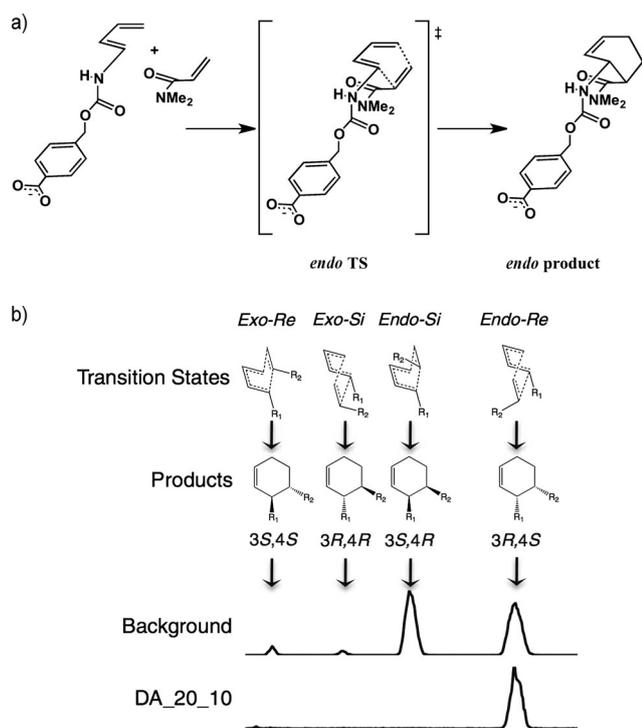


Figure 13. Diels–Alder reaction between 4-carboxybenzyl-*trans*-1,3-butadiene-1-carbamate and *N,N*-dimethylacrylamide (a), which gives only the 3*R*,4*S* *endo* product (b). Part (b) is reprinted from Ref. [172].

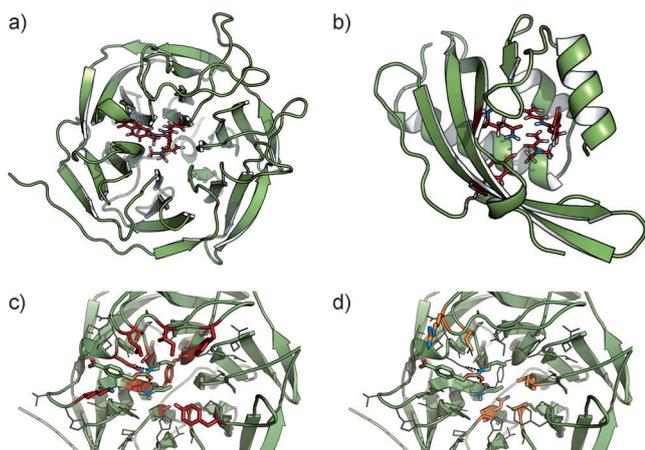


Figure 14. Computationally designed Diels–Alderase. a) DA_20_10, b) DA_42_04. Active-site Gln, Tyr, and substrates are in red sticks. c) Active site of DA_20_00 and d) of DA_20_10. Mutations are highlighted in red (for DA_20_00 compared to the native protein scaffold) and in orange (for DA_20_10 compared to DA_20_00). Parts (b) and (c) are reprinted from Ref. [172].

Tyr121 to assume the conformation required for binding and catalysis. Figure 15b shows an overlay of the active sites of DA_20_00, DA_20_10, and the QM theozyme to showcase these observations pictorially. DA_20_10 was characterized with a k_{cat} value of 2.1 h^{-1} .

The crystal structure that was solved for a variant of the DA_20 design superimposes well onto the computational design. The catalytic efficiency is comparable to that pre-

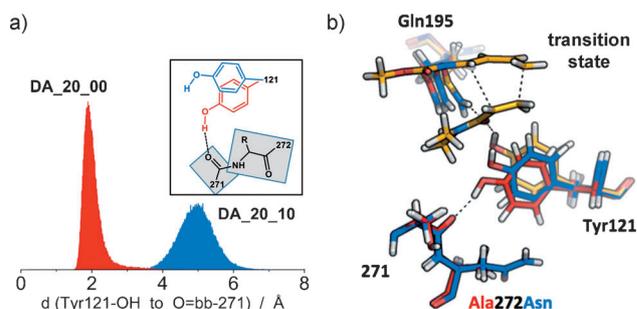


Figure 15. a) DA_20_00 shows a narrow distribution at 2 Å (hydrogen bond between Tyr121 and the carbonyl group at position 271), while DA_20_10 shows a wide distribution at 5 Å (no hydrogen bond). b) Overlay of the QM transition-state geometry (orange) with equilibrated geometries from MD simulations on DA_20_00 (red) and DA_20_10 (blue). Reprinted from Ref. [172].

viously achieved by catalytic antibodies with equal or higher catalytic rates, but a relatively weak binding of the dienophile.

4.4. Iterative Approach to “Inside-Out” Design of Enzymes

The Houk and Mayo research groups together explored an iterative variation^[191] to the Baker/Houk inside-out approach.^[171] Rather than expressing and characterizing a large number of computationally designed proteins, the efforts were focused on a single template protein (PDB-ID 1gor;^[193] Figure 16 a). As with the Röhlsberger designs, the native active site of the template was constructed to complement the transition-state geometry for the Kemp elimination of 5-NBZ (Figure 5 a). Theozyme I (Figure 5 b) served as the catalytic motif, and Phoenix^[194] rather than Rosetta was used for the design of the active-site in silico. The overall protocol was similar to that of the Röhlsberger study and was validated in its utility to generate active Kemp eliminases. However, no activity could be produced with the 1gor scaffold. HG-1 (Figure 16 b), the resulting inactive first-generation design, differs from the wild-type 1gor by seven mutations and is fully folded under the conditions of the activity assays, with a secondary structure that is very similar to the wild-type scaffold 1gor. Analysis of the structure and dynamics of HG-1, however, highlighted a number of problems. The innately flexible active-site pocket of 1gor could not be engineered to provide the necessary support for the theozyme geometry in HG-1 (compare Figure 18 a,b). Additionally it was found that a substantial number of water molecules can access the active site of HG-1—an observation that has implications for both the binding of the hydrophobic substrate (Figure 18 c) and the base strength of Glu237, the intended catalytic residue. Efforts to increase the hydrophobic character of the HG-1 active site were unsuccessful, and so a more invasive strategy was explored: rather than searching the RCSB for a protein with a native active site that is better suited for theozyme I (Figure 5 b), the focus of the computational design was shifted away from the native active site and onto a pre-existing small pocket inside the β barrel (shaded area at center-bottom of Figure 16 a,b).

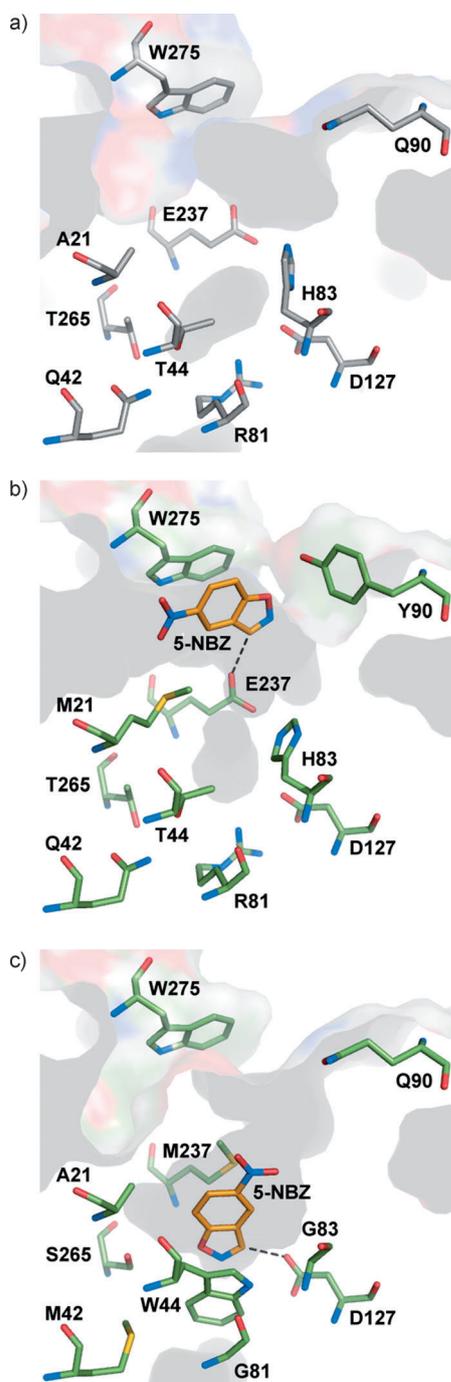


Figure 16. Variation of the active site. a) The unmodified 1gor scaffold, b) design HG-1, and c) design HG-2. The TS model is shown in orange. Reprinted from Ref. [191].

The resulting HG-2 design differs from the wild-type 1gor by 12 mutations and utilizes Asp127 as the general base (Figure 16c). The native pocket was deepened by 7 Å and tightened at the entrance, effectively generating a new active site inside the β barrel (Figure 17).

MD simulations predicted that the new design was active. They showed that HG-2 was capable of stabilizing the theozyme geometry, that it could limit the influx of water molecules to the catalytic base, and that it could support the

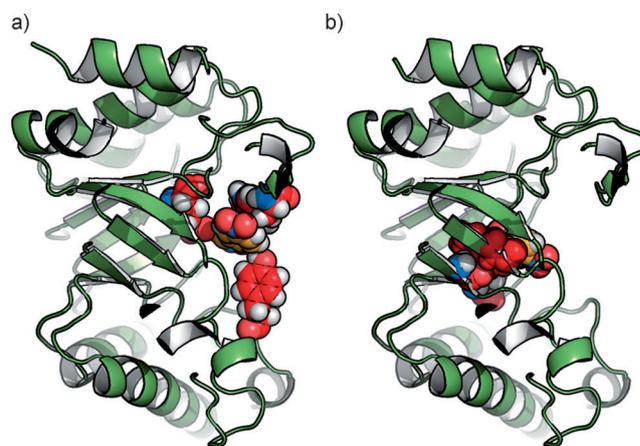


Figure 17. Active site relocated by 7 Å. a) HG-1. b) HG-2. Cutaway view with active-site residues in red and the TS model in orange.

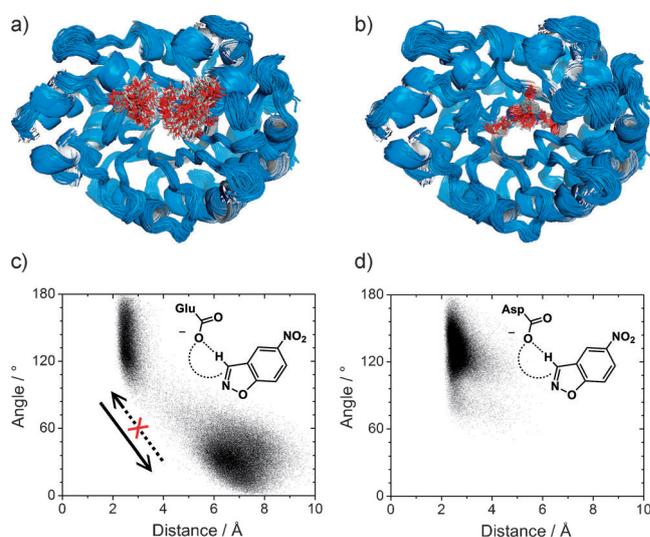


Figure 18. Dynamics of HG-1 (a) and of HG-2 (b). Equidistant snapshots from MD simulations are shown with the backbone in blue and the active-site residues in red. The backbone dynamics are of comparable magnitude in both HG-1 and HG-2, but side-chain active-site dynamics differ significantly. c) and d) show angle–distance scatter plots of the catalytic contacts between substrate and base. Modified from Ref. [191].

catalytic contact between substrate and Asp127 (Figure 18d). The protein was expressed, kinetically characterized, and its activity was confirmed with a $k_{\text{cat}}/K_{\text{M}}$ value of $123.2\text{M}^{-1}\text{s}^{-1}$, comparable to the $163\text{M}^{-1}\text{s}^{-1}$ for the most active R thlisberger design, KE59.

The crystal structure of HG-2 was solved to a resolution of 1.2 Å. A transition-state analogue (TSA) was cocrystallized and occupied the active site in two distinct but catalytically competent orientations. The structure validated the dynamics-guided computational design, but it also drew attention to the importance of accounting for alternative substrate orientations in future versions of the enzyme design protocol. Additional single-point mutations were explored with the Ser265Thr variant, and further enhanced the $k_{\text{cat}}/K_{\text{M}}$ value by a factor of three.

Table 2: Kinetic parameters of computationally designed enzymes compared to catalytic antibodies and naturally evolved enzymes.

	$k_{\text{cat}}^{[a]}$	$K_{\text{M}}^{[b]}$	$k_{\text{cat}}/K_{\text{M}}^{[c]}$	$k_{\text{cat}}/k_{\text{uncat}}$	$[k_{\text{cat}}/K_{\text{M}}]/k_{\text{uncat}}$
nat. enzymes	av 10^5	av 10^{-4}	10^6 – 10^9	10^6 – 10^{17}	10^8 – 10^{29}
cat. antibodies	10^{-2} –1	av 10^{-4}	10^2 – 10^4	10^3 – 10^6	10^5 – 10^9
Kemp elim.					
cat. antibodies	10^{-1} –1	10^{-3} – 10^{-4}	10^2 – 10^3	10^3 – 10^6	10^7 – 10^9
comp. designs	10^{-2} –1	av 10^{-3}	10 – 10^2	10^3 – 10^6	10^7 – 10^9
evolved designs	1–20	10^{-3} – 10^{-5}	10^3 – 10^6	10^6 – 10^7	10^7 – 10^{11}
Retro-Aldol					
cat. antibodies	10^{-3} – 10^{-1}	10^{-4} – 10^{-5}	10 – 10^3	10^5 – 10^6	10^7 – 10^9
comp. designs	10^{-2} – 10^{-1}	av 10^{-4}	10^{-2} – 10^{-1}	10^3 – 10^4	10^6 – 10^7
Diels-Alder^[d]					
cat. antibodies	av 10^{-5}	av 10^{-3}	av 10	av 10^3	10^9
comp. designs	10^{-5} – 10^{-4}	10^{-1} – 10^{-4}	1– 10^2	10^3 – 10^4	10^7 – 10^{11}

[a] In units of s^{-1} . [b] In units of M. [c] In units of $\text{M}^{-1} \text{s}^{-1}$. [d] $k_{\text{cat}}/(K_{\text{M-diene}} \times K_{\text{M-dienophile}})$ instead of $k_{\text{cat}}/K_{\text{M}}$ in units of $\text{s}^{-1} \text{M}^{-1} \text{M}^{-1}$ and $(k_{\text{cat}}/(K_{\text{M-diene}} \times K_{\text{M-dienophile}}))/k_{\text{uncat}}$ instead of $(k_{\text{cat}}/K_{\text{M}})/k_{\text{uncat}}$ in units of $\text{M}^{-1} \text{M}^{-1}$

most active evolved variants have rates comparable only with the least proficient of nature's enzymes.

5.1. Computational Challenges

The kinetic parameters outlined in Table 2 show that artificial catalysts developed from computational design and immunological production perform particularly poorly at binding the transition states of reactions. This can be attributed to several reasons.

One interpretation of the data is that the positioning of catalytic residues in both de novo designs and antibodies does not match the precision with which natural evolution has placed functional groups into the active sites of enzymes.^[168,199,200] Figure 3 highlights this for four computationally designed Kemp eliminases, and shows that the QM theozyme transition-state geometries are not faithfully reproduced within the designs.

The large rate enhancements produced by natural enzymes have further been attributed to the minimization of conformational motions.^[201,202] However, a comparison of computationally designed enzymes with naturally evolved ones shows that even the most active designs display a high degree of conformational mobility, as exemplified in Figure 7 a,b. Directed evolution can correct this and can increase the degree of preorganization of the active site (Figure 11 c,d), but this only works for functional designs. Those that lack catalytic function have been shown to populate nonproductive active-site configurations (Figures 7 c and 18).^[169] Efforts to improve the k_{cat} values of de novo designed enzymes thus focus on more rigorous implementations of QM theozyme geometries into scaffold proteins as well as on the routine physics-based assessment of active-site dynamics.

Nature's enzymes further depend on a balanced management of water molecules in the active site. They create finely

tuned microenvironments in which catalytic residues display significantly shifted $\text{p}K_{\text{a}}$ values and in which activated water molecules can become extensions to existing catalytic machineries. This level of control has yet to be achieved through computational or antibody catalysts. Directed evolution can help here as well, as can be seen in the example of the KE59 variants (Figure 12). An increase in the k_{cat} value was largely attributed to the improved ability of these variants to displace active-site water molecules upon substrate binding. The resulting microenvironments, however, are far from what one would call "finely tuned" and so this goal remains elusive. On the other hand, water molecules can also rescue sub-optimal enzyme active sites. Herschlag and co-workers showed this to be the case for ketosteroid isomerase (KSI), where the mutation of the catalytic Tyr16 to Phe reduced the k_{cat} value 20000-fold, while Tyr16Ser reduced the activity by only 200-fold.^[199] Crystallographic and ^{19}F NMR spectroscopic analyses show that water molecules are able to restore some of the catalytic function by acting as a substitute for the missing Tyr side chain. Conversely, three active-site water molecules facilitate catalysis in the case of catalytic antibody 13G5, and a fourth water molecule further increases the k_{cat} value 10-fold.^[203] Similarly, the improved k_{cat} values of KE07 variants are attributed to an increase in the number of water molecules in a pocket adjacent to the active site (Figure 10).^[169,185]

5.2. Experimental Challenges

Experimentally, the inside-out design of enzymes is faced with similar challenges as directed evolution. First and foremost is the setup of an effective assay with which a library of designed enzymes can be screened. It has to be both facile and sensitive, as initial designs are likely to display only weak activity, and small, single-digit increases over the uncatalyzed background rate must be detectable to guide the initial stages of a design program.

The possibility of low expression levels or insolubility of the new protein sequences poses an additional experimental challenge. Depending on the catalytic groups in the theozyme, computational de novo designs can differ from their native scaffolds by up to 15 mutations or more, each one of which generally destabilizes the protein to some extent. In line with what has been found from directed evolution experiments, the sum of mutations can then quickly cause problems for the overall integrity of the protein fold. Stable thermophilic proteins are more forgiving and thus serve as good starting points for heavy mutation loads. Alternatively, computational design can be used to increase the thermostability and half-life of enzymes by probing alternative residue identities in regions that are not associated with function.^[204] A small number of mutations (three in this case) can be sufficient to raise the melting temperature by 10°C without reducing the catalytic efficiency. Furthermore, the (re)design can be informed by related protein sequences that allow the introduction of beneficial consensus mutations.^[205]

The ever-decreasing cost of DNA synthesis, on the other hand, has greatly reduced the costs associated with testing

a new protein sequence. Whereas a typical gene would cost approximately \$50 per amino acid in 2000, custom genes can now be ordered for less than \$1 per amino acid.^[206,207] This is only a few hundred dollars for a typical enzyme, thus allowing research groups to order a number of different designs in a diverse set of scaffolds. Gene assembly approaches and custom on-chip DNA synthesis^[208] allows for even greater numbers of synthesized genes, as well as the construction of focused libraries.

The synthetic availability of substrate molecules and their toxicity profile can pose significant challenges in other cases. This is underscored by two recent studies that target the sequestration of organophosphate nerve agents. Khare et al. employed the computational inside-out protocol for the redesign of a zinc-containing deaminase with a focus on the hydrolysis of a nerve agent analogue.^[209] Hemmert et al. on the other hand, used a structure-guided approach to produce variants of carboxylesterase 1 with increased activities towards sarin, soman, and cyclosarin, the assaying for which had to be performed at a US Army Medical Research Institute for Chemical Defense.^[210]

6. Summary and Outlook

Computational methods have matured to a point at which it is now possible to produce functional proteins that can catalyze chemical reactions with no natural counterpart. So far their rate accelerations are comparatively modest. However, they display the onset of features that are typical only to nature's enzymes—such as shifted pK_a values, substrate specificity, and stereoselectivity. These can be substantially enhanced, when computational design is combined with directed evolution. The last of these has been gaining momentum in the context of industrial processes, a noteworthy example of which is the recent development of a transaminase to synthesize the diabetes drug sitagliptin.^[106] However, laboratory evolution generally requires proteins with a basal level of the desired function. When such starting points are not available in nature's repository of enzymes, computational enzyme design can be considered as a means to generate them.

So far the inside-out approach has produced catalysts for fairly simple uni- and bimolecular single-step reactions, although the construction of more involved active sites for multistep transformations is in progress.^[211] The design techniques can, however, also be used to effectively re-engineer the substrate specificities of existing enzymes, as was recently demonstrated by Gordon et al.^[212] However, with the continued advance of computational methods and computer power, algorithms are being developed that will permit incorporation of more complex and robust active sites into proteins, as well as accurate and facile physics-based simulations, with which catalyst candidates can be tested and repaired prior to the experimental workup.

The fundamental laws of physics that underlie individual components in the design of enzyme catalysts are known and—on their own—well approximated. However, when combined, they not only give rise to an exceedingly large

number of variations, as well as cross-component inconsistencies. The main difficulty in the design of a proficient enzyme then lies with rapidly computing the exact solution to each combination. One may recognize the similarity of this statement to a famous quote by Dirac: “*The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble.*”^[213]

Almost a century later, we are even more convinced that we know the physical laws that govern chemistry, including the chemistry of enzyme catalysis. Furthermore, with the help of computers, we can now solve the Dirac equation to any accuracy desired—at least for small systems. However, enzymes are not only large molecules that are immersed in water, and undergo conformational changes, but also provide finely tuned microenvironments for efficient and selective catalyses—sometimes of multiple substrates or of multiple steps, or requiring the presence of cofactors. Dirac's pessimism arose because computers were unknown at the time. However, the breakneck speed of hardware and algorithm development gives us reasons for optimism about the future of computational enzyme design. The active development of quantum mechanical methods and classical force fields ensures that the performance and precision of physical calculations continue to improve. Ideally, all of the features common to effective enzymes need to be simulated rapidly and accurately to design enzymes routinely. At present, their simultaneous application has not been achieved, but the time when this can be accomplished is within sight.

We are grateful to the Defense Advanced Research Projects Agency (DARPA), the National Institute of General Medical Sciences (NIGMS), the National Institutes of Health (NIH), the Howard Hughes Medical Institute (HHMI), and the Lawrence Scholars Program for financial support of this research.

Received: May 25, 2012

Published online: March 25, 2013

-
- [1] W. Kühne, *FEBS Lett.* **1976**, *62*, E4–E7.
 - [2] L. Pauling, *Chem. Eng. News* **1946**, *24*(10), 1375–1377.
 - [3] L. Pauling, *General Chemistry: an Introduction to Descriptive Chemistry and Modern Chemical Theory*, Freeman, San Francisco, **1947**.
 - [4] C. C. Blake, D. F. Koenig, G. A. Mair, A. C. North, D. C. Phillips, V. R. Sarma, *Nature* **1965**, *206*, 757–761.
 - [5] A. Warshel, *J. Biol. Chem.* **1998**, *273*, 27035–27038.
 - [6] A. Radzicka, R. Wolfenden, *Science* **1995**, *267*, 90–93.
 - [7] X. Zhang, K. N. Houk, *Acc. Chem. Res.* **2005**, *38*, 379–385.
 - [8] D. R. Edwards, D. C. Lohman, R. Wolfenden, *J. Am. Chem. Soc.* **2012**, *134*, 525–531.
 - [9] A. Radzicka, R. Wolfenden, *J. Am. Chem. Soc.* **1996**, *118*, 6105–6109.
 - [10] R. Wolfenden, X. Lu, G. Young, *J. Am. Chem. Soc.* **1998**, *120*, 6814–6815.
 - [11] E. A. Taylor, D. R. Palmer, J. A. Gerlt, *J. Am. Chem. Soc.* **2001**, *123*, 5824–5825.

- [12] R. A. R. Bryant, D. E. Hansen, *J. Am. Chem. Soc.* **1996**, *118*, 5498–5499.
- [13] K. Houk, A. G. Leach, S. P. Kim, X. Zhang, *Angew. Chem.* **2003**, *115*, 5020–5046; *Angew. Chem. Int. Ed.* **2003**, *42*, 4872–4897.
- [14] J. D. Stewart, *Adv. Appl. Microbiol.* **2006**, *59*, 31–52.
- [15] T. Furuya, K. Kino, *Appl. Microbiol. Biotechnol.* **2010**, *86*, 991–1002.
- [16] J. A. Gerlt, P. C. Babbitt, *Curr. Opin. Chem. Biol.* **2009**, *13*, 10–18.
- [17] *Directed Enzyme Evolution: Screening and Selection Methods* (Eds.: F. H. Arnold, G. Georgiou), Humana, New York, **2003**.
- [18] F. H. Arnold, G. Georgiou, P. C. Cirino, K. M. Mayer, D. Umeno, *Directed Evolution Library Creation*, Humana, New York, **2003**.
- [19] J. D. Bloom, M. M. Meyer, P. Meinhold, C. R. Otey, D. MacMillan, F. H. Arnold, *Curr. Opin. Struct. Biol.* **2005**, *15*, 447–452.
- [20] C. Jäckel, P. Kast, D. Hilvert, *Annu. Rev. Biophys.* **2008**, *37*, 153–173.
- [21] S. Bershtein, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **2008**, *12*, 151–158.
- [22] D. Hilvert, *Annu. Rev. Biochem.* **2000**, *69*, 751–793.
- [23] P. Wentworth, Jr., *Science* **2002**, *296*, 2247–2249.
- [24] Y. Xu, N. Yamamoto, K. D. Janda, *Bioorg. Med. Chem.* **2004**, *12*, 5247–5268.
- [25] S. C. Pan, B. List, *Ernst Schering Found Symp. Proc.* **2007**, 1–43.
- [26] T. Drepper, T. Eggert, W. Hummel, C. Leggewie, M. Pohl, F. Rosenau, S. Wilhelm, K.-E. Jaeger, *Biotechnol. J.* **2006**, *1*, 777–786.
- [27] V. Nanda, R. L. Koder, *Nat. Chem.* **2010**, *2*, 15–24.
- [28] A. Tramontano, K. D. Janda, R. A. Lerner, *Science* **1986**, *234*, 1566–1570.
- [29] S. J. Pollack, J. W. Jacobs, P. G. Schultz, *Science* **1986**, *234*, 1570–1573.
- [30] P. Wentworth, Jr., K. D. Janda, *Curr. Opin. Chem. Biol.* **1998**, *2*, 138–144.
- [31] L. Pauling, *Nature* **1948**, *161*, 707–709.
- [32] W. P. Jencks, *Catalysis in Chemistry and Enzymology*, McGraw-Hill, New York, **1969**.
- [33] M. Arnó, L. R. Domingo, *Int. J. Quantum Chem.* **2001**, *83*, 338–347.
- [34] B. G. Miller, R. Wolfenden, *Annu. Rev. Biochem.* **2002**, *71*, 847–885.
- [35] J. D. Stewart, S. J. Benkovic, *Chem. Soc. Rev.* **1993**, *22*, 213.
- [36] P. A. Patten, N. S. Gray, P. L. Yang, C. B. Marks, G. J. Wedemayer, J. J. Boniface, R. C. Stevens, P. G. Schultz, *Science* **1996**, *271*, 1086–1091.
- [37] A. Tramontano, B. Ivanov, G. Gololobov, S. Paul, *Appl. Biochem. Biotechnol.* **2000**, *83*, 233–242; discussion A. Tramontano, B. Ivanov, G. Gololobov, S. Paul, *Appl. Biochem. Biotechnol.* **2000**, *83*, 242–243.
- [38] X. Wu, T. Zhou, J. Zhu, B. Zhang, I. Georgiev, C. Wang, X. Chen, N. S. Longo, M. Louder, K. McKee, S. O'Dell, S. Perfitto, S. D. Schmidt, W. Shi, L. Wu, Y. Yang, *Science* **2011**, *333*, 1593–1602.
- [39] J. S. McLellan, M. Pancera, C. Carrico, J. Gorman, J.-P. Julien, R. Khayat, R. Louder, R. Pejchal, M. Sastry, K. Dai, et al., *Nature* **2011**, *480*, 336–343.
- [40] T. Wojcik, K. Kiec-Kononowicz, *Curr. Med. Chem.* **2008**, *15*, 1606–1615.
- [41] D. Hilvert, K. W. Hill, K. D. Nared, M. T. M. Auditor, *J. Am. Chem. Soc.* **1989**, *111*, 9261–9262.
- [42] J. Xu, Q. Deng, J. Chen, K. N. Houk, J. Bartek, D. Hilvert, I. A. Wilson, *Science* **1999**, *286*, 2345–2348.
- [43] J. Chen, Q. Deng, R. Wang, K. Houk, D. Hilvert, *ChemBioChem* **2000**, *1*, 255–261.
- [44] V. E. Gouverneur, K. N. Houk, B. de Pascual-Teresa, B. Beno, K. D. Janda, R. A. Lerner, *Science* **1993**, *262*, 204–208.
- [45] A. Heine, E. A. Stura, J. T. Yli-Kauhala, C. Gao, Q. Deng, B. R. Beno, K. N. Houk, K. D. Janda, I. A. Wilson, *Science* **1998**, *279*, 1934–1940.
- [46] C. E. Cannizzaro, J. A. Ashley, K. D. Janda, K. N. Houk, *J. Am. Chem. Soc.* **2003**, *125*, 2489–2506.
- [47] N. Bense, N. Bahr, M. Reymond, C. Schenkels, J.-L. Reymond, *Helv. Chim. Acta* **1999**, *82*, 44–52.
- [48] M. Hugot, N. Bense, M. Vogel, M. T. Reymond, B. Stadler, J.-L. Reymond, U. Baumann, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9674–9678.
- [49] A. G. Leach, K. N. Houk, J.-L. Reymond, *J. Org. Chem.* **2004**, *69*, 3683–3692.
- [50] S. P. Kim, A. G. Leach, K. N. Houk, *J. Org. Chem.* **2002**, *67*, 4250–4260.
- [51] S. N. Thorn, R. G. Daniels, M. T. Auditor, D. Hilvert, *Nature* **1995**, *373*, 228–230.
- [52] J. Na, K. N. Houk, D. Hilvert, *J. Am. Chem. Soc.* **1996**, *118*, 6462–6471.
- [53] Y. Hu, K. N. Houk, K. Kikuchi, K. Hotta, D. Hilvert, *J. Am. Chem. Soc.* **2004**, *126*, 8197–8205.
- [54] M. L. Casey, D. S. Kemp, K. G. Paul, D. D. Cox, *J. Org. Chem.* **1973**, *38*, 2294–2301.
- [55] D. S. Kemp, M. L. Casey, *J. Am. Chem. Soc.* **1973**, *95*, 6670–6680.
- [56] Y. Li, M. N. Paddon-Row, K. N. Houk, *J. Am. Chem. Soc.* **1988**, *110*, 3684–3686.
- [57] J. Wagner, R. A. Lerner, C. F. Barbas, *Science* **1995**, *270*, 1797–1800.
- [58] C. F. Barbas III, A. Heine, G. Zhong, T. Hoffmann, S. Gramatikova, R. Björnstedt, B. List, J. Anderson, E. A. Stura, I. A. Wilson, R. A. Lerner, *Science* **1997**, *278*, 2085–2092.
- [59] T. Hoffmann, G. Zhong, B. List, D. Shabat, J. Anderson, S. Gramatikova, R. A. Lerner, C. F. Barbas, *J. Am. Chem. Soc.* **1998**, *120*, 2768–2779.
- [60] B. List, C. F. Barbas, R. A. Lerner, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 15351–15355.
- [61] G. Zhong, R. A. Lerner, C. F. Barbas III, *Angew. Chem.* **1999**, *111*, 3957–3960; *Angew. Chem. Int. Ed.* **1999**, *38*, 3738–3741.
- [62] M. Arnó, L. R. Domingo, *Org. Biomol. Chem.* **2003**, *1*, 637–643.
- [63] D. S. Kemp, K. G. Paul, *J. Am. Chem. Soc.* **1975**, *97*, 7305–7312.
- [64] D. S. Kemp, D. D. Cox, K. G. Paul, *J. Am. Chem. Soc.* **1975**, *97*, 7312–7318.
- [65] H. Zipse, G. Apaydin, K. N. Houk, *J. Am. Chem. Soc.* **1995**, *117*, 8608–8617.
- [66] J. Gao, *J. Am. Chem. Soc.* **1995**, *117*, 8600–8607.
- [67] C. Lewis, T. Kramer, S. Robinson, D. Hilvert, *Science* **1991**, *253*, 1019–1022.
- [68] K. Hotta, H. Lange, D. J. Tantillo, K. N. Houk, D. Hilvert, I. A. Wilson, *J. Mol. Biol.* **2000**, *302*, 1213–1225.
- [69] G. Ujaque, D. J. Tantillo, Y. Hu, K. Houk, K. Hotta, D. Hilvert, *J. Comput. Chem.* **2003**, *24*, 98–110.
- [70] K. D. Janda, C. G. Shevlin, R. A. Lerner, *Science* **1993**, *259*, 490–493.
- [71] K. Gruber, B. Zhou, K. N. Houk, R. A. Lerner, C. G. Shevlin, I. A. Wilson, *Biochemistry* **1999**, *38*, 7062–7074.
- [72] J. Na, K. N. Houk, C. G. Shevlin, K. D. Janda, R. A. Lerner, *J. Am. Chem. Soc.* **1993**, *115*, 8453–8454.
- [73] J. Na, K. N. Houk, *J. Am. Chem. Soc.* **1996**, *118*, 9204–9205.
- [74] J. M. Coxon, A. J. Thorpe, *J. Am. Chem. Soc.* **1999**, *121*, 10955–10957.

- [75] K. D. Janda, C. G. Shevlin, R. A. Lerner, *J. Am. Chem. Soc.* **1995**, *117*, 2659–2660.
- [76] K. D. Janda, D. Schloeder, S. J. Benkovic, R. A. Lerner, *Science* **1988**, *241*, 1188–1191.
- [77] J. D. Stewart, J. F. Krebs, G. Siuzdak, A. J. Berdis, D. B. Smithrud, S. J. Benkovic, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 7404–7409.
- [78] C. Gao, B. J. Lavey, C.-H. L. Lo, A. Datta, P. Wentworth, K. D. Janda, *J. Am. Chem. Soc.* **1998**, *120*, 2211–2217.
- [79] V. A. Roberts, J. Stewart, S. J. Benkovic, E. D. Getzoff, *J. Mol. Biol.* **1994**, *235*, 1098–1116.
- [80] M. M. Thayer, E. H. Olender, A. S. Arvai, C. K. Koike, I. L. Canestrelli, J. D. Stewart, S. J. Benkovic, E. D. Getzoff, V. A. Roberts, *J. Mol. Biol.* **1999**, *291*, 329–345.
- [81] L. T. Chong, P. Bandyopadhyay, T. S. Scanlan, I. D. Kuntz, P. A. Kollman, *J. Comput. Chem.* **2003**, *24*, 1371–1377.
- [82] Y. M. Chook, H. Ke, W. N. Lipscomb, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8600–8603.
- [83] A. Y. Lee, P. A. Karplus, B. Ganem, J. Clardy, *J. Am. Chem. Soc.* **1995**, *117*, 3627–3628.
- [84] M. R. Haynes, E. A. Stura, D. Hilvert, I. A. Wilson, *Science* **1994**, *263*, 646–652.
- [85] P. Schultz, *Science* **1988**, *240*, 426–433.
- [86] O. Wiest, K. N. Houk, *J. Am. Chem. Soc.* **1995**, *117*, 11628–11639.
- [87] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, *485*, 185–194.
- [88] M. T. Reetz, P. Soni, J. P. Acevedo, J. Sanchis, *Angew. Chem.* **2009**, *121*, 8418–8422; *Angew. Chem. Int. Ed.* **2009**, *48*, 8268–8272.
- [89] M. Zumárraga, T. Bulter, S. Shleev, J. Polaina, A. Martínez-Arias, F. J. Plou, A. Ballesteros, M. Alcalde, *Chem. Biol.* **2007**, *14*, 1052–1064.
- [90] J. Karanicolas, J. E. Corn, I. Chen, L. A. Joachimiak, O. Dym, S. H. Peck, S. Albeck, T. Unger, W. Hu, G. Liu, S. Delbecq, G. Montelione, C. Spiegel, D. R. Liu, D. Baker, *Mol. Cell* **2011**, *42*, 250–260.
- [91] G. J. Williams, C. Zhang, J. S. Thorson, *Nat. Chem. Biol.* **2007**, *3*, 657–662.
- [92] S. E. Deacon, M. J. McPherson, *ChemBioChem* **2011**, *12*, 593–601.
- [93] R. D. Gupta, M. Goldsmith, Y. Ashani, Y. Simo, G. Mullokan-dov, H. Bar, M. Ben-David, H. Leader, R. Margalit, I. Silman, J. L. Sussman, D. S. Tawfik, *Nat. Chem. Biol.* **2011**, *7*, 120–125.
- [94] S. Bartsch, R. Kourist, U. T. Bornscheuer, *Angew. Chem.* **2008**, *120*, 1531–1534; *Angew. Chem. Int. Ed.* **2008**, *47*, 1508–1511.
- [95] M. T. Reetz, *Angew. Chem.* **2011**, *122*, 144–182; *Angew. Chem. Int. Ed.* **2011**, *50*, 138–174.
- [96] M. T. Reetz, B. Brunner, T. Schneider, F. Schulz, C. M. Clouthier, M. M. Kayser, *Angew. Chem.* **2004**, *116*, 4167–4170; *Angew. Chem. Int. Ed.* **2004**, *43*, 4075–4078.
- [97] S. Park, K. L. Morley, G. P. Horsman, M. Holmquist, K. Hult, R. J. Kazlauskas, *Chem. Biol.* **2005**, *12*, 45–54.
- [98] R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. A. Sheldon, G. W. Huisman, *Nat. Biotechnol.* **2007**, *25*, 338–344.
- [99] P. L. Bergquist, R. A. Reeves, M. D. Gibbs, *Biomol. Eng.* **2005**, *22*, 63–72.
- [100] G. Amitai, R. D. Gupta, D. S. Tawfik, *HFSP J.* **2007**, *1*, 67–78.
- [101] J. D. Bloom, P. A. Romero, Z. Lu, F. H. Arnold, *Biol. Direct* **2007**, *2*, 17.
- [102] B. Seelig, J. W. Szostak, *Nature* **2007**, *448*, 828–831.
- [103] J.-C. Baret, O. J. Miller, V. Taly, M. Ryckelynck, A. El-Harrak, L. Frenz, C. Rick, M. L. Samuels, J. B. Hutchison, J. J. Agresti, D. R. Link, D. A. Weitz, A. D. Griffiths, *Lab Chip* **2009**, *9*, 1850–1858.
- [104] J.-P. Goddard, J.-L. Reymond, *Curr. Opin. Biotechnol.* **2004**, *15*, 314–322.
- [105] E. M. Brustad, F. H. Arnold, *Curr. Opin. Chem. Biol.* **2011**, *15*, 201–210.
- [106] C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman, G. J. Hughes, *Science* **2010**, *329*, 305–309.
- [107] T. T. Li, J. J. Liang, A. A. Ambrogelly, T. T. Brennan, G. G. Floor, G. G. Huisman, J. J. Lalonde, A. A. Lekhal, B. B. Mijts, S. S. Muley, L. Newman, M. Tobin, G. Wong, A. Zaks, X. Zhang, *J. Am. Chem. Soc.* **2012**, *134*, 6467–6472.
- [108] M. Bocola, N. Otte, K.-E. Jaeger, M. T. Reetz, W. Thiel, *ChemBioChem* **2004**, *5*, 214–223.
- [109] M. T. Reetz, M. Puls, J. D. Carballeira, A. Vogel, K.-E. Jaeger, T. Eggert, W. Thiel, M. Bocola, N. Otte, *ChemBioChem* **2007**, *8*, 106–112.
- [110] G. A. Behrens, A. Hummel, S. K. Padhi, S. Schaetzle, U. T. Bornscheuer, *Adv. Synth. Catal.* **2011**, *353*, 2191–2215.
- [111] G. A. Strohmeier, H. Pichler, O. May, M. Gruber-Khadjawi, *Chem. Rev.* **2011**, *111*, 4141–4164.
- [112] A. S. Bommarius, J. K. Blum, M. J. Abrahamson, *Curr. Opin. Chem. Biol.* **2011**, *15*, 194–200.
- [113] G.-W. Zheng, J.-H. Xu, *Curr. Opin. Biotechnol.* **2011**, *22*, 784–792.
- [114] O. Khersonsky, D. S. Tawfik, *Annu. Rev. Biochem.* **2010**, *79*, 471–505.
- [115] A. C. Joerger, S. Mayer, A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5694–5699.
- [116] G. J. Poelarends, J. J. Almrud, H. Serrano, J. E. Darty, W. H. Johnson, M. L. Hackert, C. P. Whitman, *Biochemistry* **2006**, *45*, 7700–7708.
- [117] W. S. Yew, J. Akana, E. L. Wise, I. Rayment, J. A. Gerlt, *Biochemistry* **2005**, *44*, 1807–1815.
- [118] E. L. Wise, W. S. Yew, J. Akana, J. A. Gerlt, I. Rayment, *Biochemistry* **2005**, *44*, 1816–1823.
- [119] D. M. Z. Schmidt, E. C. Mundorff, M. Dojka, E. Bermudez, J. E. Ness, S. Govindarajan, P. C. Babbitt, J. Minshull, J. A. Gerlt, *Biochemistry* **2003**, *42*, 8387–8393.
- [120] Y. Terao, K. Miyamoto, H. Ohta, *Chem. Commun.* **2006**, 3600–3602.
- [121] Y. Ijima, K. Matoishi, Y. Terao, N. Doi, H. Yanagawa, H. Ohta, *Chem. Commun.* **2005**, 877–879.
- [122] Y. Terao, K. Miyamoto, H. Ohta, *Appl. Microbiol. Biotechnol.* **2006**, *73*, 647–653.
- [123] H. Xiang, L. Luo, K. L. Taylor, D. Dunaway-Mariano, *Biochemistry* **1999**, *38*, 7638–7652.
- [124] R. B. Hamed, E. T. Batchelar, I. J. Clifton, C. J. Schofield, *Cell. Mol. Life Sci.* **2008**, *65*, 2507–2527.
- [125] R. B. Hill, D. P. Raleigh, A. Lombardi, W. F. DeGrado, *Acc. Chem. Res.* **2000**, *33*, 745–754.
- [126] B. R. Gibney, F. Rabanal, J. J. Skalicky, A. J. Wand, P. L. Dutton, *J. Am. Chem. Soc.* **1997**, *119*, 2323–2324.
- [127] B. R. Gibney, F. Rabanal, J. J. Skalicky, A. J. Wand, P. L. Dutton, *J. Am. Chem. Soc.* **1999**, *121*, 4952–4960.
- [128] K. T. Simons, C. Kooperberg, E. Huang, D. Baker, *J. Mol. Biol.* **1997**, *268*, 209–225.
- [129] B. I. Dahiya, S. L. Mayo, *Science* **1997**, *278*, 82–87.
- [130] G. G. Dantas, B. B. Kuhlman, D. D. Callender, M. M. Wong, D. D. Baker, *J. Mol. Biol.* **2003**, *332*, 449–460.
- [131] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, D. Baker, *Science* **2003**, *302*, 1364–1368.
- [132] P.-S. Huang, J. J. Love, S. L. Mayo, *Protein Sci.* **2007**, *16*, 2770–2774.
- [133] S. J. Fleishman, T. A. Whitehead, D. C. Ekiert, C. Dreyfus, J. E. Corn, E.-M. Strauch, I. A. Wilson, D. Baker, *Science* **2011**, *332*, 816–821.

- [134] H. Yin, J. S. Slusky, B. W. Berger, R. S. Walters, G. Vilaire, R. I. Litvinov, J. D. Lear, G. A. Caputo, J. S. Bennett, W. F. DeGrado, *Science* **2007**, *315*, 1817–1822.
- [135] S. J. Shandler, I. V. Korendovych, D. T. Moore, K. B. Smith-Dupont, C. N. Streu, R. I. Litvinov, P. C. Billings, F. Gai, J. S. Bennett, W. F. DeGrado, *J. Am. Chem. Soc.* **2011**, *133*, 12378–12381.
- [136] G. Grigoryan, Y. H. Kim, R. Acharya, K. Axelrod, R. M. Jain, L. Willis, M. Drndic, J. M. Kikkawa, W. F. DeGrado, *Science* **2011**, *332*, 1071–1076.
- [137] B. S. Chevalier, T. Kortemme, M. S. Chadsey, D. Baker, R. J. Monnat, B. L. Stoddard, *Mol. Cell* **2002**, *10*, 895–905.
- [138] J. Ashworth, J. J. Havranek, C. M. Duarte, D. Sussman, R. J. Monnat, B. L. Stoddard, D. Baker, *Nature* **2006**, *441*, 656–659.
- [139] U. Y. Ulge, D. A. Baker, R. J. Monnat, *Nucleic Acids Res.* **2011**, *39*, 4330–4339.
- [140] J. Ashworth, G. K. Taylor, J. J. Havranek, S. A. Quadri, B. L. Stoddard, D. Baker, *Nucleic Acids Res.* **2010**, *38*, 5601–5608.
- [141] S. B. Thyme, J. Jarjour, R. Takeuchi, J. J. Havranek, J. Ashworth, A. M. Scharenberg, B. L. Stoddard, D. Baker, *Nature* **2009**, *461*, 1300–1304.
- [142] J. S. Marvin, H. W. Hellinga, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4955–4960.
- [143] W. Yang, L. M. Jones, L. Isley, Y. Ye, H.-W. Lee, A. Wilkins, Z.-R. Liu, H. W. Hellinga, R. Malchow, M. Ghazi, J. J. Yang, *J. Am. Chem. Soc.* **2003**, *125*, 6165–6171.
- [144] L. L. Looger, M. A. Dwyer, J. J. Smith, H. W. Hellinga, *Nature* **2003**, *423*, 185–190.
- [145] M. Allert, S. S. Rizk, L. L. Looger, H. W. Hellinga, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7907–7912.
- [146] B. Schreier, C. Stumpp, S. Wiesner, B. Hocker, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18491–18496.
- [147] F. E. Boas, P. B. Harbury, *J. Mol. Biol.* **2008**, *380*, 415–424.
- [148] H. C. Fry, A. Lehmann, J. G. Saven, W. F. DeGrado, M. J. Therien, *J. Am. Chem. Soc.* **2010**, *132*, 3997–4005.
- [149] A. Morin, K. W. Kaufmann, C. Fortenberry, J. M. Harp, L. S. Mizoue, J. Meiler, *Protein Eng. Des. Sel.* **2011**, *24*, 503–516.
- [150] K. Johnsson, R. K. Allemann, H. Widmer, S. A. Benner, *Nature* **1993**, *365*, 530–532.
- [151] R. A. Lerner, S. J. Benkovic, P. G. Schultz, *Science* **1991**, *252*, 659–667.
- [152] T. Sasaki, E. T. Kaiser, *J. Am. Chem. Soc.* **1989**, *111*, 380–381.
- [153] K. S. Broo, L. Brive, P. Ahlberg, L. Baltzer, *J. Am. Chem. Soc.* **1997**, *119*, 11362–11372.
- [154] P. Rossi, P. Tecilla, L. Baltzer, P. Scrimin, *Chem. Eur. J.* **2004**, *10*, 4163–4170.
- [155] B. Lovejoy, S. Choe, D. Cascio, D. K. McRorie, W. F. DeGrado, D. Eisenberg, *Science* **1993**, *259*, 1288–1293.
- [156] C. Tommos, J. J. Skalicky, D. L. Pilloud, A. J. Wand, P. L. Dutton, *Biochemistry* **1999**, *38*, 9495–9507.
- [157] S. Chakraborty, J. Y. Kravitz, P. W. Thulstrup, L. Hemmingsen, W. F. DeGrado, V. L. Pecoraro, *Angew. Chem.* **2011**, *123*, 2097–2101; *Angew. Chem. Int. Ed.* **2011**, *50*, 2049–2053.
- [158] C. M. Summa, M. M. Rosenblatt, J.-K. Hong, J. D. Lear, W. F. DeGrado, *J. Mol. Biol.* **2002**, *321*, 923–938.
- [159] D. N. Bolon, S. L. Mayo, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14274–14279.
- [160] C. Pabo, *Nature* **1983**, *301*, 200–200.
- [161] J. W. Ponder, F. M. Richards, *J. Mol. Biol.* **1987**, *193*, 775–791.
- [162] J. Bowie, R. Luthy, D. Eisenberg, *Science* **1991**, *253*, 164–170.
- [163] K. Yue, K. A. Dill, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 4163–4167.
- [164] H. M. Berman, *Nucleic Acids Res.* **2000**, *28*, 235–242.
- [165] A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Röthlisberger, D. Baker, *Protein Sci.* **2006**, *15*, 2785–2794.
- [166] D. Tantillo, *Stimulating Concepts in Chemistry*, Wiley-VCH, Weinheim, **2000**.
- [167] F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic, D. Baker, *PLoS ONE* **2011**, *6*, e19230.
- [168] A. J. T. Smith, R. Müller, M. D. Toscano, P. Kast, H. W. Hellinga, D. Hilvert, K. N. Houk, *J. Am. Chem. Soc.* **2008**, *130*, 15361–15373.
- [169] G. Kiss, D. Röthlisberger, D. Baker, K. N. Houk, *Protein Sci.* **2010**, *19*, 1760–1773.
- [170] L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, D. Hilvert, K. N. Houk, B. L. Stoddard, D. Baker, *Science* **2008**, *319*, 1387–1391.
- [171] D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, *Nature* **2008**, *453*, 190–195.
- [172] J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St. Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, D. Baker, *Science* **2010**, *329*, 309–313.
- [173] F. W. Studier, *Protein Expression Purif.* **2005**, *41*, 207–234.
- [174] E. Sulkowski, *Trends Biotechnol.* **1985**, *3*, 1–7.
- [175] R. Janknecht, G. de Martynoff, J. Lou, R. A. Hipskind, A. Nordheim, H. G. Stunnenberg, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 8972–8976.
- [176] *Guide to Protein Purification* (Eds.: R. R. Burgess, M. P. Deutscher), Academic Press, **2009**.
- [177] T. Gefflaut, C. Blonski, J. Perie, M. Willson, *Prog. Biophys. Mol. Biol.* **1995**, *63*, 301–340.
- [178] A. Heine, *Science* **2001**, *294*, 369–374.
- [179] J. Z. Ruscio, J. E. Kohn, K. A. Ball, T. Head-Gordon, *J. Am. Chem. Soc.* **2009**, *131*, 14111–14115.
- [180] J. K. Lassila, D. Baker, D. Herschlag, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4937–4942.
- [181] F. Hollfelder, A. J. Kirby, D. S. Tawfik, *J. Am. Chem. Soc.* **1997**, *119*, 9578–9579.
- [182] I. V. Korendovych, D. W. Kulp, Y. Wu, H. Cheng, H. Roder, W. F. DeGrado, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6823–6827.
- [183] A. J. Kirby, *Acc. Chem. Res.* **1997**, *30*, 290–296.
- [184] D. S. Kemp, *Nature* **1995**, *373*, 196–197.
- [185] O. Khersonsky, D. Röthlisberger, O. Dym, S. Albeck, C. J. Jackson, D. Baker, D. S. Tawfik, *J. Mol. Biol.* **2010**, *396*, 1025–1042.
- [186] O. Khersonsky, D. Röthlisberger, A. M. Wollacott, P. Murphy, O. Dym, S. Albeck, G. Kiss, K. N. Houk, D. Baker, D. S. Tawfik, *J. Mol. Biol.* **2011**, *407*, 391–412.
- [187] O. Khersonsky, G. Kiss, D. Röthlisberger, O. Dym, S. Albeck, K. N. Houk, D. Baker, D. S. Tawfik, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 10358–10363.
- [188] A. N. Alexandrova, D. Röthlisberger, D. Baker, W. L. Jorgensen, *J. Am. Chem. Soc.* **2008**, *130*, 15907–15915.
- [189] M. P. Frushicheva, J. Cao, Z. T. Chu, A. Warshel, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16869–16874.
- [190] M. P. Frushicheva, J. Cao, A. Warshel, *Biochemistry* **2011**, *50*, 3849–3858.
- [191] H. K. Privett, G. Kiss, T. M. Lee, R. Blomberg, R. A. Chica, L. M. Thomas, D. Hilvert, K. N. Houk, S. L. Mayo, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3790–3795.
- [192] T. Steiner, *Angew. Chem.* **2002**, *114*, 50–80; *Angew. Chem. Int. Ed.* **2002**, *41*, 48–76.
- [193] L. Lo Leggio, S. Kalogiannis, K. Eckert, S. C. Teixeira, M. K. Bhat, C. Andrei, R. W. Pickersgill, S. Larsen, *FEBS Lett.* **2001**, *509*, 303–308.
- [194] B. I. Dahiya, D. B. Gordon, S. L. Mayo, *Protein Sci.* **1997**, *6*, 1333–1337.

- [195] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović, F. Players, *Nature* **2010**, *466*, 756–760.
- [196] F. Khatib, S. Cooper, M. D. Tyka, K. Xu, I. Makedon, Z. Popović, D. Baker, F. Players, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 18949–18953.
- [197] F. Khatib, F. DiMaio, Foldit Contenders Group, Foldit Void Crushers Group, S. Cooper, M. Kazmierczyk, M. Gilski, S. Krzywda, H. Záborská, I. Pichová, J. Thompson, Z. Popović, M. Jaskolski, D. Baker, *Nat. Struct. Mol. Biol.* **2011**, *18*, 1175–1177
- [198] C. B. Eiben, J. B. Siegel, J. B. Bale, S. Cooper, F. Khatib, B. W. Shen, F. Players, B. L. Stoddard, Z. Popović, D. Baker, *Nat. Biotechnol.* **2012**, *30*, 190–192.
- [199] D. A. Kraut, P. A. Sigala, T. D. Fenn, D. Herschlag, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1960–1965.
- [200] J. Villà, A. Warshel, *J. Phys. Chem. B* **2001**, *105*, 7887–7907.
- [201] W. R. Cannon, S. F. Singleton, S. J. Benkovic, *Nat. Struct. Mol. Biol.* **1996**, *3*, 821–833.
- [202] H. Liu, A. Warshel, *Biochemistry* **2007**, *46*, 6011–6025.
- [203] E. W. Debler, R. Müller, D. Hilvert, I. A. Wilson, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18539–18544.
- [204] A. Korkegian, M. E. Black, D. Baker, B. L. Stoddard, *Science* **2005**, *308*, 857–860.
- [205] M. Lehmann, L. Pasamontes, S. F. Lassen, M. Wyss, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **2000**, *1543*, 408–415.
- [206] R. Carlson, *Nat. Biotechnol.* **2009**, *27*, 1091–1094.
- [207] M. Baker, *Nature* **2011**, *473*, 403–408.
- [208] J. Tian, K. Ma, I. Saaem, *Mol. BioSyst.* **2009**, *5*, 714–722.
- [209] S. D. Khare, Y. Kipnis, P. J. Greisen, R. Takeuchi, Y. Ashani, M. Goldsmith, Y. Song, J. L. Gallaher, I. Silman, H. Leader, J. L. Sussman, *Nat. Chem. Biol.* **2012**, *8*, 294–300.
- [210] A. C. Hemmert, T. C. Otto, R. A. Chica, M. Wierdl, J. S. Edwards, S. M. Lewis, S. L. Lewis, C. C. Edwards, L. Tsurkan, C. L. Cadieux, S. A. Kasten, J. R. Cashman, S. L. Mayo, P. M. Potter, D. M. Cerasoli, M. R. Redinbo, *PLoS ONE* **2011**, *6*, e17441.
- [211] F. Richter, R. Blomberg, S. D. Khare, G. Kiss, A. P. Kuzin, A. J. T. Smith, J. Gallaher, Z. Pianowski, R. C. Helgeson, A. Grjasnow, R. Xiao, J. Seetharaman, M. Su, S. Vorobiev, S. Lew, F. Forouhar, G. J. Kornhaber, J. F. Hunt, G. T. Montelione, L. Tong, K. N. Houk, D. Hilvert, D. Baker, *J. Am. Chem. Soc.* **2012**, *134*, 16197–16206.
- [212] S. R. Gordon, E. J. Stanley, S. Wolf, A. Toland, S. J. Wu, D. Hadidi, J. H. Mills, D. Baker, I. S. Pultz, J. B. Siegel, *J. Am. Chem. Soc.* **2012**, DOI: 10.1021/ja3094795.
- [213] P. A. M. Dirac, *Proc. R. Soc. London Ser. A* **1929**, *123*, 714–733.