UC Berkeley
UC Berkeley Previously Published Works

Title
Computational Design of Synthetic Enzymes

Permalink
https://escholarship.org/uc/item/9jv6c9j9

Journal
Chemical Reviews, 119(11)

ISSN
0009-2665

Authors
Welborn, Valerie Vaissier
Head-Gordon, Teresa

Publication Date
2019-06-12

DOI
10.1021/acs.chemrev.8b00399

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Computational Design of Synthetic Enzymes

Valerie Vaissier Welborn[†,‡,¶] and Teresa Head-Gordon[*,†,‡,§, ,¶]

†*Kenneth S. Pitzer Center for Theoretical Chemistry*

‡*Department of Chemistry, University of California, Berkeley*

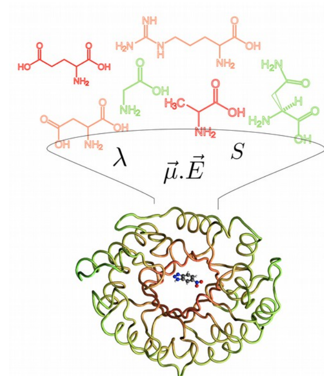
¶*Chemical Sciences Division, Lawrence Berkeley National Labs, Berkeley*

§*Department of Chemical and Biomolecular Engineering and*

Department of Bioengineering, University of California, Berkeley

*Corresponding author, E-mail: thg@berkeley.edu

We review the standard model for *de novo* computational design of enzymes, which primarily focuses on the development of an active site geometry composed of protein functional groups in orientations optimized to stabilize the transition state for a novel chemical reaction not found in nature. Its emphasis is placed on the structure and energetics of the active site embedded in an accommodating protein that serves as a physical support that shields the reaction chemistry from solvent, which is typically improved upon using laboratory directed evolution. We also provide a review of design strategies that move beyond the standard model, by placing more emphasis on the designed enzyme as a whole catalytic construct. Starting with complete *de novo* enzyme design examples, we consider additional design factors such as entropy of individual residues, correlated motion between side chains (mutual information), dynamical correlations of the enzyme motions that could aid the reaction, reorganization energy, and electric fields as a way to exploit the entire protein scaffold to improve upon the catalytic rate, thereby providing directed evolution with better starting sequences for increasing the biocatalytic performance.



1. Introduction

1.1 Enzymes in Biology

1.2 Mimicking Enzymes for Reactions Not Found in Nature

1.3 Laboratory Directed Evolution

1.4 Computational Enzyme Design

2. The Standard Model for Enzyme Design

2.1 Design Protocol for Active Site Chemistry

2.2 Recognized Limitations and Improvement Strategy

2.3 Active Site Positioning and Molecular Dynamics

3. Beyond the Standard Model for Enzyme Design

3.1 Complete de novo design

3.2 Entropic Effects on Activation Free Energy

3.3 Dynamical Effects

3.4 Minimizing the Reorganization Energy

3.5 Electric Field Optimization

4. Conclusion

5. Bibliography

6. Biographies

1. Introduction

1.1 Enzymes in biology

Enzymes accelerate biological reactions under mild conditions with efficiencies that remain unrivaled to this day.[1,2] Their remarkable performance is attributed to the active site, a pocket within the protein where reactants are shielded from bulk solvent and any undesired competing reactions.[2,3] The segregation of the reactants in the active site also induces specific molecular interactions that promote the formation of the reaction transition state, thereby lowering the activation energy compared to the uncatalyzed reaction in water. Although diverse in nature, these substrate-enzyme interactions are specific, making enzymes highly selective, i.e. exhibiting stereo-, regio- and chemoselectivity.

Quantifying and comparing enzymatic activity is typically formulated through the mechanistic Michaelis-Menten equation[4,5]:

$$E+S \underset{k_r}{\overset{k_f}{\leftrightarrows}} E\cdot S \xrightarrow{k_2} E+P \quad (1)$$

where E , S , $E\cdot S$ and P is the enzyme, substrate (or reactant), enzyme · substrate complex, and product, respectively. The kinetic constants k_f and k_r are the forward and reverse rates associated with the formation of the $E\cdot S$ complex, and once formed, k_2 is the forward rate to product and catalyst turnover. While Michaelis-Menten treated the first step of Eq. (1) as a pre-equilibrium[4], Briggs and Haldane[6] proposed the steady state approximation to derive the velocity, v_0 , i.e. the rate of change of product formation with time.

$$v_0=\frac{k_2[E]_0}{1+\frac{k_2+k_r}{k_f[S]_0}}=\frac{k_2[E]_0[S]_0}{[S]_0+K_M} \quad (2)$$

In Eq. (2), $[E]_0$ and $[S]_0$ are the initial concentrations of enzyme and substrate, and K_M is the Michaelis constant that is defined when half of the active sites of an enzyme are filled, and the rate of product formation is half of its maximum velocity, v_{max} .

The analysis of Eq. (2) can give us different measures of the quality of an enzyme's performance that will be useful to consider when we evaluate synthetic enzymes. For example, $k_2 = v_{max}$ is attained when the catalytic sites on the enzyme are saturated with substrate. Hence $k_{cat} = v_{max}/[E]_0$ is the turnover number, that is, the measure of the acceleration of the true chemical step depending on enzyme concentration. While enzymatic rate accelerations are often quantified with respect to the uncatalyzed reaction in water,

$$v_0 \propto k_{cat}/k_{uncat} \tag{3}$$

the ratio k_{cat}/K_M is a measure of the catalytic efficiency

$$v_0 = \frac{k_{cat}[E]_0[S]_0}{K_M} = \frac{k_{cat}[E]_0[S]_0 k_f}{k_{cat} + k_r} \tag{4}$$

Consequently, the most perfect enzymes are ones where $k_{cat} \gg k_r$, whose meaning from Eq. (4) is that only a small fraction of available enzyme is substrate-bound, and therefore limited by the diffusion of the reactants to the active site rather than the formation of the $E \cdot S$ complex.

While in practice the performance of an enzyme also frequently depends on electron or functional group carriers which binds the active site together with the reactants, called coenzymes, the Michaelis-Menten equation is a useful way to compare many natural enzymes that span a wide range of efficiencies.[2] Superoxide dismutase for example[7,8], catalyzes the superoxide radical to more benign products with a k_{cat}/K_M of 10^9 $M^{-1}s^{-1}$ that is close to diffusion controlled. Other enzymes are much less efficient, like the digestive enzyme Chymotrypsin that is characterized by a k_{cat}/K_M of 9.3 $M^{-1}s^{-1}$.[9] It is worth noting that the

activity of natural enzymes is not always optimized for catalytic speed or efficiency, but rather that it can be regulated to adapt to varying physiological conditions in living cells for functional purposes. For example, product formation can sometimes slow down the reaction via a feedback inhibition loop.[10-12] Nevertheless, an average enzyme provides 10^8-10^{11} improvement in reaction rate compared to the uncatalyzed reaction in water[13] and can reach up to 10^{20} in the remarkable case of arginine decarboxylase.[14]

1.2 Mimicking enzymes for reactions not found in nature

The high level of efficiency and selectivity of enzymes, as well as their mild operating conditions, make them relevant to the industrial world where facile high throughput catalysis is required.[15-17] Long used in food and detergent industries, they are now a growing part of a broader range of applications such as biomedicine, drug production, bioremediation of pollutants or green chemistry.[18-21] However, the use of enzymes in broader ranges of chemistry demands that we expand the range of reactions they can catalyze[17,22], as well as requiring enzymes to function in alternative solvents or over a wider range of temperature and pressure[23]. Note that the goal is not to reproduce the complex architecture and interactions of biomolecular network assemblies, within which an enzyme is adapted to suit the homeostasis of living organisms. Rather, we would like to isolate, understand, and exploit the main contributors to their catalytic performance and in particular their exceptional values of k_{cat} and K_M .

Theoretical estimates of the catalytic rate (of any catalyst) are typically formulated within the framework of transition state theory (TST)[3]:

$$k_{cat}/K_M = A(T)\exp(-\Delta G^{\dagger}/RT) \quad (5)$$

where $\Delta G^{\dagger}$ is the rate-limiting activation free energy, and the pre-exponential term A(T) that depends on the temperature T incorporates additional model factors such as the transmission

coefficient (adding quantum effects by setting a non-zero probability for crossing the transition state below the barrier height) and friction (accounting for solvent dynamics that govern the climbing of the reaction barrier), as well as ways to incorporate the diminishment of the upper bound rate factor due to recrossing events (e.g. when the reaction does not proceed to the products after crossing the transition state but rather turns back to the reactant well).[24,25]

However, equation (5) clearly states that reducing the free energy barrier has the most impact on the reaction rate given its exponential dependence. To do that, we would need an atomistic picture of the $E \cdot S$ complex and all the molecular interactions involving the reactant and transition states that make up the relative free energy. Consequently, some researchers rely on approaches that do not require a detailed understanding of these factors to tailor and create novel enzymes, which is best exemplified by laboratory directed evolution.[26]

1.3 Laboratory Directed Evolution

Laboratory Directed Evolution (LDE)[26] makes use of the natural evolvability of proteins to create new functional states by high-throughput screening of large libraries to tackle a diverse range of protein function including binding selectivity, protein stability, and biocatalysis.[15,27-29] For the case of biocatalysis, LDE exploits the fact that, although highly selective to specific reactants, enzymes also exhibit low levels of activity for side reactions involving non-native substrates or product enantiomers.[30] This catalytic promiscuity allows for evolutionary paths that create new enzymes. This is illustrated in Figure 1 where a high peak in efficiency is shown for enzymes catalyzing their native reactions (on the left in green). The important point is that this efficiency does not die off entirely as we move away from the original protein sequence (via random mutations). When these mutated enzymes fall into a region in sequence space that is also characterized by a weak efficiency for a non-native reaction (overlap region in red in Figure 1 or

region of catalytic promiscuity), there is the possibility to preserve mutations that lead to the enhancement of the catalytic performance for the new reaction (evolutionary trajectory).

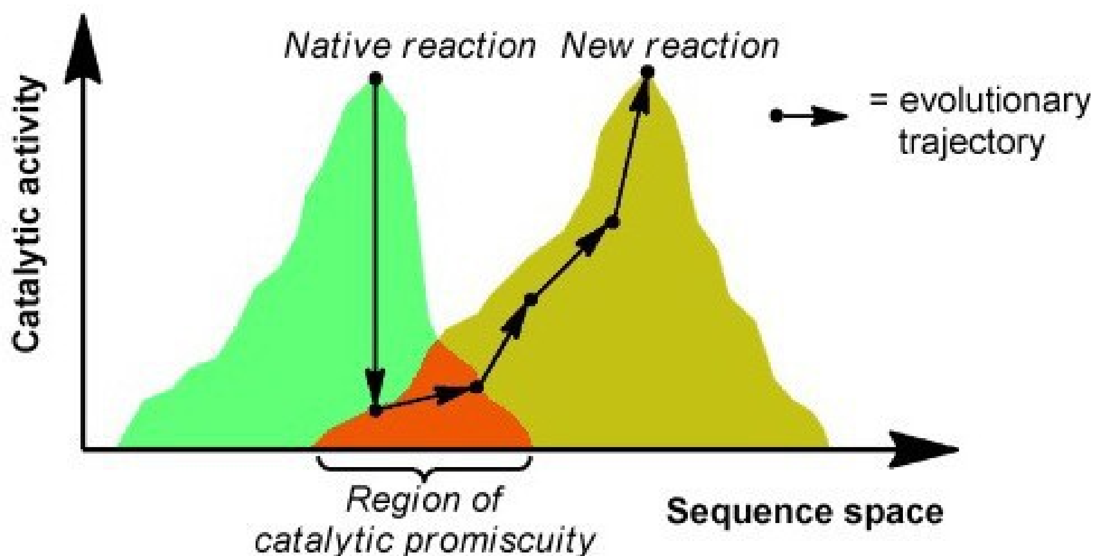


Figure 1: *Natural evolution of enzymes*. Spontaneous mutations lead to the subtle change of sequence space within the enzyme's region of promiscuity. An uphill evolutionary walk then leads to a new enzyme that efficiently catalyzes a new reaction.[30] From Renata et al. Angew. Chem. Int. Ed. 2015, 54, 3351. Copyright © 2015 by John Wiley Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.

While natural evolution of spontaneous mutations occurs over millions of years, LDE operates on a much shorter timescale that is achieved by techniques that accelerate the exploration of a gene's sequence space[31] using multiple rounds of mutagenesis to quickly create new variants.[32] In addition, since the desired new function is different from natural evolution, methods are required to identify and isolate variants with the desired properties (i.e., functional screening). Taken together, LDE is an iterative method that provides a stepwise, and often substantial, improvement in a desired targeted property such as catalysis.[15,33,34] Two recent reviews by Renata el al.[30] and Packer et al.[31] extensively describe the principles and techniques of LDE and how it has become the go-to method for refining biocatalysts.

One of the many successful examples of LDE is the tuning of wild-type halohydrin dehalogenase (HHDH) to catalyze the conversion of ethyl (R)-4-cyano-3-hydroxybutyrate (HN) for commercial purposes[15], to replace energy inefficient chemical cyanation performed at elevated temperatures and pH. Fox and coworkers worked from the versatility of HHDH's active site and tuned its activity to accept the non-natural nucleophile CN^{-}. They combined the traditional hit-shuffling approach with a method based on protein sequence activity relationship (ProSAR), which they deemed necessary to identify beneficial mutations even in variants with reduced functionality. With ∼35 mutations, the volumetric productivity for the desired product was enhanced 4000 fold, reaching the necessary commercial levels of efficiency. Another exceptional example of LDE is the improvements made to the computationally designed Kemp Eliminase HG3[35] that reached a natural enzyme efficiency after 17 rounds of directed evolution[36].

1.4 Computational enzyme design

Despite LDE's undeniable advantages, the fact that it can be immune to understanding the enzyme mechanics is ultimately a limitation for rational and robust enzyme design. This is not to say that attempts to rationalize LDE have not been put forward, or that useful strategies to make LDE more efficient have not been advanced.[28,37-39] But ultimately LDE performs best when given a good starting point, which is a very difficult task without the knowledge of what makes enzymes efficient, because the quality of said starting point is unknown. However, this problem can be addressed by state-of-the-art computational techniques that can be used to test and refine our understanding of enzymatic activity at the molecular level.[40-43]

The origin of enzymes' remarkable performance has been the subject of extensive theoretical and computational research over the past several decades to design synthetic biocatalysts. As reviewed by Korendovych and DeGrado in 2014, computational design of

catalytic proteins has had a boom and bust cycle, progressing from catalytic antibodies through to design of complete protein folds with catalytic function (especially for metalloenzymes) to so-called *de novo* enzyme design.[41] The most recent computational design effort is exemplified by the work of Baker and colleagues who created a series of synthetic enzymes: for the retro-aldol reaction based on four different catalytic motifs for the catalyzed breaking of a carbon-carbon bond[44], the Kemp Elimination reaction as a prototype for catalyzed proton abstraction from carbon[45], the stereoselective Diels Alder reaction involving two substrates forming two new carbon bonds[46], as well as catalytic triads with nucleophilic serine for hydrolases[47]. Overall the computationally designed enzymes do not perform very well, sometimes performing no better than small molecular catalysts[41], and must ultimately rely on LDE to reach respectable enzyme activity outcomes.

However, a great deal is being learned to overcome the limitations, and the remainder of this review covers the progress made for computational enzyme design over the last ~ 5 years. We start in Section 2 with enzyme design approaches that focus on active site optimization, which we refer to as the Standard Model, in the spirit of Sauer and colleagues who have advanced a similar computational protocol for zeolites that also strongly emphasizes the detailed active site.[48] For biocatalysts, the Standard Model focuses on methods that allow the precise chemical positioning of the substrate in the active site to promote enzymatic efficiency[49] including desolvation effects[50], ground-state destabilization[51-53], side chain packing interactions, and hydrogen bonding networks[51,54,55]. It is these molecular effects involving the active site that are the most considered aspect of enzyme design, and rightfully so, given the spatial proximity to the substrate.

In Section 3 we consider enzyme design criteria that move beyond the Standard Model by considering the entire enzyme as a complete catalytic construct. We begin this Section with a brief review of complete *de novo* design approaches that build an artificial enzyme in its entirety. But one of the more lively debates[56] to explain natural enzyme performance, and thus which might be exploitable in biocatalytic design, centers around the pre-organization concept for transition state stabilization introduced by Warshel and colleagues[57-59] vs. concerted dynamical organization of enzyme motions that aid the reaction progress to the product state[60,61].

Since the mid-1980's the preorganization view has advanced the important idea that electrostatic interactions that are unique to the complete enzyme environment contribute substantially to the reduction of the reaction activation energy relative to that in water, by stabilizing the charge distribution of the substrate in the transition state.[57,58] This effectively reduces the reorganization cost associated with the charge redistribution occurring during the enzymatic reaction, a cost that must be paid for in the uncatalyzed reaction in water. However, it is only recent that reorganization energy minimization and electric field optimization have been considered and developed as part of a biocatalytic design strategy.

More recently, enzyme motion that is dynamically organized to aid the catalytic step is thought to be another and equally important factor in enzyme performance.[61] In this case one needs to define what is meant by dynamics. Most often it actually refers to entropic thermodynamic forces that are manifested as the natural statistical fluctuations of a thermalized system.[62,63] By contrast, others have illustrated how dynamical timescales[64], sometimes formulated as time correlation functions of specific vibrational states[65,66], can promote the reactive chemistry. Finally, a third definition is based on a chemical dynamics picture[67,68], i.e. the net reactive flux through the dividing surface at the transition state[69], which also accounts for the

recrossing effects not captured through a simple description based on transition state theory.[70-74] Ultimately, all of these molecular forces and dynamical effects play in concert to more or less effect on the catalytic step, but it is the current dichotomy of electrostatic preorganization and dynamical effects that we review in this last Section, which we demonstrably show can impact our ability to optimize synthetic enzymes.

2. The Standard Model for Enzyme Design

2.1 Design protocol for active site chemistry

We begin by describing the accepted and standard protocol for computational enzyme design, which revolves around the careful positioning of an active site chemistry into a natural scaffold. The general standard model protocol consists of multiple steps which are illustrated in Figure 2. First, a theozyme is constructed with high level quantum chemistry methods that resolve the structure of the transition state as well as a few residues with functional groups assisting the chemistry of the reaction.[75] The geometry of the system is then optimized so as to mimic enzymatic mechanics for stabilizing the transition state.

Second, a search is carried out to place the theozyme into the predefined active site of an existing protein scaffold; a number of straightforward algorithms used to achieve this task include ORBIT from the Mayo group[76] and Selection of Active/binding Sites for Enzyme Design (SABER) from the Houk group[77,78] which searches the Protein Data Bank (PDB) for proteins that already have the catalytic functionality of the theozyme in place. These algorithms are based on a statistical analysis of structural similarity[79] where atomic geometries in a protein are searched to locate specific 3D arrangements that correspond to the designed theozyme. Once a match is found, mutations are sought to accommodate the new substrate in its transition state geometry within the protein scaffold. Another similar algorithmic approach is RosettaMatch[80-82] where the search is able to accommodate both the new substrate and the active site residues of the

theozyme. Although more mutations are necessary with RosettaMatch than with SABER (that only incorporates the substrate of the theozyme), risking the potential to severely disrupt the stability or the performance of the resulting enzyme, it is often seen as a superior technique for designing enzyme for non-biological transformations since the entire theozyme can be incorporated as a whole, and is more likely to catalyze the desired reaction.[82]

Once a match is found for the new substrate in its transition state geometry and/or the new residues directly involved in the chemistry of the reaction, the remainder of the active site has to be reoptimized for both its sequence and structure. This is usually performed with automated software such as RosettaDesign[80-82], which is based on a Monte-Carlo algorithm to search for alternative rotamers for the active site residues that pack well onto the target active site structure at room temperature and satisfying native-like hydrogen bonding.

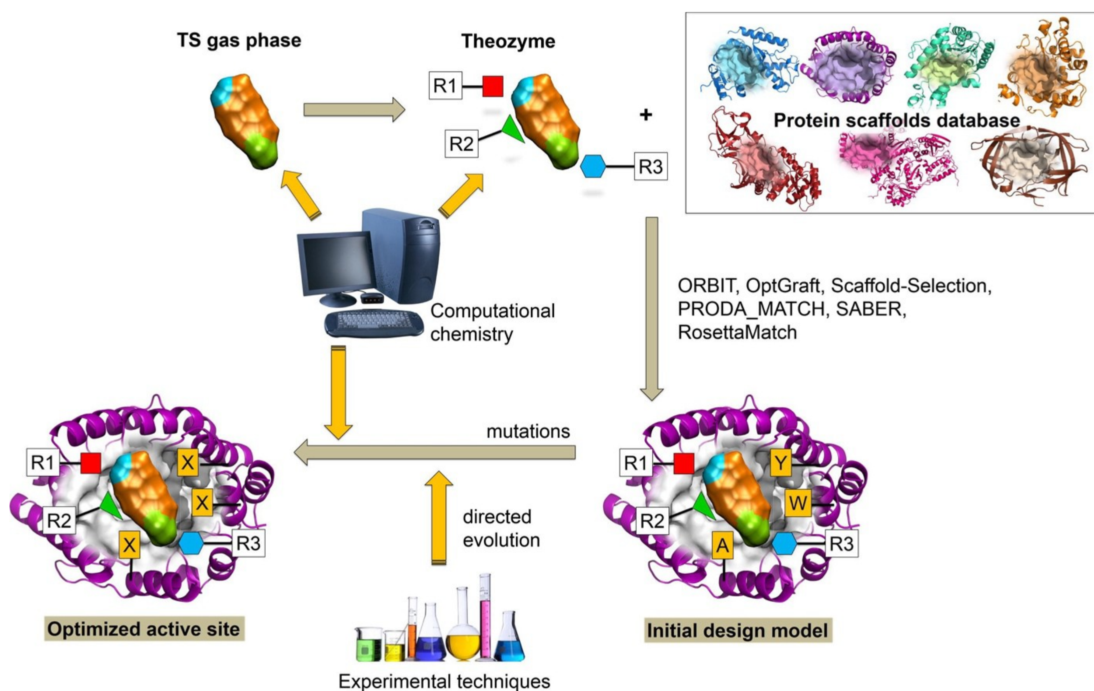


Figure 2. *Standard Model for de novo enzyme design.*[83] The theozyme is built around the transition state geometry resolved by quantum chemistry methods and matched to a pre-existing scaffold. Final designs to be tested experimentally are then generated by re-optimizing the protein side chains to better accommodate the new active site. Not represented in this diagram is

an intermediate step where the design sequences are assessed for catalytic performance prior to experimental validation. Reproduced with permission from Swiderek et al. Arch. Biochem. Biophys. 2015, 582, 68. Elsevier.

These three steps are usually performed for multiple theozymes that involve different motifs for the targeted catalytic reaction.[44,45] For example, the Kemp Eliminases from the Baker group are designed around multiple types of catalytic bases including aspartic acid, glutamic acid, or a His-Asp dyad for direct proton abstraction from the 5-nitroxibenzole substrate.[44] Additional theozyme features are chemical groups that can help align the substrate or support a charged transition state, which can be accomplished with different amino acid side chain chemistries.[82]

An additional and vital final step of the design process is to computationally assess the performance of these enzymes before committing experimental resources to perform the molecular biology to create the designed sequence. Ruscio and co-workers used all atom molecular dynamics (MD) simulations to ascertain whether the expected catalytic activity was possible or whether undesired water accessibility to the active site was impeding the design of *de novo* retroaldolases[84], as did Alexandrova et al who used QM/MM Monte Carlo with Free Energy Perturbation (FEP) for testing the design of Kemp Eliminases[85]. Although using MD to test synthetic enzyme performance has become more routine[82,86], the computational design field still relies on experimental testing, screening and LDE to select the best performing synthetic enzyme.

2.2. Recognized limitations and improvement strategy

Since the first artificial enzymes were proposed, a significant effort has gone into analyzing their performance and understanding their shortcomings. Largely, the goal of a robust Standard Model

for enzyme design – active site accuracy – was not sufficiently executed such that many of the *de novo* enzymes failed to show significant catalytic activity.[41] Often catalytic activity enhancements are realized by subjecting them to LDE, demonstrated especially for the impressive improvements for the HG3 sequence evolved to HG.17.[36] But for some computational designs, catalytic improvements under LDE rapidly reached a plateau with additional LDE rounds, as seen for the KE07 and KE70 Kemp Eliminase designs.[87,88] Often analysis of the LDE sequence trajectories can provide valuable feedback for the design algorithms[89], but lessons drawn from poorly designed starting sequences should be viewed with caution. For example natural enzymes nearly always exploit transition state stabilization as the means for lowering the activation free energy barrier, but many of the Kemp Eliminases that were improved under LDE were shown to have benefitted mostly from destabilizing the reactant state.[90]

Although some of these limitations come from the fact that the active site is designed in the absence of the surrounding scaffold, other more easily resolved issues were identified. For example, MD studies of designed enzymes that failed to show significant catalytic activity revealed that hydrogen bond networks are not always maintained.[35,46,82] This has driven the development of automated protocols that would include non-bonded interactions in the design process, like Foldit[91] or EDGE[82,92].

Similarly, some of the early *de novo* designed Kemp Eliminase enzymes[45] were produced with earlier versions of RosettaMatch and RosettaDesign, both of which have evolved to incorporate features that were identified as necessary for enzyme performance, such as steric packing or π−π stacking.[81] Rajagopalan and co-workers found that the actual geometry of the catalytic triad for hydrolases match the design models much better than their earlier Kemp Eliminase designs, validating the changes they made in their method when using quantum

mechanics/molecular mechanics (QM/MM) potentials for the design.[47] It is also now recognized that iterative design rounds are more beneficial. This means that the rotamer sampling from RosettaDesign for example should be performed multiple times, interspersed with minimization of the side chains, backbone, substrate conformation and rigid body positions.[82,93-96]

Another limitation that was identified from MD studies was the presence of water in the active site that hinders the chemical reaction.[97,98] This is especially important for catalysis of reactions that depends on a carboxylate base for example, such as the Kemp Eliminase series. The solvent-like pKa of the designed catalytic base in the native active site of HG3 led Privett et al to shift the theozyme into a pre-existing small pocket deeper inside the beta barrel instead.[35]

Additionally, it has been observed that the design sequences were not dynamic enough to accommodate the substrate structural changes as the reaction occurs.[99] This suggests that dynamics simulations need to be integrated within the design protocol to produce sequences that better model the true enzyme motions.[82,84,85] Significant improvements can also be made by considering at least partial flexibility of the protein backbone that serves as a scaffold or the use of non-canonical amino-acids. These flexible models are now possible thanks to advances in quantum chemistry methods and QM/MM in particular, which provide a more accurate potential for enzymatic complexes.[83] This also unlocks new possibilities and access to novel protein folds. For example, the creation of a new topology, the top7 alpha/beta was attributed to the flexible backbone minimization step in the design protocol.[100,101]

2.3 Active site positioning and molecular dynamics

A good example of a rational redesign of an active site to address the limitations discussed in Section 2.2 is illustrated using the retroaldolase (RA) design RA22 which exhibited a k_{cat} / K_M = 0.018 ± 0.006 M^{-1} s^{-1}.[44] The retro-aldol reaction involves two primary steps: an initial

nucleophilic attack to form a covalently bound complex followed by proton transfer and subsequent carbon-carbon cleavage using a His-Asp dyad as the catalytic base.[44] Even though the crystal structure of the same RA protein with a single point mutation showed that the RA22 enzyme satisfied the design criteria for substrate positioning, Ruscio et al determined a multitude of problems in the active site that explained RA22's poor performance that showed up under a dynamical assessment.[84]

The first identified issue is that the nearest attack conformation (NAC)[102,103], requiring that lysine ε-amino group and the substrate ketone carbon are at a distance (< 3.25Å) that resembles the bond to be formed in the transition state (Figure 3a), was not satisfied for the intended design for which the substrate is π-stacked in a T-shaped geometry with respect to the his-asp dyad (Figure 3b). Instead, the substrate when dynamically sampled was found to prefer a π-π stacking arrangement with the catalytic base, with participation of scaffold residue Phe211 (Figure 3c), which in fact did meet the NAC criteria for the formation of the imine.

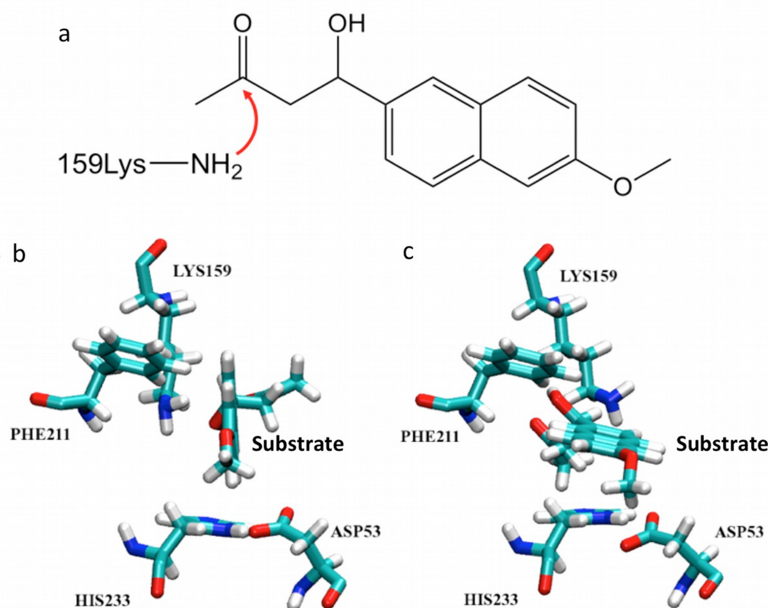


Figure 3. *The computational design of the retro-aldol reaction catalyzed by de novo design enzyme RA22.*[84] (a) the first step is the nucleophilic attack of lysine to create the imine. (b) The

original active site design by Jiang and co-workers[44] did not satisfy the nearest attack conformation (NAC) under the molecular dynamics simulation. (c) The alternative and highly preferred active site determined from Ruscio et al using molecular dynamics simulation did satisfy the NAC for formation of the imine. Adapted with permission from J. Z. Ruscio et al. J. Amer. Chem. Soc., 2009, 131, 14111-14115. Copyright 2009 American Chemical Society.

However the unintended consequence of the π-π stacking arrangement is that it is subsequently poorly organized for the second step, in which the geometric criteria for substrate positioning of the original intended design[44] (Figure 4a) is grossly violated[84]. Furthermore, the MD revealed the excessive fragility of the hydrogen bonded networks (d_1 and d_2) to thermal fluctuations in the active site, such that it failed to form the His-Asp dyad, thereby weakening the ability of the catalytic base to perform proton abstraction from the reactant imine intermediate. Furthermore, a large empty space forms between Phe211 and the enzyme active site, which results in undesired exposure to aqueous solvent.

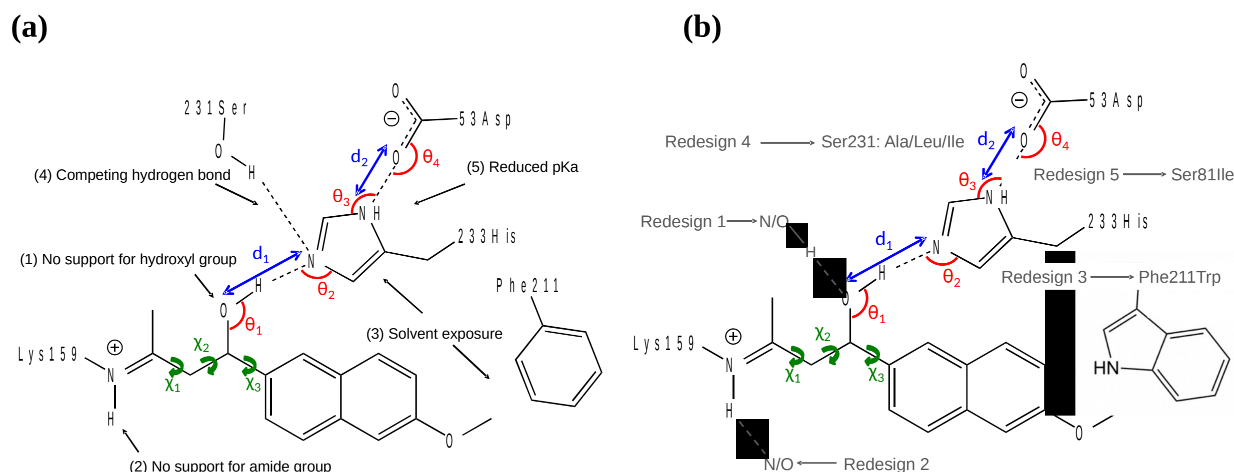


Figure 4. *Active site redesign of de novo enzyme RA22 using the Standard Model.* (a) The enzyme-ligand complex for proton transfer and subsequent carbon-carbon cleavage is not well organized due to poor chemical positioning of the substrate, an empty cavity in the active site filled with solvent waters that disrupt the proton transfer step, and competing hydrogen-bonding to His 233 that destabilizes the His-Asp dyad. (b) Proposed redesign of RA22 enzyme active site alleviating the problems identified in (a).

Consequently, a redesign of the RA22 enzyme would ideally aim to reach the following goals: (1) reorient the ligand into the T-shaped geometry through mutations that can form a hydrogen bond with the substrate hydroxyl, (2) simultaneously improve the NAC under the T-shaped orientation through a scaffold residue that can hydrogen bond to the substrate amide group, (3) introduce large and bulky side chain that can shield the active site from solvent waters, (4) eliminate the competing hydrogen bond from Ser231 that disrupts the hydrogen bonds between the ligand hydroxyl group and Asp53 with His233, and (5) the (trivial) increase in pKa of the His-Asp dyad through creation of a more hydrophobic environment (Figure 4b). All of these criteria are also required to remain robust to the expected protein thermal fluctuations. Nonetheless, this redesign approach is missing conceptual factors beyond the Standard Model for enzyme design that is further considered in Section 3.

3. Beyond the Standard Model for Enzyme Design

What makes enzymes unique relative to other man-made catalysts is that they have changed over evolutionary timescales to be entire catalytic entities. We therefore must look beyond the enthalpic interactions of the active site and more fully consider the greater environment, which can play a highly non-trivial role in achieving gains in catalytic activity.[104] In this Section we look at specific additional entropic and dynamical factors and whole enzyme features that could be used in the design of synthetic enzymes to improve their performance beyond the Standard Model approach.

3.1 Complete de novo design

Artificial metalloenzymes, consisting of a scaffold that hosts a catalytically active transition metal complex[19], are popular synthetic targets for design given their industrial importance. Recently, Hartwig and coworkers reported performance similar to those of natural enzymes after

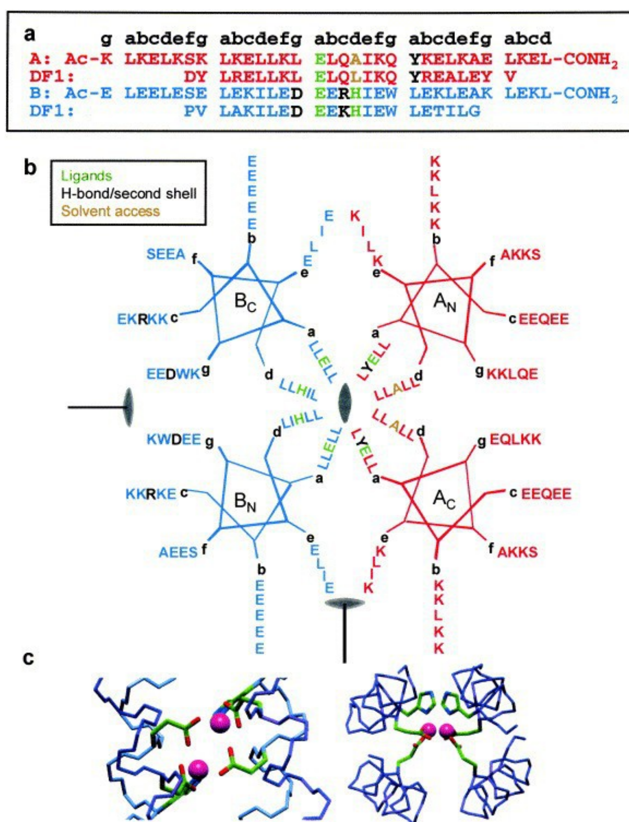
the iron center was replaced by an iridium porphyrin in an existing protein scaffold for an artificial metalloenzyme.[105] We refer the reader to two recent reviews for additional relevant background on artificial metalloenzyme design.[106,107]

When there isn't a natural enzyme to start from, new biological catalysts have to be created from scratch. This requires us to address the very challenging inverse protein-folding problem, meaning that we need to find protein sequences that fold into specified three-dimensional structures. To reduce the degree of complexity of this problem, researchers have focused on α-helical barrels, with helices organized around a central channel with a metal center.[108] The key to efficient metalloenzymes is thought to be in the design of the second co-ordination sphere provided by the biological scaffold[109,110], i.e. a requirement that considers the greater environment beyond the active site.

De Grado and coworkers designed hererotetrameric four-helix bundles to probe the structural basis for the diversity of natural diiron proteins.[111] The A_2B_2 sequence was designed, starting by specifying the backbone, with a novel computational approach that considers the stabilization of the desired fold as well as the destabilization of likely alternatives.[112,113] The structure is shown in Figure 5 and exhibits approximate D_2 symmetry with three orthogonal 2-fold rotational axes. Note also that the active site binds Zn(II) and Co(II) and shows ferroxidase activity under single turnover conditions, as in naturally occurring diiron proteins.

Figure 5. *Four-helix hererotetrameric A_2B_2 bundle designed as a diiron protein model by De Grado and coworkers.*[111] The designed sequence (a), structure (b) and active site (c) is reproduced with permission from Summa et al. J. Mol. Biol. 2002, 321, 923-938. Elsevier.[111]

In a later design of model diiron proteins for the phenol oxidase reaction, they turned their original four-chain heteroatomic helical bundle, DFtet, into a symmetric complex with



identical subunits.[40] In this case, the active site was then designed to allow for the binding of the 4-aminophenol substrate. This included reducing the steric bulk of a couple of residues to fully accommodate the new substrate within this pocket. The catalytic rate of this design was found to be higher than previous *de novo* proteins, with a $k_{cat} / K_M = 1500\ M^{-1}min^{-1}$, thereby representing an excellent step forward in the complete and whole design of a synthetic biocatalyst. Another significant advance was made by Burton et al. who designed a fully *de novo* construct for the hydrolysis of p-nitrophenyl acetate whose catalytic efficiency matches those of redesigned hydrolysases based on natural protein scaffolds.[114] Their strategy was to build an alpha helical barrel with seven helices arranged around an accessible channel and propose mutations to polar residues within the lumen of the barrel to assist the reaction. Overall these fully realized enzyme constructs offer an important step forward beyond the Standard Model of enzyme design which focuses on active site chemistry only.

3.2 Entropic effects on activation free energy

From a thermodynamic perspective, the enzymatic activation energy in Eq. (5) can be decomposed as $\Delta G^{\dagger} = \Delta H^{\dagger} - T\Delta S^{\dagger}$, where $\Delta H^{\dagger}$ and $\Delta S^{\dagger}$ are the enthalpy and entropy change associated with the catalyzed reaction, respectively. The Standard Model of enzyme design has traditionally focused on development of active site models that optimize enthalpic

interactions while minimizing a more flexible active site to yield better values of k_{cat}/K_M. Typically the MD introduced in Section 2 helps validate that the enthalpic interactions are at least optimal enough to withstand the thermal fluctuations under which enzymes must operate. The question we explore in this sub-section is whether entropic effects can play a more active role in the design process to improve the catalytic rate.

The most popular rationalization of entropic effects in enzymes is the so-called Circe effect, which states that the energy gain upon the favorable binding of the substrate is, at least partially, spent to pay the entropic penalty of destabilizing the ground state.[115] This destabilization mostly comes from translational, rotational and conformational substrate entropies and results in a smaller entropy loss than in solution upon overcoming the activation barrier.[116] However, this principle is not verified in general as Kasemi et al. showed for cytidine deaminase.[117] They argue that while the activation entropy is calculated to be close to zero for the rate-limiting acquisition of the transition state, it is due to a change in the reaction mechanism compared to the uncatalyzed reaction in water rather than from an excess energy from substrate binding.

The thermodynamic drive for substrate binding to the enzyme was also revisited by Boehr et al., in which they found that many protein conformations co-exist and the ensemble experiences a population shift upon substrate binding that redistributes the conformational states.[118] This implies that the ligand merely binds to the most favored conformation and that equilibrium statistical fluctuations are the key to substrate binding and molecular recognition in general.[118] Some of those fluctuations can be trapped by static experimental methods such as room temperature X-ray crystallography, in which Fraser and co-workers found evidence for interconverting rotamer sub-states of the catalytic turnover for cyclophilin A.[63]

Joshua Wand and coworkers used NMR relaxation methods to quantify the contribution of changes in protein conformational entropy to the free energy of ES complexes.[119,120] This is achieved by using NMR measures of motions between conformational states on the subnanosecond timescale as a proxy for conformational entropy that can then be separated from the entropy of solvent. With this technique, they showed that the large reservoirs of entropy that characterize folded enzymes are coupled to the binding of ligands.[121] Further, they demonstrate that this effect is large, yet variable, providing a clearer picture of the mechanics of molecular recognition. In another study, they perturbed Ubiquitin with kilobar pressures, enough to affect side chains and backbone dynamics, but well below the denaturation pressure.[122] They showed that NMR resolved fast internal motions decreases with increasing pressure, thereby decreasing the conformational entropy. These experiments were crucial to building the picture of the thermodynamic architecture underlying enzyme stability and function.

It is worth noting that entropy also impacts the stability of proteins, as evidenced when comparing thermolabile to thermostable enzymes. Indeed, enzymes with high activity at room temperature often exhibit a flexible structure (high entropy) resulting in enhanced thermolability. This was also observed in the *de novo* enzymes KE07 and KE70 where entropy was found to destabilize the ground state and stabilize the transition state.[62] When decomposing the state function into contributions from individual residues (entropy) and correlated motion between side chains (mutual information) across an ensemble of backbone conformations and side chain packings, the Head-Gordon lab found that high mutual information residues annihilate activity when mutated.[62] In contrast, enzymes with rigid structures (low entropy) are thermostable but characterized by a low activity at room temperature. This gives rise to the activity-stability trade-off where thermally adapted proteins are made unstable by improving their activity.[123-125]

While this isn't a necessity for natural enzymes, biotechnological applications will require non-natural enzymes to simultaneously display high activity and thermostability.[126] A clever strategy to address this issue is to increase the entropy change of the reaction by displacing more water molecules in the transition state relative to the ground state, thereby enhancing the catalytic activity.[126-129] This was proven possible in a GTPase activating protein where a positive activation entropy of the rate-limiting step resulted in a 105-fold increase in the rate.[130] Computer simulations later rationalize this entropy increase to be associated with the displacement of five water molecules into the bulk solvent upon formation of the transition state due to the rearrangement of the side chain of an arginine residue in the active site. It is worth nothing that less dramatic structural rearrangements can also lead to an increase of the entropy change by breaking the hydrogen bonds of the water molecules around the substrate, which are weakened in hydrophobic active sites.[126]

This is to be contrasted with the entropic effects of water in the catalytic site of the improved KE07 enzyme when subjected to LDE.[87] Using a 3D spatially resolved two-phase entropy analysis method of a MD trajectory of KE07, Belsare et al found that the active site of the best evolved enzyme KE07-R7.2 enzyme held fewer hydrated waters than the original design in the ligand-bound state.[131] However they were more tightly bound than in the original design such as to contribute to a more negative solvation entropy, thereby destabilizing the reactant state[131] as opposed to stabilizing the transition state as most natural enzymes do[90,132]. The contrast between the rational design of more native-like entropic effects in the GTPase activating protein, and the unintended entropic factors using LDE applied to the Kemp Eliminases, suggests that more can be done at the computational design stage of new synthetic enzymes using entropic principles.

3.3 Dynamical correlations

The first step of biocatalyzed reactions is the shift of the enzyme conformational ensemble upon binding of the substrate.[62,118] This is followed by a thermal search of the conformational space towards the reactive transition state and terminates with the rare barrier crossing event that occurs on a much faster time scale. Thus, dynamical effects will impact the speed at which the system explores the potential energy surface. Indeed, it has been long recognized that, close to barrier crossing, statistical enzyme motions enable progress along the reaction coordinate.[70]

However, there remains a controversy over the true dynamical nature of explicit selective modes, primarily because detangling the time scale of these processes is not trivial.[56,133] For example, imagine these rate promoting vibration (RPV) modes are not in thermal equilibrium; then they will be characterized by an excess energy that would dissipate more slowly.[134] Taking the hydride transfer reaction catalyzed by lactate dehydrogenase (LDH) for example, the donor-acceptor distance fluctuations decay within 200 fs.[56] Donor-acceptor distance fluctuation is a widely used mode in this case because it was shown that an incomplete compression/relaxation of this variable compromises the reaction (Figure 6a).[135] The Schwartz group followed this study by a microsecond-scale all-atom MD simulation of the Michaelis complex of LDH, revealing that it actually consists of different substrates that are interconverting with each other on different time scales. They demonstrated that the dynamics can be treated as two independent loop motions in a relatively rigid framework (Figure 6b).

(a) **(b)**

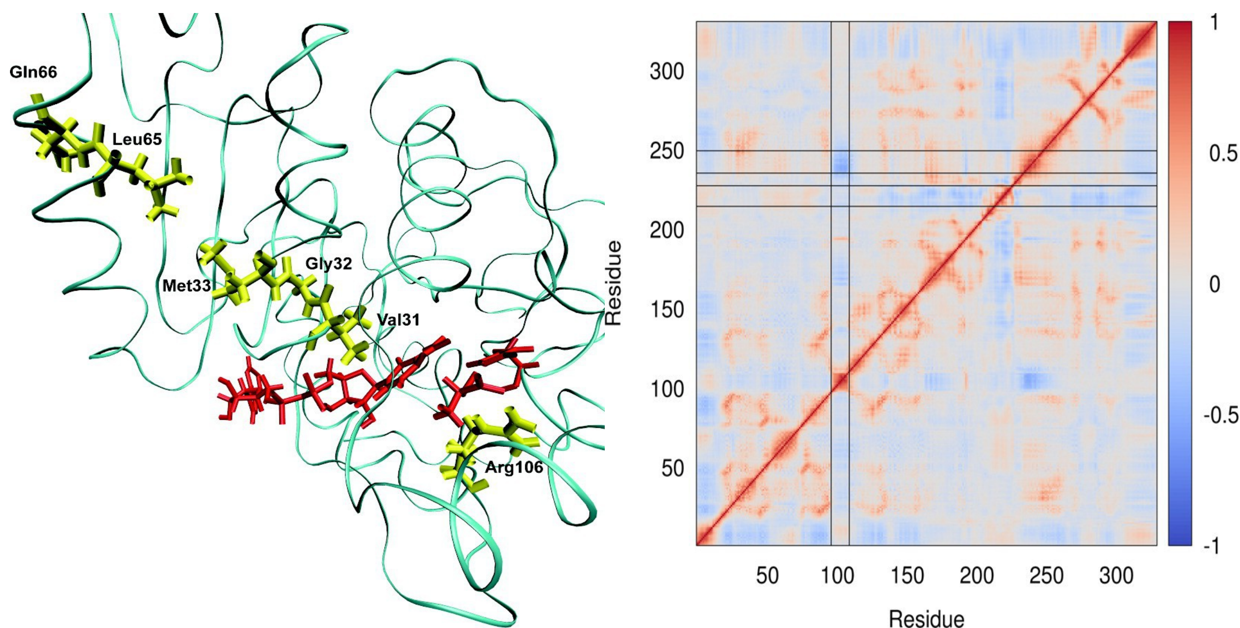


Figure 6. *Dynamical motions of the monomer of human heart lactate dehydrogenase.*[136] (a) The donor (nicotinamide ring) and acceptor (lactate) are showed in red while the promoting vibration residues, are shown in yellow. Reprinted with permission from D. Antoniou et al. J. Phys. Chem. B, 2011, 115, 15147-15158. Copyright 2011 American Chemical Society. (b) the Cα dynamic cross-correlation map[137] shows the intersections of the active site loop (residues 98-110) with the nonactive site loop (residues 215-227) and the contacting helix (residues 236-249) are highlighted. The two loop regions show very weak dynamical correlation, while the correlation between the dynamics of the active site loop and the contacting helix is relatively strong. Reprinted with permission from X. Pan et al. J. Phys. Chem. B, 2015, 119, 5430-5436. Copyright 2015 American Chemical Society.

In another study, they looked at LDHs from two different organisms, which have several structural differences including a five amino acid insertion in one of them. They used transition path sampling and committor distribution to show that RPVs are found across different type of LDHs, suggesting that it is a beneficial feature preserved through evolution. However these vibrations are modulated by structural changes near the active site that can hinder catalytically important motions.[138] To build up on these findings, they proposed a mutation for the aromatic amine dehydrogenase that enhances the natural promoting vibrations while preserving the

electrostatics of the reactive event. Their mutant is expected to lower the free energy barrier of the reaction, hence exhibits higher acceleration rate compared to the natural enzyme.[139]

Focusing on correlated motions with the active site loop, Pan and Schwartz explained how the strong conformational heterogeneity may open alternative mechanisms for the chemical step that need to be accounted for when designing *de novo* enzymes.[137] They extended their analysis to designed Kemp Eliminases and their improved variants under LDE, for which they found that LDE selected mutations favor a more dynamic active site that makes use of the donor-acceptor compression.[140] They propose the fact that since the Kemp Eliminases were designed without accounting for these dynamical effects, it may explain why the design sequences were so poorly efficient.

It is worth noting that the work of Klinman and co-workers on gating motions is very similar to the one of Schwartz et al., but their interpretation of these motions is in terms of statistical fluctuations.[141-143] Similarly, Boekelheide et al. used ring polymer molecular dynamics that includes nuclear quantization effects to look at the role of vibrating motions in the catalytic activity of DHFR. They showed that dynamical correlations are fast but also local, vanishing at distances beyond 4-6 Å. Slow thermal fluctuations from the protein scaffold gate the fast dynamics in the active site, thereby modulating the instantaneous rate for the intrinsic reaction.[74]

Hammes-Schiffer and co-workers[70] proposed a different picture where dynamics play a role on longer timescales, after identifying slower coupled promoting motions for the hydride transfer of dihyfrofolate reductase (DHFR) from the entire protein, and not just the active site (Figure 7). This would suggest that millisecond timescales are critical to the barrier passage. While some of these conclusions were motivated by genomic analysis, Hammes-Schiffer and colleagues also performed QM/MM simulations of DHFR combined with rank correlation

analysis. They observed that single, double and triple DHFR mutants are characterized by different distributions of coupled motions correlated to the catalyzed reaction.[73] Although their method cannot distinguish between motions playing an active role in catalysis and motions responding to alterations caused by catalysis, they argue that since the entire enzyme is coupled via long-range electrostatics and hydrogen bond network, surface mutations can modify the probability of sampling active site conformations that will lead to the transition state.

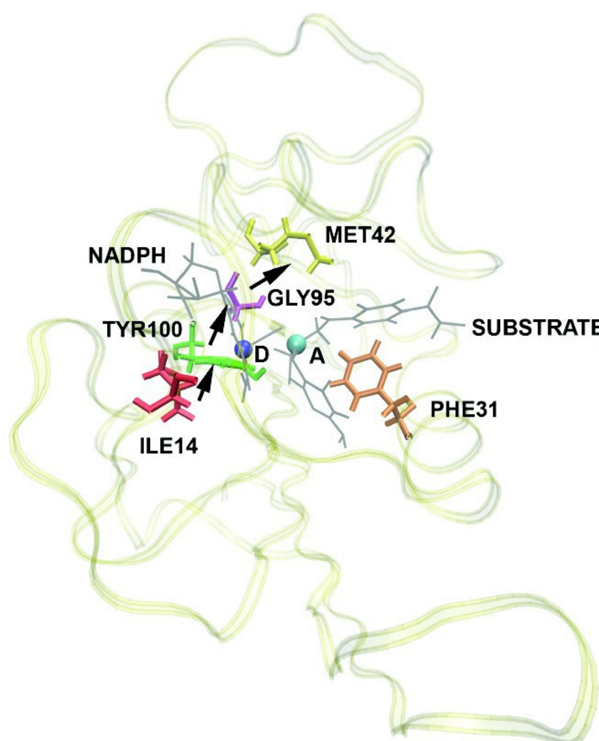


Figure 7: *Diagram showing the residues and their direction of movement as a reaction crosses the separatrix in the DHFR reaction*.[136] Donor and acceptor are shown as a blue and green ball, respectively. There is no well-defined promoting vibration, and the limited protein scaffold prevents the formation of the type of compressive promoting vibration seen in LDH. Reprinted with permission from D. Antoniou et al. J. Phys. Chem. B, 2011, 115, 15147-15158. Copyright 2011 American Chemical Society.

3.4 Minimizing the reorganization energy

Lowering the activation energy via transition state stabilization is one of the main design criteria for efficient synthetic enzymes.[90,132] An important ingredient for this stabilization is embodied in the reorganization energy, the energy cost paid as the enzyme organizationally responds to the changes in substrate structure and charge distribution as the reaction proceeds along the reaction

corrdinate.[99] The ease with which these structural and electronic changes occur along the reaction coordinate will contribute to an enhancement of the enzyme performance.

The reorganization energy due to the changing substrate state can in principle be quantified by Marcus' theory in the weak coupling, high temperature limit.[144-146] Indeed, the use of Marcus' theory implies that the potential energy landscape of the adiabatic process that is an enzymatic reaction, can be approximately described by a pair of crossing diabatic surfaces.[144,145] This is non-trivial because diabatic states are states that do not change character along the reaction coordinate and as such are not easily mapped onto an adiabatic state, which change constantly as to remain the lowest energy state at all times. This means that using diabatic states to describe an adiabatic reaction, which is called the non-adiabatic approximation, only works under specific conditions, namely when the reactant and product states are weakly coupled (weak coupling regime).

Historically, Marcus quantified the kinetics of electron transfer reactions using this non-adiabatic approximation adopting two parabolas for the reactant and product states and the energy gap, q, as the reaction coordinates (Figure 8). Therefore, we can write the diabatic states within this model as $G_{RS}(q)=4\lambda(q-q_{RS})$ and $G_{PS}(q)=G_{RS}(q)+q$, where G_{RS} and G_{PS} are the potential energy curves of the reactant and product states respectively, q_{RS} is the value of the energy gap at the equilibrium of the reactant state and λ is the reorganization energy, related to the curvature of the parabolas. As shown in Figure 8, the reorganization energy can also be defined as the vertical energy difference between the reactant and product diabatic curves, at the product equilibrium point.

Figure 8: *Marcus model for reorganization energy applied to enzyme catalysis.* Diabatic potential energy surfaces of the reactant and product states, which are separated along the energy gap reaction coordinate. In the weak coupling regime, the true adiabatic surfaces only differ from

their diabatic counterparts near q_c, where they avoid crossing. The reorganization energy (λ) quantifies the cost of reorganizing the medium around the changed product charge distribution. Reproduced from[147] with permission from the PCCP Owner Societies.

In the case of enzymes, the reorganization energy quantifies the electrostatic preorganization originally put forward by Warshel to explain the power of naturally occurring enzymes.[99] In this model, the transition state is located at the intersection of the two parabolas, where the two states are degenerate. The rate of reaction can then be derived from Fermi's golden rule in the linear response approximation (LRA), which results in defining the activation energy as:

$$\Delta G^{\dagger} = (\Delta G^0 + \lambda)^2 / 4\lambda \tag{6}$$

where ΔG^0 is the reaction free energy. Alvarez-Paggi and coworkers combined theory and experiments to characterize the reorganization energy of electron transfer reaction in Cytochrome C.[148] They showed that a stiff scaffold helps to minimize the reorganization energy by keeping solvent water molecules far from the redox active site. They further demonstrated that Cytochrome C switches between a high and low reorganization energy conformation via electrostatic interactions. While performed on model electrostatic complexes, this work helps to identify the protein-protein interactions that could eventually minimize λ, hence the energy barrier of the reaction, when interacting with natural redox partners, in agreement with other studies.[149]

Although Marcus theory can be very powerful, it has some limitations for enzyme design. Indeed, many enzymatic reactions cannot be accurately described in the diabatic regime. This is because in condensed phase, chemical reactions exhibit strong coupling between the reactant and product states and the free energy landscape is best described by a continuous surface. Furthermore, it is also traditionally very challenging to reliably compute the reorganization energy, even for small molecules,[150,151] whereas for enzymatic reactions, the effect of both the scaffold and the solvent has to be considered.

However, it is worth noting that the reorganization energy is readily available from empirical valence bond (EVB) methods that have been developed for decades by Warshel and colleagues to quantify enzymatic catalytic effects. In this case, the energy landscape is built by diagonalizing the EVB matrix whose diagonal elements are the state energies. Although applicable to many states, two are generally used, one for the reactant and the other for the product state. The ith diagonal element is then defined as:

$$H_{ii}=\alpha_{gas}^{i}+U_{intra}^{i}(\vec{R},\vec{Q})+U_{inter}^{i}(\vec{R},\vec{Q},\vec{r},\vec{q})+U_{solvent}^{i}(\vec{r},\vec{q}) \tag{7}$$

where $\vec{R},\vec{Q}$ are the atomic coordinates and charges of the reactants and products; $\vec{r},\vec{q}$ are the atomic coordinates and charges of the solvent or protein; U_{intra}^{i} is the intramolecular potential of the solute system relative to its minimum; U_{inter}^{i} the interaction energy between the solute and solvent; $U_{solvent}^{i}$ is the potential energy of the solvent; and α_{gas}^{i} is the energy of the ith diabatic state in the gas phase. The coupling between the states is often approximated to be constant, but a more accurate form involves a functional dependence on the reaction coordinate, $\Delta R'$ (usually taken as distance between reacting atoms);

$$H_{ij}=A e^{-a \vee \Delta R' \vee \text{¿}} \tag{8}$$

The adjustable parameters A and a are used to fit either the reference quantum calculations or experiments. We refer readers to more authoritative reviews on EVB applied to enzyme catalysis for more detail.[58,152]

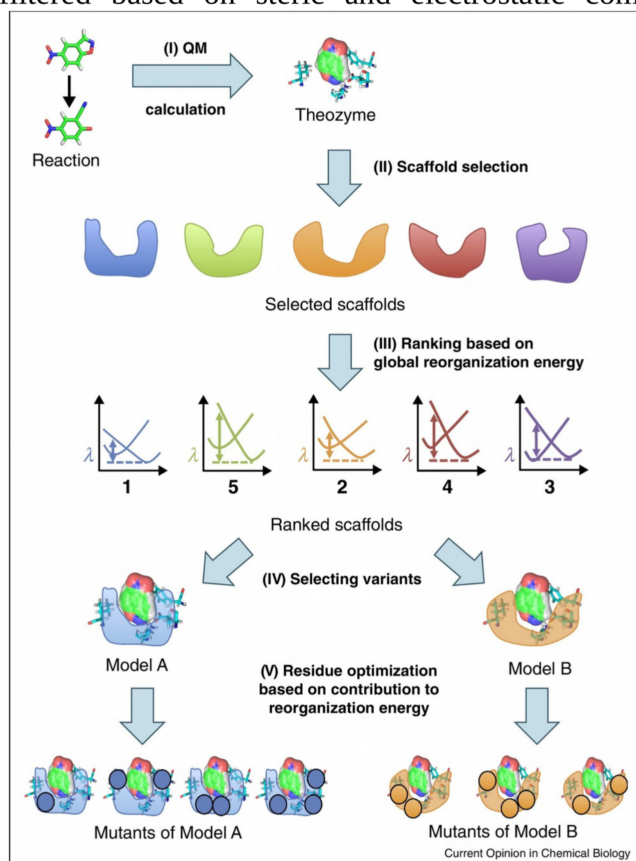
Fuxreiter et al. used the EVB method and all-atom free energy perturbation to characterize the catalytic effects of acetylcholinesterase.[153] Compared to the reaction in water, the enzyme reduces the energy barrier of the acylation step of acetylcholine hydrolysis via electrostatic effects, which come from a reduction of both the energy of the charged intermediate and the reorganization energy. This preorientation results in a smaller reorganization energy and thus reaction activation energy, which contrasts with the higher reorganization energy cost for the reference reaction in water.[154]

Warshel and colleagues have also analyzed the LDE trajectory of the synthetic enzyme KE07 and shown that beneficial mutations resulted in a better electrostatic preorganization around the reaction transition state, hence reduced reorganization energy.[90,132] Labas et al. corroborated these findings by calculating the reorganization energy of the KE07 design and its improved LDE variants.[155] They derived the reorganization energy directly from the EVB diabatic states with which they constructed the free energy pathways of the various mutants. They concluded that the improved KE07 variants are characterized by a reduced reorganization energy, which ultimately drives the evolution of the enzyme.[155] Further, they found that some mutations can cause a decrease in reorganization energy even if not directly beneficial for catalysis.

This suggests that reorganization energy could become an independent property to account for in the functional screening of variants (Figure 9). It is particularly important for distant mutations that link the theozyme to the rest of the scaffold, which are currently

disregarded after screening. This is in line with other studies that show that electrostatic preorganization, hence reorganization energy, could be tuned in the LDE process.[36,44,46] This means that a better starting point for LDE would be a design where reorganization energy is already optimized. Therefore, systematic reorganization energy calculations could help design better starting sequences by changing how we evaluate and rank scaffolds.[99]

Figure 9: *Flowchart of computational enzyme design using reorganization energy calculations.*[99] The theozyme (I) determined by QM methods is docked into different scaffolds (II) and a set filtered based on steric and electrostatic complementarity (III). Reorganization energy is computed for the selected scaffolds, using short MD simulations within the LRA framework (IV). Top-ranked candidates are subjected to further refinements. In addition to TS binding energy and electrostatics, residue contributions to reorganization energy are also evaluated and used for scoring (V). Reproduced with permission[99] Curr. Opin. Chem. Biol. Elsevier.



More specifically, integrating the reorganization energy in the design process would allow us to account for two important factors. First, it would help rank the ability of an enzyme to respond to the changes in charge distribution from the ground to transition state. Second, it would quantify the steric strain on the scaffold if significant geometric deformations occur between the ground and transition states. The design strategy incorporating scaffold rankings based on reorganization energy calculations proposed by Fuxreiter *et al* is shown in Figure 9 is thought to improve on the current

protocol that prioritizes shape and charge complementarity in the active site. This way, the functional screening of proposed variants using reorganization energy will select residues that enable structural changes as the reaction occurs, even if they are not directly involved in catalysis.

3.5 Electric field optimization

The pioneering work of Warshel and colleagues have established that an enzymatic reduction in $\Delta G^{\dagger}$ occurs in natural enzymes because they are electrostatically preorganized to stabilize the transition state relative to the reactant state.[57] While the minimization of reorganization energy is well-appreciated, getting an experimental handle on the exact nature of the preorganized electrostatic environment is only indirectly inferred through measurements of k_{cat}. However, recent developments in spectroscopy have started to remedy this situation by characterizing electrostatic interactions in enzyme-substrate complexes.[156-160]

The most direct advances in electrostatic characterization of proteins have been made by the Boxer group[161], who brought forward the idea that electric fields are an ideal probe for intermolecular interactions introduced by the environment for a substrate molecule modeled as a dipole, which is illustrated for a simple reaction center embedded in water in Figure 10.[162] Within this electrostatic picture, Fried and Boxer suggest that configurations of minimal energy are achieved by the best alignment of the substrate dipole to the field they experience from the surrounding protein scaffold that implicitly takes into account excluded volume, hydrogen-bonding, and other non-covalent factors.[163] Exploiting this concept, Boxer and colleagues developed and used vibrational Stark spectroscopy to measure electric fields in the active site of several enzymes. By transitioning from the traditional optical probes to vibrational probes, which

minimally perturb the system, they are able to probe electric fields in enzyme with high spatial (< 1 Å) and field (< 1 MV/cm) resolution.[51] Such experiments are done in two steps.

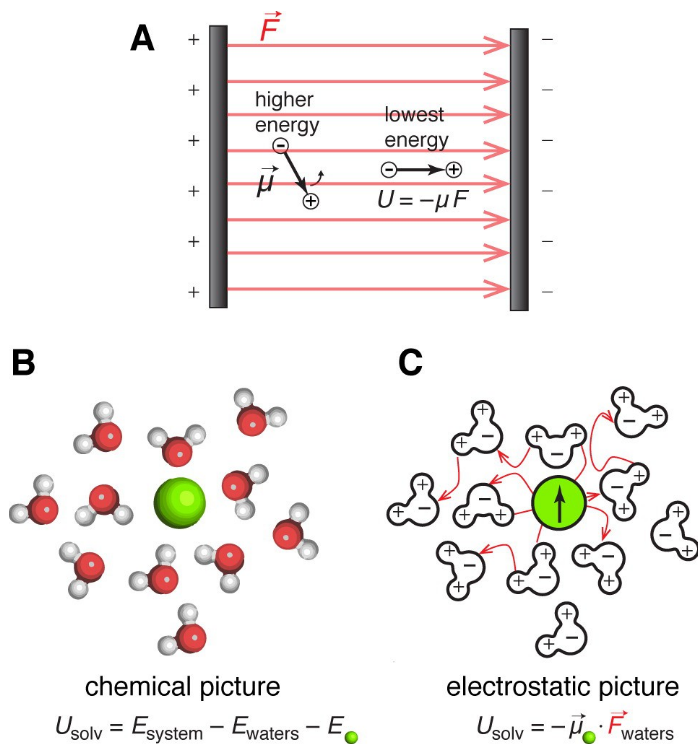


Figure 10: *Connection between electric fields and molecular interactions.[162]* (A) A uniform electric field is created by applying a voltage between parallel plates and interacts with dipole moments. (B) Chemical picture of a molecule interacting with other molecules through non-covalent interactions. (C) Conversion of the chemical picture into an electrostatic one. Reprinted with permission. Copyright 2015 American Chemical Society.

First, the difference dipole ($\Delta\vec{\mu}$) of a given molecule is calibrated by measuring the change in frequency ($\Delta\acute{v}$) caused by a known applied electric field ($\vec{F}_{ext}$) according to:

$$hc\,\Delta\acute{v} = -\Delta\vec{\mu}\cdot\vec{F}_{ext} \tag{9}$$

where h is Planck's constant and c is the speed of light.[51,164] Note that from this equation, the difference dipole (e.g. the difference between the transition state and ground state dipole moments) can be seen as the sensitivity of the molecule to the electric field.[164] This means that the magnitude of the electric field is modulated by the dipole that changes between ground and transition states.

Once the difference dipole is known for a specific molecule, it can be used as a probe in an "inverted" Stark experiment where frequency shifts due to unknown electric fields are

measured. For example, by placing the probe in the active site of an enzyme, the electric field due to the scaffold and surrounding solvent can be quantified according to:

$$\Delta E = hc\,\Delta\acute{v}_{obs} = -\Delta\vec{\mu}\cdot\Delta\vec{F}_{protein} \quad (10)$$

Many such experiments have been performed over the past years to perfect the technique and determine the best vibrational probes for proteins.[158,159,161,164,165] For example, the experiments were found to be helped by calculations of electric fields that various solvents project onto a given probe, which provides a calibration of solvent-induced shifts (solvatochromism).[163,166,167] This leads to a careful mapping of the vibrational frequencies to absolute electric fields.

More recently, Fried et al. used the carbonyl vibration of the inhibitor 19-nortestosterone to probe the electric field of the natural enzyme Ketosteroid Isomerase (KSI), which is characterized by a fast k_{cat} (~10^4-10^5 s^{-1}).[51,160] Looking at various KSI mutants, they demonstrated that an increase in the activation barrier was linearly correlated to the extent of electric field decrease, directly linking magnitude of electric fields to catalytic performance. They concluded that the enzyme produces large electric fields that stabilize the increased dipole of the transition state. Furthermore, they estimated that these electrostatic effects were responsible for 70% of the total rate acceleration of KSI, which is over two orders of magnitude more than the contribution from entropic effects associated with the precise positioning of the catalytic base.[51,160] Although others have contested the assertion that a majority of the rate acceleration comes from electric fields[168-170], it is generally agreed that the Stark measurements support the view that electrostatic interactions from the enzyme play a pivotal role in biocatalysis.[171,172]

We would like to remark that these two views, namely that electrostatic effects (view 1) or chemical positioning (view 2) as the main contributor to enzyme catalysis, are not as contradictory as this debate would suggest. Indeed, strong electric fields emanate from the active

site residues that are actively and directly involved in the chemical reaction. This was specifically demonstrated by Wang et al. who calculated that 98% of the electric field in the active site of KSI comes from the network of short hydrogen bond residues that stabilizes the reaction intermediate.[55] However, since both the magnitude and orientation of these fields matter, a slight change in the position of the catalytic residues will have large consequences on the strength of electrostatic interactions. This means that, in the active site, precise chemical positioning and electric fields are intimately coupled and disrupting one will disrupt the other. In addition, the electric fields from residues in the active site should not be the sole focus of our attention. Electric field contributions from individual residue in the scaffold and hydration water, while individually small, will integrate to something substantial, since every 30 MV/cm stabilization in the exponential is an order of magnitude in rate.[173]

For interpreting or using Stark experiments in the context of enzyme design, at first glance the dipole model for the substrate may seem restrictive or even unwarranted. We would assert that it can serve to express the electronic changes of bonding that are being reformulated within the substrate molecule along the reaction coordinate, that can be modeled as a superposition of changing bond dipoles that differ in the reactant and transition state.[173,174] How these bond dipoles interact with the electric field created by all other residues can give rise to electrostatic stabilization. How much stabilization is determined by the folded structure of the enzyme which has shaped the electric field environments in the active site.[173]

The implications of this view are quite profound for the future viability of the Standard Model of enzyme design, which in most cases has encased the theozyme in an arbitrary protein environment. The Head-Gordon group showed that for the synthetic Kemp Eliminases that the TIM barrel scaffold does not support the reaction in the active site of the designed KE07 and

KE70 enzymes.[173] In fact, they showed that the scaffold and surrounding solvent actually disfavor the Kemp eliminase reaction, because these enzymes were designed without integrating the interactions between the scaffold and the active site, resulting in electric fields that are only locally optimized by the catalytic base Glu-101for KE07 and the His-Asp dyad for KE70. Such an approach can't be completely rescued by LDE as the only viable option available is to optimize the sequence locally near the active site. Otherwise the scaffold would have to be reorganized entirely to create an environment commensurate with the catalytic reaction at the active, which is equivalent to saying that the reorganization energy is so enormously high that LDE is not a practical solution.

This work provided the basis for exploiting at least local electric fields as a design criterion for the active site (in line with the high local electric fields in the active site of KSI [51,55]), and to replace the early rounds of LDE to computationally evolve a synthetic enzyme. Vaissier *et al.* developed an electric field optimization scheme to improve the efficiency of the synthetic Kemp Eliminase KE15, for which no LDE was performed.[175] This was achieved by computing the electric fields emanating from the enzyme's active site and scaffold as well as surrounding water molecules in both the ground and transition states. Considering the sensitivity of the substrate to an external electric field (i.e. the magnitude and direction of the bond dipoles), mutations were proposed as to optimize the electrostatic free energy stabilization of the transition state:

$$\Delta G^{\dagger}_{elec} = -\sum_i \left(\vec{\mu}^i_{EL^\dagger} \cdot \vec{E}^i_{EL^\dagger} - \vec{\mu}^i_{EL} \cdot \vec{E}^i_{EL} \right) \quad (11)$$

Where $\vec{\mu}^i_x$ is the bond dipole of bond i and $\vec{E}^i_x$ is the electric field at *i* evaluated in the ground (x =EL) and transition (x =EL†) states. In this equation, the sign convention is

chosen so that the positive direction of the dipole is aligned with the one of the electric field (promoting the flow of electron in the transition state).

Starting from the design that has a k_{cat}/K_M of 27 $M^{-1}s^{-1}$, individual mutations were proposed to enhance the magnitude and alignment of the electric fields in the transition state relative to the ground state. Remarkably, only 4 computationally targeted mutations were needed to yield a k_{cat}/K_M of 403 $M^{-1}s^{-1}$, with almost all of the enzyme improvement realized through a 43-fold improvement in k_{cat} as measured experimentally (Figure 11). More importantly, most of this improvement comes from a stabilization of the transition state (by about 2.25 kcal/mol) rather than a destabilization of the ground state (~0.7 kcal/mol).

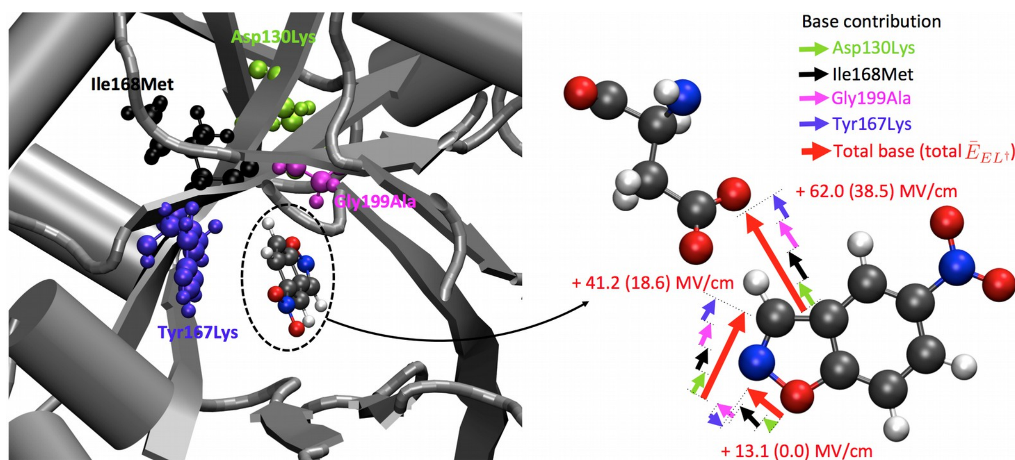


Figure 11. *Optimization of electric fields in the KE15 Kemp Eliminase synthetic enzyme*. Shown is the location of the 4 mutations of KE15 best variant (left) and the total electric field contributions to the transition state relative to the original design (right).[175] The positive direction for the bond dipole is the same as for the electric field (from C to H, from N to C and from O to N for the CH, CN and ON bond respectively with magnitude +1 , +0.4 and +2.3 Debye). The efficiency of these computationally designed variants was measured and confirmed experimentally. The quadruple mutant yields a k_{cat}/K_M of 403 $M^{-1}s^{-1}$ relative to a k_{cat}/K_M of 27 $M^{-1}s^{-1}$ for the design.

The demonstration that true catalytic enhancement can be achieved by incorporating electric field optimization in the design process of enzymes is an important future direction in a range of catalytic systems.[104] While the Head-Gordon lab has created a top-down approach for optimizing electrostatics of an existing enzyme scaffold through mutations[175], Sokalski and co-workers have developed a bottom-up strategy of proposing what would be the optimal electric field environment for a given active site chemistry.[176-178] Motivated by the role of electric fields in biocatalysis, Coote and co-workers have taken electrostatic design in exciting new directions ranging from synthetic chemistry illustrated by the Diels Alder reaction[179] to electrocatalysis[180]. The future import in which electric fields can influence new catalytic constructs and reaction chemistry have also been reviewed very recently.[104,181]

4. Conclusion

Computational enzyme design for chemical reactions that are not part of nature's repertoire has typically served as a plausible beginning for subsequent laboratory directed evolution, which enhances performance through rapid exploration of sequence space using mutation and recombination. While computational design followed by LDE is the obvious and popular strategy for many protein engineering studies, there is no rational way to improve on these constructs given the random sequence mutations introduced. In fact, LDE can also be inherently compromised by the underlying performance of a designed enzyme that has a poorly optimized active site at the start.

The Standard Model for designing synthetic biocatalysts typically revolves around the optimization of the active site chemistry, in most cases almost independently from its integration within the greater protein environment. However, the knowledge acquired over the past few years has the potential to spark a revolution in the field of computational enzyme design. Using

higher-level simulation techniques and analyzing successful as well as failed designed sequences, a few axes have been identified as promising to both improve upon as well as move beyond the standard model for enzyme design (Table 1).

As is always the case, the accurate description of the active site is a paramount – with no well-designed active site there is no designed enzyme for performing catalysis. Hence better theoretical chemistry methods that create more accurate chemical positioning of the substrate is always a desirable improvement, such as obvious use of enhanced QM models that better describe molecular interactions. In fact, any exploitation of the greater protein environment of an enzyme for design will be fruitless without accurate chemical positioning, for example the inability of a correctly organized electrostatic environments to align appropriately due to a poorly organized catalytic active site.

Table 1. *Criteria for computational enzyme design and their potential impact.* We qualitatively rank the impact on the catalytic activity of a synthetic enzyme by considering the transition state theory rate and whether the effect acts on the pre-exponential or the exponential term. Note that the impact is also evaluated given our current approach to enzyme design and by no means sets a ceiling on the potential impact each factor could have.

Criterion	Definition	Relative impact on catalytic activity (today)	References
Transition state stabilization under the standard model (theozyme)	Structure determination of a few active site residues with functional groups assisting the chemistry of the reaction, i.e. stabilizing the transition state.	Large	[44,45,75,82]
Enzyme dynamics	Conformational heterogeneity due to thermal motion of the enzyme and water.	Moderate	[70,118,136,137]
Reorganization energy	Energy cost of reorganizing the enzyme in response to the change in charge distribution as the reaction occurs.	Moderate	[99,148,154,155]
Reaction entropy	Change in entropy as the reaction	Potentially large	[119,120,126,130,62]

change	proceeds, which adds onto the enthalpic stabilization of the transition state.		
Electric fields	Probe for short- and long-range intermolecular interactions stabilizing the transition state in the active site of the enzyme.	Large	[51,104,175]

An additional set of axes take a more wholistic view of the enzyme as a complete catalyst by considering the greater environment in which the active site resides, meaning that the protein is no longer considered as a simple physical support system for the energetics of the active site. These axes include entropic and dynamical effects that span a variety of time and length scales. At the very least, design sequences have to sustain the active site positioning against the natural statistical fluctuations from the thermalized conformational ensemble of the enzyme. We could go further and design a scaffold that accommodates the specific vibrational modes that promote the chemistry of the reaction in the active site. Finally, long-range electrostatic effects can be accounted for by incorporating metrics such as the reorganization energy or electric fields in the design protocol. The latter is especially important because not only does it allow a more accurate model of enzyme activity, it also reconciles our design strategy with our present understanding of enzymatic machinery; that is that enzymes are pre-organized environments that electrostatically stabilize the reaction transition state. This paves the way for building a unifying picture of the molecular interactions responsible for enzymes' incredible performance, hence managing efficient design and enabling further methods development for the accurate simulation of large biomolecular catalysts.

ACKNOWLEDGMENTS. VVW and THG thank the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division under Contract No. DE-AC02-05CH11231. The material on theory and methods is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) program. We thank Po Hu for his earlier work on the RA22 improvements.

5. Bibliography

(1) Mak, W. S.; Siegel, J. B. Computational enzyme design: Transitioning from catalytic proteins to enzymes. *Curr. Opin. Struct. Biol.* **2014,** *27*, 87-94.

(2) Wolfenden, R.; Snider, M. J. The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res.* **2001,** *34*, 938-945.

(3) Fersht, A. *Enzyme structure and mechanism*; WH Freeman: New York, New York, 1985.

(4) Michaelis, L.; Menten, M. L.; Johnson, K. A.; Goody, R. S. The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochem.* **2011,** *50*, 8264-8269.

(5) Michaelis, L.; Menten, M. M. The kinetics of invertin action. *FEBS Lett* **2013,** *587*, 2712-2720.

(6) Briggs, G. E.; Haldane, J. B. A note on the kinetics of enzyme action. *Biochem. J.* **1925,** *19*, 338-339.

(7) Hsu, J.-L.; Hsieh, Y.; Tu, C.; O'Connor, D.; Nick, H. S.; Silverman, D. N. Catalytic properties of human Manganese Superoxide Dismutase. *J. Biol. Chem.***1996,** *271*, 17687-17691.

(8) Becker, M. N.; Greenleaf, W. B.; Ostrov, D. A.; Moyer, R. W. Amsacta moorei entomopoxvirus expresses an active superoxide dismutase. *J. Virol.* **2004,** *78*, 10265-10275.

(9) Wilcox, P. E. Chymotrypsinogens—chymotrypsins. *Methods Enzym.* **1970,** *19*, 64-108.

(10) Umbarger, H. E. Evidence for a negative-feedback mechanism in the biosynthesis of Isoleucine. *Science* **1956,** *123*, 848-848.

(11) Weber, G.; Convery, H. J. H.; Lea, M. A.; Stamm, N. B. Feedback inhibition of key glycolytic enzymes in liver: action of free fatty acids. *Science* **1966,** *154*, 1357-1360.

(12) Datta, P.; Gest, H. Control of enzyme activity by concerted feedback inhibition. *Proc. Natl. Acad. Sci.***1964,** *52*, 1004-1009.

(13) Zhang, X.; Houk, K. N. Why enzymes are proficient catalysts beyond the Pauling paradigm. *Acc. Chem. Res.* **2005,** *38(5)*, 379-385.

(14) Marti, S.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J. Computational design of biological catalysts. *Chem. Soc. Rev.* **2008,** *37*, 2634-2643.

(15) Fox, R. J.; Davis, S. C.; Mundorff, E. C.; Newman, L. M.; Gavrilovic, V.; Ma, S. K.; Chung, L. M.; Ching, C.; Tam, S.; Muley, S.et al. Improving catalytic function by ProSAR-driven enzyme evolution. *Nature Biotechnol.* **2007,** *25*, 338-344.
(16) Armstrong, C. T.; Watkins, D. W.; Anderson, J. L. R. Constructing manmade enzymes for oxygen activation. *Dalton Trans.* **2013,** *42*, 3136-3150.
(17) Giacomo, C.; Sergio, R. Properties and synthetic applications of enzymes in organic solvents. *Angew. Chem. Int. Ed.* **2000,** *39*, 2226-2254.
(18) Klibanov, A. M. Improving enzymes by using them in organic solvents. *Nature* **2001,** *409*, 241-246.
(19) Dürrenberger, M.; Ward, T. R. Recent achievments in the design and engineering of artificial metalloenzymes. *Curr. Opin. Chem. Biol.* **2014,** *19*, 99-106.
(20) Ro, D.-K.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006,** *440*, 940-943.
(21) Kirk, O.; Borchert, T. V.; Fuglsang, C. C. Industrial enzyme applications. *Curr. Opin. Chem. Biol.* **2002,** *13*, 345-351.
(22) Denard, C. A.; Ren, H.; Zhao, H. Improving and repurposing biocatalysts via directed evolution. *Curr. Opin. Chem. Biol.* **2015,** *25*, 55-64.
(23) Chokhawala, H. A.; Roche, C. M.; Kim, T.-W.; Atreya, M. E.; Vegesna, N.; Dana, C. M.; Blanch, H. W.; Clark, D. S. Mutagenesis of Trichoderma reesei endoglucanase I: impact of expression host on activity and stability at elevated temperatures. *BMC Biotechnol.* **2015,** *15*, 11-23.
(24) Hynes, J. T. Chemical reaction dynamics in solution. *Ann. Rev. Phys. Chem.* **1985,** *36*, 573-597.
(25) Tunon, I.; Laage, D.; Hynes, J. T. Are there dynamical effects in enzyme catalysis? Some thoughts concerning the enzymatic chemical step. *Arch. Biochem. Biophys.* **2015,** *582*, 42-55.
(26) Brustad, E. M.; Arnold, F. H. Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol.* **2011,** *15*, 201-210.
(27) Cherry, J. R.; Lamsa, M. H.; Schneider, P.; Vind, J.; Svendsen, A.; Jones, A.; Pedersen, A. H. Directed evolution of a fungal peroxidase. *Nature Biotechnol.* **1999,** *17*, 379-384.
(28) Voigt, C. A.; Mayo, S. L.; Arnold, F. H.; Wang, Z. G. Computational method to reduce the search space for directed protein evolution. *Proc. Natl. Acad. Sci.* **2001,** *98*, 3778-3783.
(29) Wu, S.; Acevedo, J. P.; Reetz, M. T. Induced allostery in the directed evolution of an enantioselective Baeyer-Villiger monooxygenase. *Proc. Natl. Acad. Sci.* **2010,** *107*, 2775-2780.
(30) Renata, H.; Wang, J. Z.; Arnold, F. H. Expanding the enzyme universe: accessing non-natural reactions by mechanism guided directed evolution. *Angew. Chem. Int. Ed.* **2015,** *54*, 3351-3367.
(31) Packer, M. S.; Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **2015,** *16*, 379-394.
(32) Tawfik, D. S. Directed molecular evolution of proteins: Or how to improve enzymes for biocatalysis, edited by Susanne Brackmann and Kay Johnsson. 2002. Weinheim, Germany: Wiley-VCH. 357 pages. *Methods Mol. Bio.* **2004,** *230*, 2836-2837.

(33) Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **1994,** *370*, 389-391.
(34) Castle, L. A.; Siehl, D. L.; Gorton, R.; Patten, P. A.; Chen, Y. H.; Bertain, S.; Cho, H.-J.; Duck, N.; Wong, J.; Liu, D.et al. Discovery and directed evolution of a Glyphosate tolerance gene. *Science* **2004,** *304*, 1151-1154.
(35) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci.* **2012,** *109*, 3790-3795.
(36) Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grütter, M. G.; Privett, H. K.; Mayo, S. L.; Hilvert, D. Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* **2013,** *503*, 418-421.
(37) Saunders, C. T.; Baker, D. Evaluation of structural and evolutionary contributions to deleterious mutation prediction. *J. Mol. Bio.* **2002,** *322*, 891-901.
(38) Fleishman, S. J.; Khare, S. D.; Koga, N.; Baker, D. Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci.* **2011,** *20*, 753-757.
(39) Smith, A. J.; Müller, R.; Toscano, M. D.; Kast, P.; Hellinga, H. W.; Hilvert, D.; Houk, K. N. Structural reorganization and preorganization in enzyme active sites: comparisons of experimental and theoretically ideal active site geometries in the multistep serine esterase reaction cycle. *J. Amer. Chem. Soc.* **2008,** *130*, 15361-15373.
(40) Kaplan, J.; DeGrado, W. F. De novo design of catalytic proteins. *Proc. Natl. Acad. Sci.* **2004,** *101*, 11566-11570.
(41) Korendovych, I. V.; DeGrado, W. F. Catalytic efficiency of designed catalytic proteins. *Curr. Opin. Struct. Bio.* **2014,** *27*, 113-121.
(42) Bolon, D. N.; Mayo, S. L. Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci.* **2001,** *98*, 14274-14279.
(43) Nanda, V.; Koder, R. L. Designing artificial enzymes by intuition and computation. *Nature Chem.* **2010,** *2*, 15-24.
(44) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.et al. De novo computational design of retro-aldol enzymes. *Science* **2008,** *319*, 1387-1391.
(45) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.et al. Kemp elimination catalysts by computational enzyme design. *Nature* **2008,** *453*, 190-198.
(46) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St.Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **2010,** *329*, 309-313.
(47) Rajagopalan, S.; Wang, C.; Yu, K.; Kuzin, A. P.; Richter, F.; Lew, S.; Miklos, A. E.; Matthews, M. L.; Seetharaman, J.; Su, M.et al. Design of activated serine--containing catalytic triads with atomic-level accuracy. *Nature Chem. Bio.***2014,** *10*, 386-393.
(48) Sauer, J.; Freund, H.-J. Models in catalysis. *Catal. Lett.* **2015,** *145*, 109-125.
(49) Hu, Y.; Houk, K. N.; Kikuchi, K.; Hotta, K.; Hilvert, D. Nonspecific medium effects versus specific group positioning in the antibody and Albumin

catalysis of the base-promoted ring-opening reactions of benzisoxazoles. *J. Amer. Chem. Soc.***2004,** *126*, 8197-8205.
(50) Debler, E. W.; Ito, S.; Seebeck, F. P.; Heine, A.; Hilvert, D.; Wilson, I. a. Structural origins of efficient proton abstraction from carbon by a catalytic antibody. *Proc. Natl. Acad. Sci.* **2005,** *102*, 4984-4989.
(51) Fried, S. D.; Boxer, S. G. Electric fields and enzyme catalysis. *Ann. Rev. Biochem.* **2017,** *86*, 387-415.
(52) Xiang, S.; Short, S. A.; Wolfenden, R.; Carter, C. W., Jr. The structure of the cytidine deaminase-product complex provides evidence for efficient proton transfer and ground-state destabilization. *Biochem.* **1997,** *36*, 4768-4774.
(53) Miller, B. G.; Butterfoss, G. L.; Short, S. A.; Wolfenden, R. Role of enzyme−Ribofuranosyl contacts in the ground state and transition state for Orotidine 5‘-Phosphate Decarboxylase: a role for substrate destabilization? *Biochem.* **2001,** *40*, 6227-6232.
(54) Wang, L.; Fried, S. D.; Boxer, S. G.; Markland, T. E. Quantum delocalization of protons in the hydrogen-bond network of an enzyme active site. *Proc. Natl. Acad. Sci.* **2014,** *111*, 18454-18459.
(55) Wang, L.; Fried, S. D.; Markland, T. E. Proton network flexibility enables robustness and large electric fields in the Ketosteroid Isomerase active site. *J. Phys. Chem. B* **2017,** *121*, 9807-9815.
(56) Kohen, A. Role of dynamics in enzyme catalysis: substantial versus semantic controversies. *Acc. Chem. Res.* **2015,** *48*, 466-473.
(57) Warshel, A. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J. Biol. Chem.* **1998,** *273*, 27035-27038.
(58) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H. B.; Olsson, M. H. M. Electrostatic basis for enzyme catalysis. *Chem. Rev.* **2006,** *106*, 3210-3235.
(59) Villà, J.; Štrajbl, M.; Glennon, T. M.; Sham, Y. Y.; Chu, Z. T.; Warshel, A. How important are entropic contributions to enzyme catalysis? *Proc. Natl. Acad. Sci.* **2000,** *97*, 11899-11904.
(60) Klinman, J. *Dynamics in enzyme catalysis*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013.
(61) Nagel, Z. D.; Klinman, J. P. A 21st century revisionist's view at a turning point in enzymology. *Nature Chem. Bio.* **2009,** *5*, 543-550.
(62) Bhowmick, A.; Sharma, S. C.; Honma, H.; Head-Gordon, T. The role of side chain entropy and mutual information for improving the de novo design of Kemp eliminases KE07 and KE70. *Phys. Chem. Chem. Phys.* **2016,** *18*, 19386-19396.
(63) Fraser, J. S.; Clarkson, M. W.; Degnan, S. C.; Erion, R.; Kern, D.; Alber, T. Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **2009,** *462*, 669-673.
(64) Henzler-Wildman, K. a.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **2007,** *450*, 913-916.
(65) Caratzoulas, S.; Mincer, J. S.; Schwartz, S. D. Identification of a protein-promoting vibration in the reaction catalyzed by horse liver alcohol dehydrogenase stavros. *J. Amer. Chem. Soc.* **2002,** *124*, 3270-3276.
(66) Mincer, J. S.; Schwartz, S. D. A computational method to identify residues important in creating a protein promoting vibration in enzymes. *J. Phys. Chem.* **2003,** *107*, 366-371.

(67) Truhlar, D. G.; Garrett, B. C. Variational transition-state theory. *Acc. Chem. Res.* **1980,** *13*, 440-448.
(68) Pu, J.; Gao, J.; Truhlar, D. G. Multidimensional tunneling, recrossing, and the transmission coefficient for enzymatic reactions. *Chem. Rev.***2006,** *106*, 3140-3169.
(69) Peters, B. *Reaction rate theory and rare events simulations*; Elsevier: Amsterdam, 2017.
(70) Agarwal, P. K.; Billeter, S. R.; Rajagopalan, P. T.; Benkovic, S. J.; Hammes-Schiffer, S. Network of coupled promoting motions in enzyme catalysis. *Proc. Natl. Acad. Sci. U S A* **2002,** *99*, 2794-2799.
(71) Hammes-Schiffer, S.; Benkovic, S. J. Relating protein motion to catalysis. *Ann. Rev. Biochem.* **2006,** *75*, 519-541.
(72) Nashine, V. C.; Hammes-Schiffer, S.; Benkovic, S. J. Coupled motions in enzyme catalysis. *Curr. Opin. Chem. Bio.* **2010,** *14*, 644-651.
(73) Wong, K. F.; Selzer, T.; Benkovic, S. J.; Hammes-Schiffer, S. Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase. *Proc. Nat. Acad. Sci.* **2005,** *102(19)*, 6807-6812.
(74) Boekelheide, N.; Salomon-Ferrer, R.; Miller, T. F., 3rd. Dynamics and dissipation in enzyme catalysis. *Proc. Natl. Acad. Sci.* **2011,** *108*, 16159-16163.
(75) Tantillo, D. J.; Chen, J.; Houk, K. N. Theozymes and compuzymes: theoretical models for biological catalysis. *Curr. Opin. Chem. Bio.* **1998,** *2*, 743-750.
(76) Dahiyat, B. I.; Mayo, S. L. Protein design automation. *Protein Sci.* **1996,** *5*, 895-903.
(77) Shirvanyants, D.; Alexandrova, A. N.; Dokholyan, N. V. Rigid substructure search. *Bioinform.* **2011,** *27*, 1327-1329.
(78) Nosrati, G. R.; Houk, K. N. SABER: A computational method for identifying active sites for new reactions. *Protein Sci.* **2012,** *21*, 697-706.
(79) Barker, J. A.; Thornton, J. M. An algorithm for constraint-based structural template matching: application to 3D templates with statistical analysis. *Bioinform.* **2003,** *19*, 1644-1649.
(80) Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Röthlisberger, D.; Baker, D. New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **2006,** *15*, 2785-2794.
(81) Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D. De novo enzyme design using Rosetta3. *PLoS One* **2011,** *6*, e19230.
(82) Kiss, G.; Celebi-Olcum, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational enzyme design. *Angew. Chem. Int. Ed.* **2013,** *52*, 5700-5725.
(83) Świderek, K.; Tuñón, I.; Moliner, V.; Bertran, J. Computational strategies for the design of new enzymatic functions. *Arch. Biochem. Biophys.* **2015,** *582*, 68-79.
(84) Ruscio, J. Z.; Kohn, J. E.; Ball, K. A.; Head-Gordon, T. The influence of protein dynamics on the success of computational enzyme design. *J. Amer. Chem. Soc.* **2009,** *131*, 14111-14115.
(85) Alexandrova, A. N.; Rothlisberger, D.; Baker, D.; Jorgensen, W. L. Catalytic mechanism and performance of computationally designed enzymes for Kemp elimination. *J. Amer. Chem. Soc.* **2008,** *130*, 15907-15915.
(86) Khersonsky, O.; Kiss, G.; Röthlisberger, D.; Dym, O.; Albeck, S.; Houk, K. N.; Baker, D.; Tawfik, D. S. Bridging the gaps in design methodologies by

evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc. Natl. Acad. Sci.* **2012,** *109*, 10358-10363.

(87) Khersonsky, O.; Röthlisberger, D.; Dym, O.; Albeck, S.; Jackson, C. J.; Baker, D.; Tawfik, D. S. Evolutionary optimization of computationally designed enzymes: Kemp Eliminases of the KE07 Series. *J. Mol. Biol.* **2010,** *396*, 1025-1042.

(88) Khersonsky, O.; Röthlisberger, D.; Wollacott, A. M.; Murphy, P.; Dym, O.; Albeck, S.; Kiss, G.; Houk, K. N.; Baker, D.; Tawfik, D. S. Optimization of the in-silico-designed Kemp Eliminase KE70 by computational design and directed evolution. *J. Mol. Biol.* **2011,** *407*, 391-412.

(89) Kries, H.; Blomberg, R.; Hilvert, D. De novo enzymes by computational design. *Curr. Opin. Chem. Biol.* **2013,** *17*, 221-228.

(90) Frushicheva, M. P.; Cao, J.; Chu, Z. T.; Warshel, A. Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proc. Natl. Acad. Sci.* **2010,** *107*, 16869-16874.

(91) Cooper, S.; Khatib, F.; Treuille, A.; Barbero, J.; Lee, J.; Beenen, M.; Leaver-Fay, A.; Baker, D.; Popovic, Z.; Players, F. Predicting protein structures with a multiplayer online game. *Nature* **2010,** *466*, 756-760.

(92) Wang, J.; Korambath, P.; Kim, S.; Johnson, S.; Jin, K.; Crawl, D.; Altintas, I.; Smallen, S.; Labate, B.; Houk, K. N. Theoretical enzyme design using the Kepler scientific workflows on the Grid. *Proc. Comp. Sci.* **2010,** *1*, 1175-1184.

(93) Haddadian, E. J.; Gong, H.; Jha, A. K.; Yang, X.; Debartolo, J.; Hinshaw, J. R.; Rice, P. A.; Sosnick, T. R.; Freed, K. F. Automated real-space refinement of protein structures using a realistic backbone move set. *Biophys. J.* **2011,** *101*, 899-909.

(94) Murphy, G. S.; Mills, J. L.; Miley, M. J.; Machius, M.; Szyperski, T.; Kuhlman, B. Increasing sequence diversity with flexible backbone protein design: the complete redesign of a protein hydrophobic core. *Structure* **2012,** *20*, 1086-1096.

(95) Leaver-Fay, A.; Jacak, R.; Stranges, P. B.; Kuhlman, B. A generic program for multistate protein design. *PLoS One* **2011,** *6*, e20937.

(96) Sterner, R.; Hocker, B. Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold. *Chem. Rev.* **2005,** *105*, 4038-4055.

(97) Blaha-Nelson, D.; Kruger, D. M.; Szeler, K.; Ben-David, M.; Kamerlin, S. C. Active site hydrophobicity and the convergent evolution of paraoxonase activity in structurally divergent enzymes: the case of Serum Paraoxonase 1. *J. Amer. Chem. Soc.* **2017,** *139*, 1155-1167.

(98) Mortier, J.; Rakers, C.; Bermudez, M.; Murgueitio, M. S.; Riniker, S.; Wolber, G. The impact of molecular dynamics on drug design: applications for the characterization of ligand-macromolecule complexes. *Drug Discov. Today* **2015,** *20*, 686-702.

(99) Fuxreiter, M.; Mones, L. The role of reorganization energy in rational enzyme design. *Curr. Opin. Chem. Bio.* **2014,** *21*, 34-41.

(100) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a novel globular protein fold with atomic-level accuracy. *Science* **2003,** *302*, 1364-1368.

(101) Park, S.; Yang, X.; Saven, J. G. Advances in computational protein design. *Curr. Opin. Struct. Bio.* **2004,** *14*, 487-494.

(102) Hur, S.; Bruice, T. C. The near attack conformation approach to the study of the chorismate to prephenate reaction. *Proc. Natl. Acad. Sci.* **2003,** *100*, 12015-12020.
(103) Scheiner, S.; Lipscomb, W. N.; Kleier, D. A. Molecular orbital studies of enzyme activity. 2. Nucleophilic attack on carbonyl systems with comments on orbital steering. *J. Amer. Chem. Soc.***1976,** *98*, 4770-4777.
(104) Vaissier-Welborn, V.; Pestana, L. R.; Head-Gordon, T. Computational optimization of electric fields for better catalysis design. *Nature Catal.* **2018,** *in press*.
(105) Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y. E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. An artificial metalloenzyme with the kinetics of native enzymes. *Science* **2016,** *354*, 102-106.
(106) Lewis, J. C. Artificial metalloenzymes and metallopeptide catalysts for organic synthesis. *ACS Catal.* **2013,** *3*, 2954-2975.
(107) Takafumi, U.; Hiroyasu, T.; Yuya, T. Artificial metalloenzymes constructed from hierarchically-assembled proteins. *Chem. Asian J.* **2013,** *8*, 1646-1660.
(108) Thomson, A. R.; Wood, C. W.; Burton, A. J.; Bartlett, G. J.; Sessions, R. B.; Brady, R. L.; Woolfson, D. N. Computational design of water-soluble alpha-helical barrels. *Science* **2014,** *346*, 485-488.
(109) Bos, J.; Roelfes, G. Artificial metalloenzymes for enantioselective catalysis. *Curr. Opin. Chem. Biol.* **2014,** *19*, 135-143.
(110) Zastrow, M. L.; Pecoraro, V. L. Designing functional metalloproteins: From structural to catalytic metal sites. *Coord. Chem. Rev.* **2013,** *257*, 2565-2588.
(111) Summa, C. M.; Rosenblatt, M. M.; Hong, J. K.; Lear, J. D.; DeGrado, W. F. Computational de novo design, and characterization of an A(2)B(2) diiron protein. *J. Mol. Bio.***2002,** *321*, 923-938.
(112) Marsh, E. N.; DeGrado, W. F. Noncovalent self-assembly of a heterotetrameric diiron protein. *Proc. Natl. Acad. Sci.* **2002,** *99*, 5150-5154.
(113) Lombardi, A.; Summa, C. M.; Geremia, S.; Randaccio, L.; Pavone, V.; DeGrado, W. F. Retrostructural analysis of metalloproteins: Application to the design of a minimal model for diiron proteins. *Proc. Natl. Acad. Sci.***2000,** *97*, 6298-6305.
(114) Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N. Installing hydrolytic activity into a completely de novo protein framework. *Nat. Chem.* **2016,** *8*, 837-844.
(115) Jencks, W. P. Binding energy, specificity, and enzymic catalysis: the Circe effect. **2006**, in Advances in Enzymology, Altom Meister, Ed. Wiley, Ch. 4, 219-410.
(116) Hansen, J. L.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. Structural insights into peptide bond formation. *Proc. Natl. Acad. Sci.* **2002,** *99*, 11670-11675.
(117) Kazemi, M.; Himo, F.; Aqvist, J. Enzyme catalysis by entropy without Circe effect. *Proc. Natl. Acad. Sci.* **2016,** *113*, 2406-2411.
(118) Boehr, D. D.; Nussinov, R.; Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **2009,** *5*, 789-796.
(119) Wand, A. J.; Moorman, V. R.; Harpole, K. W. A surprising role for conformational entropy in protein function. *Top. Curr. Chem.* **2013,** *337*, 69-94.
(120) Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. The role of conformational entropy in molecular recognition by calmodulin. *Nature Chem. Bio.* **2010,** *6*, 352-358.

(121) Wand, A. J. The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Curr. Opin. Struct. Bio.* **2013,** *23*, 75-81.
(122) Fu, Y.; Kasinath, V.; Moorman, V. R.; Nucci, N. V.; Hilser, V. J.; Wand, A. J. Coupled motion in proteins revealed by pressure perturbation. *J. Amer. Chem. Soc.* **2012,** *134*, 8543-8550.
(123) Counago, R.; Wilson, C. J.; Pena, M. I.; Wittung-Stafshede, P.; Shamoo, Y. An adaptive mutation in adenylate kinase that increases organismal fitness is linked to stability-activity trade-offs. *Protein Eng. Des. Sel.* **2008,** *21*, 19-27.
(124) Siddiqui, K. S.; Cavicchioli, R. Cold-adapted enzymes. *Ann. Rev. Biochem.***2006,** *75*, 403-433.
(125) Feller, G. Protein stability and enzyme activity at extreme biological temperatures. *J. Phys. Cond. Matt.***2010,** *22*, 323101.
(126) Siddiqui, K. S. Defying the activity-stability trade-off in enzymes: taking advantage of entropy to enhance activity and thermostability. *Crit. Rev. Biotech.* **2017,** *37*, 309-322.
(127) Siddiqui, K. S.; Williams, T. J.; Wilkins, D.; Yau, S.; Allen, M. A.; Brown, M. V.; Lauro, F. M.; Cavicchioli, R. Psychrophiles. *Ann. Rev. Earth Planetary Sci.* **2013,** *41*, 87-115.
(128) Guerriero, G.; Hausman, J. F.; Strauss, J.; Ertan, H.; Siddiqui, K. S. Destructuring plant biomass: focus on fungal and extremophilic cell wall hydrolases. *Plant Sci.* **2015,** *234*, 180-193.
(129) Siddiqui, K. S.; Poljak, A.; De Francisci, D.; Guerriero, G.; Pilak, O.; Burg, D.; Raftery, M. J.; Parkin, D. M.; Trewhella, J.; Cavicchioli, R. A chemically modified alpha-amylase with a molten-globule state has entropically driven enhanced thermal stability. *Protein Eng. Des. Sel.* **2010,** *23*, 769-780.
(130) Kotting, C.; Kallenbach, A.; Suveyzdis, Y.; Wittinghofer, A.; Gerwert, K. The GAP arginine finger movement into the catalytic site of Ras increases the activation entropy. *Proc. Natl. Acad. Sci.* **2008,** *105*, 6260-6265.
(131) Belsare, S.; Pattni, V.; Heyden, M.; Head-Gordon, T. Solvent entropy contributions to catalytic activity in designed and optimized Kemp Eliminases. *J. Phys. Chem. B* **2018,** *122*, 5300-5307.
(132) Frushicheva, M. P.; Cao, J.; Warshel, A. Challenges and advances in validating enzyme design proposals: the case of kemp eliminase catalysis. *Biochem.* **2011,** *50*, 3849-3858.
(133) Bhabha, G.; Biel, J. T.; Fraser, J. S. Keep on moving: discovering and perturbing the conformational dynamics of enzymes. *Acc. Chem. Res.* **2015,** *48*, 423-430.
(134) Welborn, V. V.; Van Voorhis, T. Non-radiative deactivation of cytosine derivatives at elevated temperature. *Mol. Phys.* **2018**, DOI:10.1080/00268976.2018.1457806 10.1080/00268976.2018.1457806, 1-8.
(135) Basner, J. E.; Schwartz, S. D. How enzyme dynamics helps catalyze a reaction in atomic detail: a transition path sampling study. *J. Amer. Chem. Soc.* **2005,** *127*, 13822-13831.
(136) Antoniou, D.; Schwartz, S. D. Protein dynamics and enzymatic chemical barrier passage. *J. Phys. Chem. B* **2011,** *115*, 15147-15158.
(137) Pan, X.; Schwartz, S. D. Free energy surface of the Michaelis complex of lactate dehydrogenase: a network analysis of microsecond simulations. *J. Phys. Chem. B* **2015,** *119*, 5430-5436.

(138) Varga, M. J.; Dzierlenga, M. W.; Schwartz, S. D. Structurally linked dynamics in Lactate Dehydrogenases of evolutionarily distinct species. *Biochem.* **2017,** *56*, 2488-2496.
(139) Zoi, I.; Antoniou, D.; Schwartz, S. D. Incorporating fast protein dynamics into enzyme design: a proposed mutant aromatic Amine Dehydrogenase. *J. Phys. Chem. B* **2017,** *121*, 7290-7298.
(140) Chen, X.; Schwartz, S. D. Directed evolution as a probe of rate promoting vibrations introduced via mutational change. *Biochem.* **2018,** *57*, 3289-3298.
(141) Schramm, V. L.; Schwartz, S. D. Promoting vibrations and the function of enzymes. Emerging theoretical and experimental convergence. *Biochem.* **2018**, DOI:10.1021/acs.biochem.8b00201 10.1021/acs.biochem.8b00201.
(142) Klinman, J. P. The role of tunneling in enzyme catalysis of C-H activation. *Biochim. Biophys. Acta* **2006,** *1757*, 981-987.
(143) Knapp, M. J.; Klinman, J. P. Environmentally coupled hydrogen tunneling. *Eur. J. Biochem.* **2002,** *269*, 3113-3121.
(144) Marcus, R. A.; Sutin, N. Electron transfers in chemistry and biology. *Biochim. Biophys. Acta* **1985,** *811*, 265-322.
(145) Marcus, R. A. Chemical and electrochemical electron-transfer theory. *Ann. Rev. Phys. Chem.* **1964,** *15*, 155-196.
(146) Vaissier, V.; Barnes, P.; Kirkpatrick, J.; Nelson, J. Influence of polar medium on the reorganization energy of charge transfer between dyes in a dye sensitized film. *Phys. Chem. Chem. Phys.* **2013,** *15*, 4804-4814.
(147) Aikawa, K.; Sumita, M.; Shimodo, Y.; Morihashi, K. Theoretical studies of molecular orientation and charge recombination in poly-paraphenylenevinylene light-emitting diodes. *Phys. Chem. Chem. Phys.* **2015,** *17*, 20923-20931.
(148) Alvarez-Paggi, D.; Castro, M. A.; Tortora, V.; Castro, L.; Radi, R.; Murgida, D. H. Electrostatically driven second-sphere ligand switch between high and low reorganization energy forms of native cytochrome c. *J. Amer. Chem. Soc.* **2013,** *135*, 4389-4397.
(149) Galinato, M. G.; Kleingardner, J. G.; Bowman, S. E.; Alp, E. E.; Zhao, J.; Bren, K. L.; Lehnert, N. Heme-protein vibrational couplings in cytochrome c provide a dynamic link that connects the heme-iron and the protein surface. *Proc. Natl. Acad. Sci.* **2012,** *109*, 8896-8900.
(150) Vaissier, V.; Frost, J. M.; Barnes, P. R. F.; Nelson, J. Influence of intermolecular interactions on the reorganization energy of charge transfer between surface-attached dye molecules. *J. Phys. Chem. C* **2015,** *119*, 24337-24341.
(151) Maggio, E.; Martsinovich, N.; Troisi, A. Evaluating charge recombination rate in dye-sensitized solar cells from electronic structure calculations. *J. Phys. Chem. C* **2012,** *116*, 7638-7649.
(152) Åqvist, J.; Warshel, A. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem. Rev.* **1993,** *93*, 2523-2544.
(153) Fuxreiter, M.; Warshel, A. Origin of the catalytic power of Acetylcholinesterase: computer simulation studies. *J. Amer. Chem. Soc.***1998,** *120*, 183-194.
(154) Yadav, A.; Jackson, R. M.; Holbrook, J. J.; Warshel, A. Role of solvent reorganization energies in the catalytic activity of enzymes. *J. Amer. Chem. Soc.* **1991,** *113*, 4800-4805.

(155) Labas, A.; Szabo, E.; Mones, L.; Fuxreiter, M. Optimization of reorganization energy drives evolution of the designed Kemp eliminase KE07. *Biochim. Biophys. Acta* **2013,** *1834*, 908-917.
(156) Zoi, I.; Antoniou, D.; Schwartz, S. D. Electric fields and fast protein dynamics in enzymes. *J. Phys. Chem. Lett.* **2017,** *8*, 6165-6170.
(157) Lehle, H.; Kriegl, J. M.; Nienhaus, K.; Deng, P.; Fengler, S.; Nienhaus, G. U. Probing electric fields in protein cavities by using the vibrational stark effect of carbon monoxide. *Biophys. J.* **2005,** *88*, 1978-1990.
(158) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. Electric fields at the active site of an enzyme: direct comparison of experiment with theory. *Science* **2006,** *313*, 200-204.
(159) Webb, L. J.; Boxer, S. G. Electrostatic fields near the active site of human aldose reductase: 1. New inhibitors and vibrational stark effect measurements. *Biochem.* **2008,** *47*, 1588-1598.
(160) Fried, S. D.; Bagchi, S.; Boxer, S. G. Extreme electric fields power catalysis in the active site of ketosteroid isomerase. *Science* **2014,** *346*, 1510-1514.
(161) Park, E. S.; Andrews, S. S.; Hu, R. B.; Boxer, S. G. Vibrational Stark spectroscopy in proteins: a probe and calibration for electrostatic fields. *J. Phys. Chem. B* **1999,** *103*, 9813-9817.
(162) Fried, S. D.; Boxer, S. G. Measuring electric fields and noncovalent interactions using the vibrational stark effect. *Acc. Chem. Res.* **2015,** *48*, 998-1006.
(163) Fried, S. D.; Bagchi, S.; Boxer, S. G. Measuring electrostatic fields in both hydrogen-bonding and non-hydrogen-bonding environments using carbonyl vibrational probes. *J. Amer. Chem. Soc.* **2013,** *135*, 11181-11192.
(164) Suydam, I. T.; Boxer, S. G. Vibrational Stark effects calibrate the sensitivity of vibrational probes for electric fields in proteins. *Biochem.* **2003,** *42*, 12050-12055.
(165) Andrews, S. S.; Boxer, S. G. Vibrational stark effects of nitriles I. Methods and experimental results. *J. Phys. Chem. A* **2000,** *104*, 11853-11863.
(166) Levinson, N. M.; Fried, S. D.; Boxer, S. G. Solvent-induced infrared frequency shifts in aromatic nitriles are quantitatively described by the vibrational Stark effect. *J. Phys. Chem. B* **2012,** *116*, 10470-10476.
(167) Bagchi, S.; Fried, S. D.; Boxer, S. G. A solvatochromic model calibrates nitriles' vibrational frequencies to electrostatic fields. *J. Amer. Chem. Soc.* **2012,** *134*, 10373-10376.
(168) Chen, D.; Savidge, T. Comment on “Extreme electric fields power catalysis in the active site of ketosteroid isomerase”. *Science* **2015,** *349*, 936.
(169) Fried, S. D.; Boxer, S. G. Response to Comments on “Extreme electric fields power catalysis in the active site of ketosteroid isomerase”. *Science* **2015,** *349*, 936.
(170) Natarajan, A.; Yabukarski, F.; Lamba, V.; Schwans, J. P.; Sunden, F.; Herschlag, D. Comment on “Extreme electric fields power catalysis in the active site of ketosteroid isomerase”. *Science* **2015,** *349*, 936.
(171) Warshel, A. Energetics of enzyme catalysis. *Proc. Natl. Acad. Sci.***1978,** *75*, 5250-5254.
(172) Warshel, A.; Weiss, R. M. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Amer. Chem. Soc.***1980,** *102*, 6218-6226.

(173) Bhowmick, A.; Sharma, S. C.; Head-Gordon, T. The importance of the scaffold for de novo enzymes: a case study with Kemp Eliminase. *J. Amer. Chem. Soc.* **2017,** *139*, 5793-5800.
(174) Liu, C. T.; Layfield, J. P.; Stewart, R. J.; French, J. B.; Hanoian, P.; Asbury, J. B.; Hammes-Schiffer, S.; Benkovic, S. J. Probing the Electrostatics of Active Site Microenvironments along the Catalytic Cycle for Escherichia coli Dihydrofolate Reductase. *J. Amer. Chem. Soc.* **2014,** *136*, 10349-10360.
(175) Vaissier, V.; Sharma, S. C.; Schaettle, K.; Zhang, T.; Head-Gordon, T. Computational optimization of electric fields for improving catalysis of a designed Kemp Eliminase. *ACS Catal.* **2017,** *8*, 219-227.
(176) Beker, W.; van der Kamp, M. W.; Mulholland, A. J.; Sokalski, W. A. Rapid Estimation of Catalytic Efficiency by Cumulative Atomic Multipole Moments: Application to Ketosteroid Isomerase Mutants. *J. Chem. Theory Comput.* **2017,** *13*, 945-955.
(177) Chojnacka, M.; Feliks, M.; Beker, W.; Sokalski, W. A. Predicting substituent effects on activation energy changes by static catalytic fields. *J. Mol. Model.* **2017,** *24*, 28.
(178) Szefczyk, B.; Mulholland, A. J.; Ranaghan, K. E.; Sokalski, W. A. Differential transition-state stabilization in enzyme catalysis: quantum chemical analysis of interactions in the chorismate mutase reaction and prediction of the optimal catalytic field. *J. Amer. Chem. Soc.* **2004,** *126*, 16148-16159.
(179) Aragonès, A. C.; Haworth, N. L.; Darwish, N.; Ciampi, S.; Bloomfield, N. J.; Wallace, G. G.; Diez-Perez, I.; Coote, M. L. Electrostatic catalysis of a Diels–Alder reaction. *Nature* **2016,** *531*, 88.
(180) Zhang, L.; Laborda, E.; Darwish, N.; Noble, B. B.; Tyrell, J. H.; Pluczyk, S.; Le Brun, A. P.; Wallace, G. G.; Gonzalez, J.; Coote, M. L.et al. Electrochemical and Electrostatic Cleavage of Alkoxyamines. *J. Amer. Chem. Soc.* **2018,** *140*, 766-774.
(181) Ciampi, S.; Darwish, N.; Aitken, H. M.; Díez-Pérez, I.; Coote, M. L. Harnessing electrostatic catalysis in single molecule, electrochemical and chemical systems: a rapidly growing experimental tool box. *Chem. Soc. Rev.* **2018,** *47*, 5146-5164.

6. Biographies

Valerie Vaissier Welborn received her Ph.D. in Physics in 2014 from the EPSRC Center for Doctoral Training in Theory and Simulation of Materials at Imperial College London. She worked as a postdoctoral fellow with Troy Van Voorhis at Massachusetts Institute of Technology from 2015-2016. In January 2017, she joined the group of Teresa Head-Gordon as a postdoctoral fellow at the University of California, Berkeley and Lawrence Berkeley National Laboratory. Her work focuses on quantum chemical simulations of complex condensed phase systems with wide range of applications including energy, functional materials and catalysis.

Teresa Head-Gordon received her Ph.D. in Theoretical Chemistry from Carnegie Mellon University in 1989. She was a Postdoctoral Member of Technical Staff at AT&T Bell Laboratories from 1990-1992. She currently is Chancellor's Professor in the Departments of Chemistry, Bioengineering, and Chemical and Biomolecular Engineering at the University of California, Berkeley and Faculty Scientist at Lawrence Berkeley National Laboratory. Her research program has advanced significant developments in theoretical models and statistical mechanical methodology applied to molecular liquids, proteins and protein assemblies, polymers, general interfacial systems, and catalysis.