ACCOUNTS of chemical research

Article

pubs.acs.org/accounts

Molecular Dynamics Explorations of Active Site Structure in Designed and Evolved Enzymes

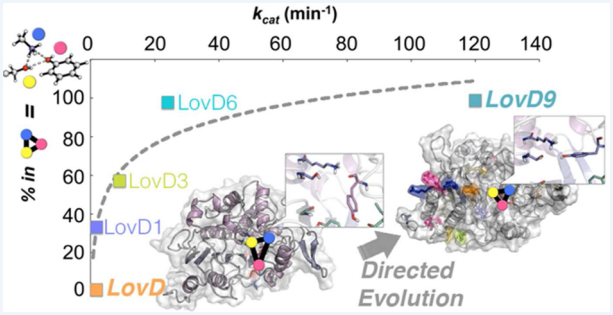
Published as part of the Accounts of Chemical Research special issue "Protein Motion in Catalysis".

Sílvia Osuna,[†,‡] Gonzalo Jiménez-Osés,[‡] Elizabeth L. Noey,[‡] and K. N. Houk*[,‡]

[†]Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, Campus Montilivi s/n, 17071 Girona, Spain

[‡]Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, United States

CONSPECTUS: This Account describes the use of molecular dynamics (MD) simulations to reveal how mutations alter the structure and organization of enzyme active sites. As proposed by Pauling about 70 years ago and elaborated by many others since then, biocatalysis is efficient when functional groups in the active site of an enzyme are in optimal positions for transition state stabilization. Changes in mechanism and covalent interactions are often critical parts of enzyme catalysis. We describe our explorations of the dynamical preorganization of active sites using MD, studying the fluctuations between active and inactive conformations normally concealed to static crystallography. MD shows how the various arrangements of active site residues influence the free energy of the transition state and relates the populations of the catalytic conformational ensemble to the enzyme activity. This Account is organized around three case studies from our laboratory. We first describe the importance of dynamics in evaluating a series of computationally designed and experimentally evolved enzymes for the Kemp elimination, a popular subject in the enzyme design field. We find that the dynamics of the active site is influenced not only by the original sequence design and subsequent mutations but also by the nature of the ligand present in the active site. In the second example, we show how microsecond MD has been used to uncover the role of remote mutations in the active site dynamics and catalysis of a transesterase, LovD. This enzyme was evolved by Tang at UCLA and Codexis, Inc., and is a useful commercial catalyst for the production of the drug simvastatin. X-ray analysis of inactive and active mutants did not reveal differences in the active sites, but relatively long time scale MD in solution showed that the active site of the wild-type enzyme preorganizes only upon binding of the acyl carrier protein (ACP) that delivers the natural acyl group to the active site. In the absence of bound ACP, a noncatalytic arrangement of the catalytic triad is dominant. Unnatural truncated substrates are inactive because of the lack of protein−protein interactions provided by the ACP. Directed evolution is able to gradually restore the catalytic organization of the active site by motion of the protein backbone that alters the active site geometry. In the third case, we demonstrate the key role of MD in combination with crystallography to identify the origins of substrate-dependent stereoselectivities in a number of Codexis-engineered ketoreductases, one of which is used commercially for the production of the antibiotic sulopenem. Here, mutations alter the shape of the active site as well as the accessibility of water to different regions of it. Each of these examples reveals something different about how mutations can influence enzyme activity and shows that directed evolution, like natural evolution, can increase catalytic activity in a variety of remarkable and often subtle ways.



1. INTRODUCTION

Understanding the enormous catalytic power of enzymes is a grand challenge for chemical biology. In the past decade, our group has collaborated with biologists and computational groups to design new enzymes with functions different from those evolved by natural enzymes.[1] Our successes in enzyme design showed that we understand which catalytic groups will accelerate the rates of reactions and that quantum-mechanical (QM) calculations predict the correct positioning of these groups.[2] We also have the computational tools, such as Baker's Rosetta or Mayo's Phoenix, to incorporate these groups into well-characterized stable protein scaffolds.[3]

In spite of these successes, our understanding is not sufficient to rationally tune the global architecture of the protein that ultimately determines catalysis in order to achieve efficiencies rivaling those of naturally evolved proteins. Instead, experimental directed evolution (DE) is often used to achieve several orders of magnitude of acceleration beyond that achieved by the designed protein.[4] Furthermore, many designs fail altogether. Even in successful cases, most of the designs are not active. We need to

Received: December 19, 2014
Published: March 4, 2015

DOI: 10.1021/ar500452q
Acc. Chem. Res. 2015, 48, 1080−1089

develop additional computational tools that provide significant acceleration and selectivity.

The goal of our recent research is to learn how to substitute, at least in part, computational design for these costly DE experiments. To do this, we first need to know much more about how mutations that do not modify the catalytic groups nevertheless influence the catalytic activity to a large extent.[5−8] Mutagenesis can alter not only the shape but also the flexibility of the active site. These dynamic consequences have important effects on catalysis that are often concealed to structural techniques like X-ray crystallography.

All molecules possess potential and kinetic energy and thus are in constant motion, each atom moving as much as 1 Å in 10^{-14} s. MD uses classical Newtonian mechanics to determine how all of the atoms in a system move as a result of forces acting on the atoms. Since there are thousands of atoms in a protein and the associated solvent molecules and ions, parametrized force fields such as AMBER[9] are used to efficiently calculate the forces on each atom that determine their movement. The forces change every time the atoms move, and they must be recalculated after very brief time steps (less than 5 fs). Millions or even billions of calculations are necessary to describe motions happening at short times such as nanoseconds or microseconds. These studies have been made more efficient by technological breakthroughs such us graphics processing unit (GPU)-based computing[10] and special-purpose machines such as the Anton Supercomputer,[11] both of which were used in the work described herein.

In chemistry, dynamics describes the motions of atoms, molecules, and even condensed media.[12] These dynamical fluctuations take place on a wide range of time scales (Figure 1).[13,14] Many time scales like this have been published,[13,14] and

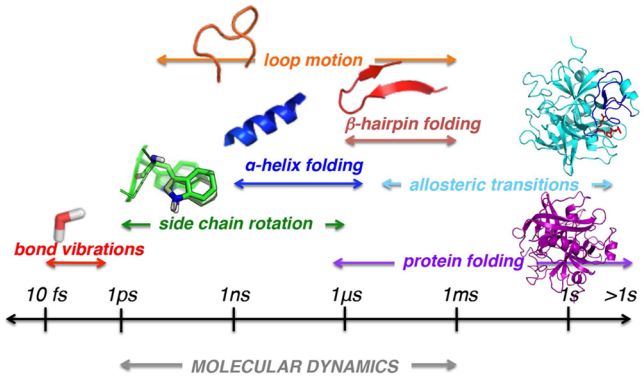


Figure 1. Time scales of different types of protein motions.

we show our own version here. These motions include bond vibrations (10−100 fs), side-chain rotations (ps to μs), local domain fluctuations (ns to ms), allosteric transitions (μs to s), and the overall folding of the structural motifs of proteins (μs to s). Each of these motions either precedes or accompanies chemical reactions.

Chemical dynamics involves studies of rates and mechanisms of reactions at the molecular level, generally in a time-resolved fashion.[12] Molecular dynamicists, including our group, have studied the dynamics of simple reactions in the gas phase and in solution.[12,15,16] The time scales of processes involved in chemical reactions are represented in Figure 2. The time it takes to make or break bonds in a reaction (50−100 fs) is the same as that of most molecular vibrations. The quantity k_BT/h (~50 fs) is the pre-exponential factor in Eyring's transition state theory equation[17] and represents the highest possible rate at which a reaction with no free energy barrier can occur. Other quantities of interest for catalytic reactions are the time for association of substrates to catalysts and the dissociation of catalyst−product complexes. Diffusion limits how fast the association of substrate and enzyme can occur in a condensed phase. Diffusion in water occurs with a bimolecular rate constant of ~10^9 M^{-1} s^{-1}. This translates to association periods of 1 s to 10^{-3} s when the enzyme concentration is micromolar and the substrate concentration is millimolar to 1 M.[18] Other time scales emphasizing uncatalyzed reactions have been published.[19] Dissociation usually occurs in 10^{-5} to 10^{-2} s. The overall time scale of enzyme-catalyzed reactions is quite broad (1 to 10^{-8} s on the basis of measured k_{cat} values).

As discussed in many Accounts in this Special Issue, the time scale for bond breaking and formation (tens of fs) is much shorter than the time scales of collisions with the protein or solvent or of intramolecular vibrational relaxation. Vibrations of the substrate and the active site are indeed necessary to acquire the energy necessary to surmount a potential energy barrier,[12] but this vibrational energy transfer occurs much more slowly than the bond changing events.[14,19−21] Nevertheless, preorganization of active sites is crucial for catalysis, and dynamic motions determine the probability that the active site will be appropriately positioned to stabilize the bond-changing events in reactions.[22−25] These aspects of catalysis and dynamics are the focus of this Account. In addition, protein conformational changes[26] and motions involving mobile loops can act as gates for substrate access or product release[27] from the active site, but we have not studied these.

Much of this Account involves MD simulations aimed at understanding how remote (or at least noncatalytic) mutations influence catalysis. DE involves multiple rounds of evolution followed by screening to accumulate beneficial mutations, which are often scattered around the whole enzyme.[4] Important contributions to this field have been reported recently by Codexis, Inc.[28,29] and the Hilvert,[30] Mayo,[31] Tawfik,[32,33] and Arnold[34−36] laboratories. One of the goals of our research is to understand how mutations of amino acids not directly involved in catalysis nevertheless influence the activity and to build this knowledge into the enzyme design procedure.

2. MD EVALUATION OF DESIGNED ENZYMES FOR THE KEMP ELIMINATION

The "inside-out" enzyme design protocol[1] that our group developed with the Baker lab involves testing the stability of the protein active sites designed through our quantum-mechanical *theozyme*, match, and Rosetta design strategy and their ability to bind the substrate. These tests involve short MD simulations to screen out poor designs. This was used extensively to evaluate Kemp eliminases[7] that were designed by Rosetta.[7,31,33] The Kemp elimination is shown in Figure 3a. MD simulations revealed that unsuccessful computational designs failed to maintain key catalytic hydrogen bonds (Figure 3b). Besides filtering out inactive designs, the time-dependent behavior of the catalytic active site residues was used to guide directed evolution.[31,32]

In a subsequent study,[6] we performed longer MD simulations (100 ns) using ligands resembling the transition state (TS) to different extents (Figure 4a,b), and a broader range of protein folds (Figure 4c−e). The modestly active enzymes KE70 and KE59 were designed through QM calculations in our group and Rosetta in Baker's lab[37] and then evolved experimentally by

DOI: 10.1021/ar500452q
Acc. Chem. Res. 2015, 48, 1080−1089

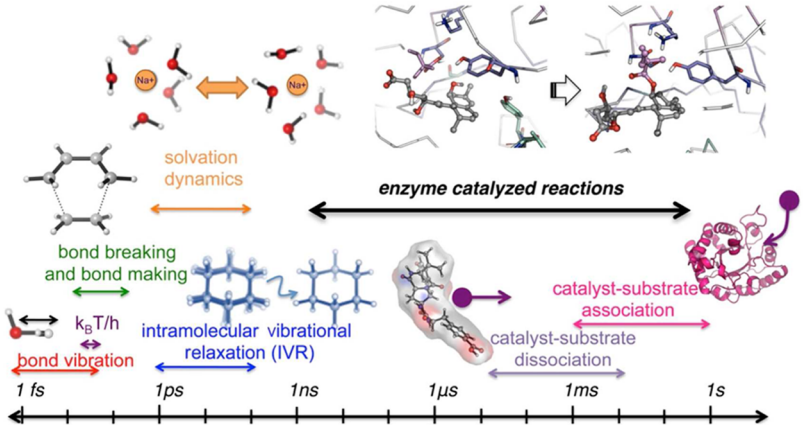


Figure 2. Time scales of events related to reactions.

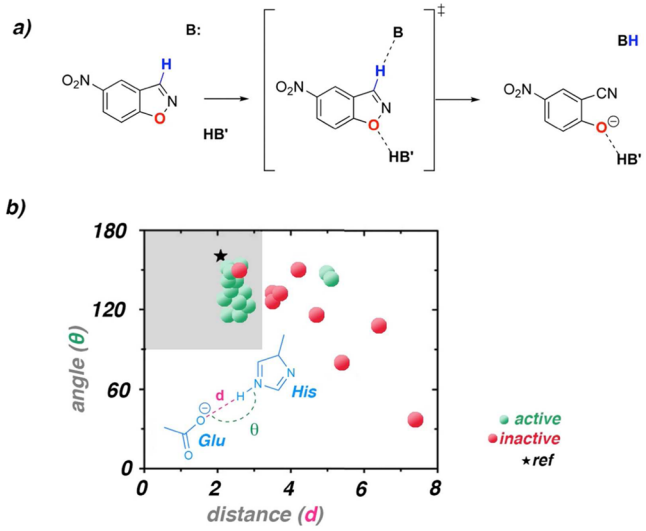


Figure 3. (a) The Kemp elimination. (b) Active site preorganization analyzed for all active and inactive variants by monitoring the H-bond distance (d) and angle (θ) of the catalytic His-Glu dyad along 20 ns MD trajectories. Green circles denote active designs and red circles inactive ones. The star indicates the *theozyme* values of d and θ.

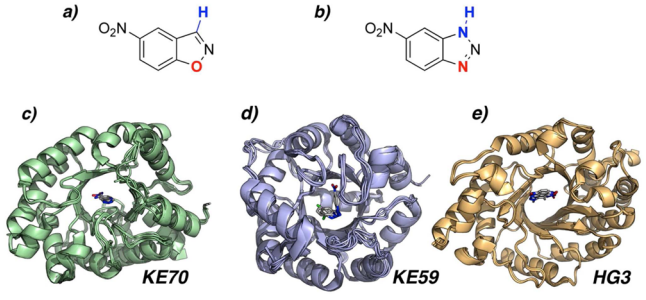


Figure 4. Structures of (a) 5-nitrobenzisoxazole (**1**), the experimental substrate; (b) 6-nitrobenzotriazole (**2**), a frequently cocrystallized inhibitor; and (c–e) the three different Kemp eliminases studied through MD simulations: (c) KE70, (d) KE59, and (e) HG3.

Tawfik.[32,33] These designs feature a His-Asp dyad (KE59) or Glu (KE70) as the general base ("B" in Figure 3a) and Ser as a hydrogen-bond donor ("HB" in Figure 3a; also see Figures 5a and 6a). The available X-ray structures of KE70 variants bound to different substrates and TS analogues indicated a broad distribution of catalytic contacts and often an unproductive

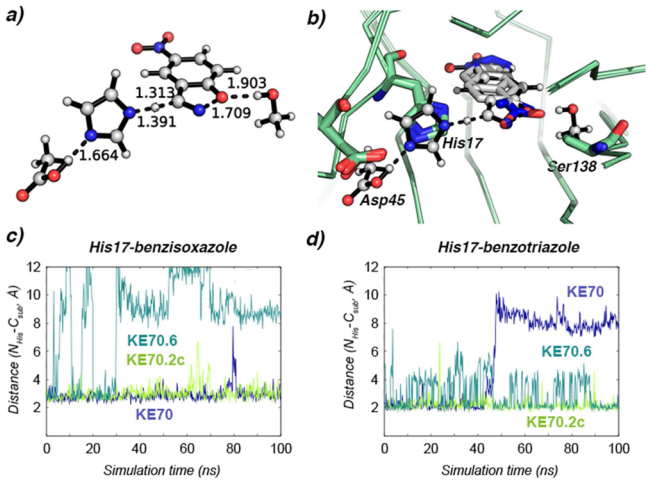


Figure 5. (a) DFT *theozyme* optimized for KE70 enzymes.[33] (b) Overlay of the active site structures of the crystallized KE70.5 (fifth round of DE) and KE70.6 (sixth round) mutants (in sticks) and the *theozyme* structure (in balls and sticks). (c, d) MD traces of the distances between the reactive N atom of His17 and (c) the C atom of benzisoxazole **1** and (d) the N atom of benzotriazole **2** for selected KE70 mutants.

disposition of the substrate with respect to the catalytic residues (Figure 5b). After DE, both enzymes experienced similar improvements of around 400-fold in catalytic efficiency (k_{cat}/K_M).

The more proficient enzyme HG3 was first designed by our group and Mayo[31] and was subsequently evolved experimentally by Hilvert, who achieved a 181-fold increase in catalytic activity.[38] It features an Asp and originally a Lys (afterward mutated to His and finally to Gln) as the general base and the hydrogen-bond donor, respectively. A very preorganized active site close to the original computational design was characterized crystallographically in the final mutant.

The less active, unevolved enzymes KE70 and KE59 showed adequate active site–ligand contacts with the actual substrate, benzisoxazole **1**, in the MD simulations (Figures 5c and 6c); surprisingly, the evolved, most active mutants exhibited worse substrate binding due to migration to either the water solution or secondary binding sites not observed in the crystallographic structures. However, with the more polarized, TS-like benzotriazole **2** as a substrate, the trend was reversed (Figures 5d and 6d). Here, after the second round of DE, the ligand maintained a proper orientation in the active site of KE70, and the catalytic

DOI: 10.1021/ar500452q
Acc. Chem. Res. 2015, 48, 1080–1089

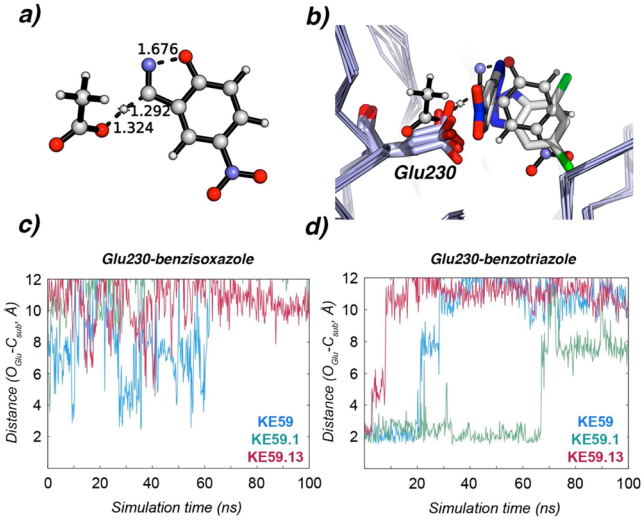


Figure 6. (a) DFT *theozyme* optimized for KE59 enzymes.[32] (b) Overlay of the active site structures of the crystallized KE59.1 (first round of DE) and KE59.13 (13th round) mutants (in sticks) and the *theozyme* structure (in balls and sticks). (c, d) MD traces of the distances between the reactive O atom of Glu230 and (c) the C atom of benzoisoxazole **1** and (d) the N atom of benzotriazole **2** for selected KE59 mutants.

contacts overall got closer to the optimal calculated values (Figure 5d).

With HG3 enzymes, simulations with the reacting substrate **1** were again not consistent with the *theozyme*-fidelity paradigm (Figure 7a): the two most active mutants, HG3.14 and HG3.17,

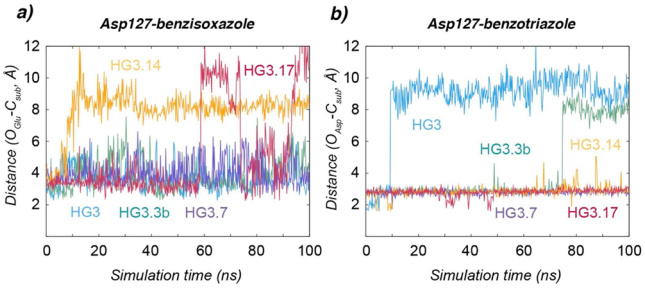


Figure 7. Evolution of the distances between the reactive O atom of Asp127 and (a) the C atom of benzoisoxazole **1** and (b) the N atom of benzotriazole **2** along the MD simulations for all of the HG3 mutants: the original design (HG3, in blue), two intermediate DE mutants (the third HG3.3b and seventh HG3.7, in teal and violet, respectively), the 14th-DE-round mutant (HG3.14, in orange), and finally the most evolved variant (HG3.17, in magenta).

showed worse substrate binding than the less evolved variants, which displayed a more ordered arrangement of the substrate and the catalytic residues. Again, when the more polarized, cocrystallized benzotriazole **2** was used, the most active mutants maintained the best catalytic contacts (Figure 7b): the catalytic distance between the substrate N−H bond (a polarized mimic of the reactive C−H bond) and Asp127 was very distorted (~10 Å) in the first variant as a result of substrate translocation but was restored along the DE pathway from HG3 to HG3.17 by decreasing the population of the unproductive binding modes. In the seventh round of DE, the catalytic distance stayed constant at the optimal value observed in the X-ray structure of the final mutant and in the computed *theozyme* (2.5 Å; see Figure 8a,b). These results agree with Hilvert and Mayo's definition of *precision*[38] and Warshel's preorganization,[22,23] which is recognized as a requirement for enhanced catalysis.

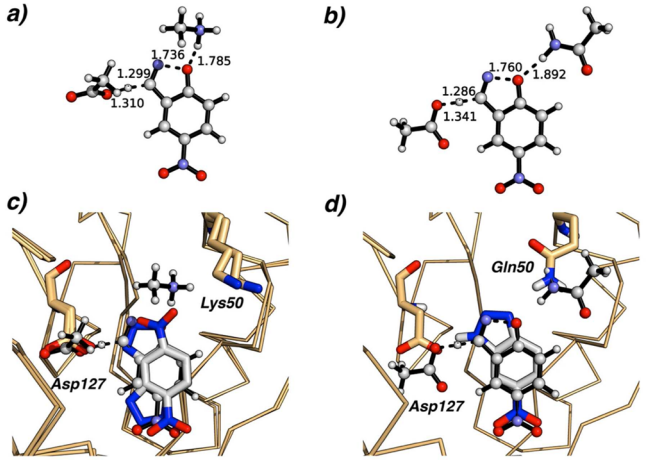


Figure 8. (a, b) DFT *theozymes* with Asp as the base and (a) Lys or (b) Gln as hydrogen-bond donors, optimized for HG3 enzymes. (c, d) Overlays of the *theozyme* structure (in balls and sticks) with the X-ray structures (in sticks) of the (c) computationally designed HG1 and HG2 enzymes[31] and (d) the final evolved HG3.17 variant.[38]

The MD trajectories performed for multiple designs and evolved variants of enzymes for the Kemp elimination indicated that the geometric similarity to the computed *theozymes* could only be achieved in the ground state using *polarized* substrates that mimic the charge distribution in the TS. These observations are fully consistent with the fact that the designs are made to provide maximum stabilization of the polar transition state and not the nonpolar reactant. MD shows very clearly that the active site is dynamic and assembles very tightly around the polar transition state but not around the reactant. The systematic study of different protein scaffolds and substrates by MD simulations has determined how the average structure of active sites can change significantly upon binding of the substrate or TS with respect to the apo state; these conformational changes usually take tens or hundreds of nanoseconds and are very often undetectable by crystallography. Evolution gradually populates the catalytically competent arrangements of inherently flexible active sites. In view of these results, we propose that the systematic evaluation of the conformational flexibility/rigidity of designed active sites through MD simulations, using polarized TS models, will aid in the design of more preorganized and active enzymes.

3. MD ANALYSIS OF THE ROLE OF REMOTE MUTATIONS IN LOVD ACTIVE SITE DYNAMICS

LovD from *Aspergillus terreus* converts monacolin J acid (MJA) into the cholesterol-lowering drug lovastatin (LVA, acid form) via acylation of the *α*-*S*-methylbutyrate side chain (Figure 9). In the natural pathway, LovD interacts with the acyl carrier protein (ACP) domain of its binding partner protein LovF. In this process, LovF acylates Ser76 of LovD following a ping-pong mechanism[39] to deliver the *α*-*S*-methylbutyrate side chain. The latter is then transferred to MJA (Figure 9). The acylation−deacylation reactions are assisted by the catalytic residues Tyr188 and Lys79 (Figures 10 and 11a). Simvastatin (SVA, acid form), the active pharmaceutical ingredient in Zocor, differs from LVA by one methyl group.[40] However, the wild-type LovD enzyme

DOI: 10.1021/ar500452q
Acc. Chem. Res. 2015, 48, 1080−1089

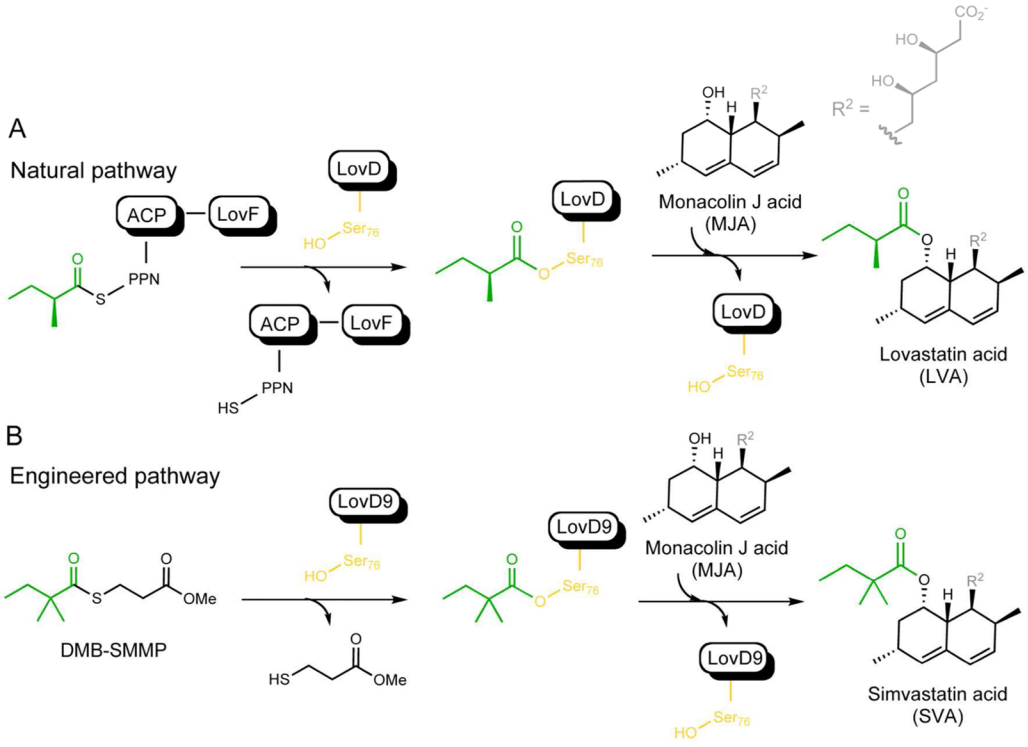


Figure 9. (A) Natural and (B) engineered biosynthetic pathways of LovD.

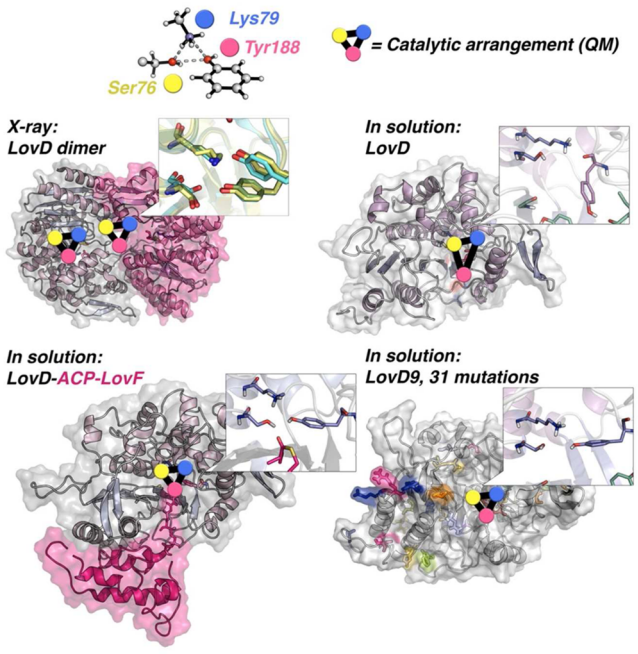


Figure 10. X-ray structure of wild-type LovD (dimer) and representative snapshots from the MD simulations of LovD monomer, the LovD−ACP complex, and the final mutant LovD9 in solution. The insets show zoomed-in views of the active site, and the catalytically (in)competent arrangement(s) of the Ser76-Tyr188-Lys79 triad are represented with cartoon triangles.

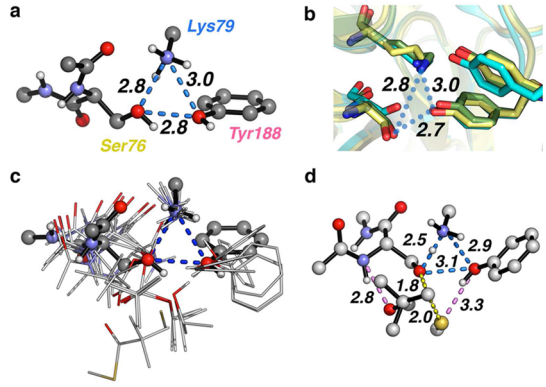


Figure 11. (a) DFT-optimized catalytic Ser76-Lys79-Tyr188 triad. (b) Overlay of the X-ray structures of wild-type LovD (yellow), LovD6 (cyan), and LovD9 (green). (c, d) DFT structures of (c) all of the stationary points along the transacylation reaction pathway and (d) the rate-determining TS for Ser76 acylation.

exhibits poor activity toward non-natural truncated acyl donors. Nine rounds of DE generated LovD9, which now accepts a small free acyl thioester without requiring the ACP domain of LovF.[5,40] LovD9 is 1000-fold more efficient than wild-type LovD in the synthesis of simvastatin and presents 31 beneficial mutations, most of them located outside the active site.[5] The X-ray structures of several mutants with different activities toward SVA showed an almost identical catalytically competent arrangement of the active site, especially for the catalytic Ser76-Tyr188-Lys79 residues (Figures 10 and 11b). Similarly, <100 ns MD simulations did not provide an explanation of the observed activities. In contrast, microsecond MD simulations performed in ANTON[11] provided remarkable insights for the improvement of the catalytic proficiency of the laboratory-generated variants along the directed evolution pathway.

The ∼1.5 μs MD simulations performed on the apo monomeric structures showed that along the DE pathway, the catalytically competent arrangement of residues Ser76, Tyr188, and Lys79 predicted by QM calculations (Figure 11a) was progressively stabilized (Figure 12). This ideal arrangement

DOI: 10.1021/ar500452q
Acc. Chem. Res. 2015, 48, 1080−1089

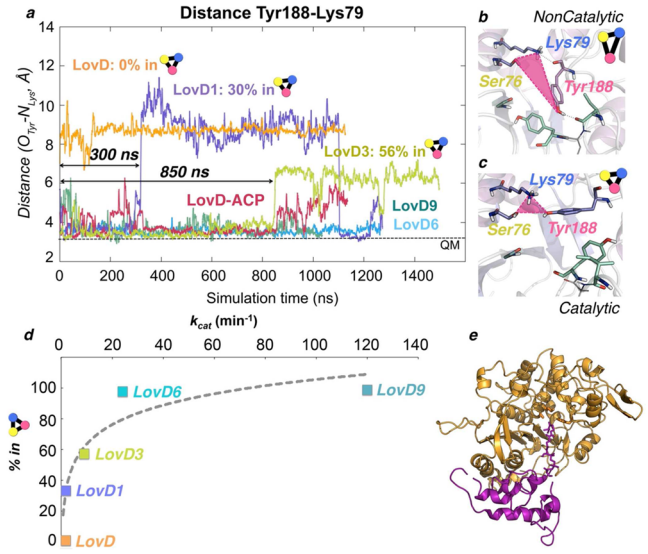


Figure 12. (a) Evolution of the O(Tyr188)−N(Lys79) distance along the MD simulations of all DE-engineered LovD mutants. (b, c) Representative snapshots of (b) wild-type LovD and (c) the final LovD9 mutant. (d) Correlation between the populations of the triad in catalytically competent conformations and the catalytic activities of the mutants. (e) LovD−ACP(LovF) complex.

calculated for the catalytic triad in the unbound state was similar to the TS geometry of the rate-determining step of the process (Figure 11d). It is noteworthy that the time scales of the conformational changes leading to unproductive arrangements of the catalytic triad were around 300 and 850 ns for the first variant, LovD1, and the third variant, LovD3; in the highly evolved LovD6 and LovD9 mutants, the ideal catalytic triad was maintained the whole simulation time. These progressively slower motions translated into increasingly larger populations of the productive conformation (30%, 56%, and 100% for LovD1, LovD3, and the final LovD6 and LovD9 mutants, respectively). Such a conformer distribution correlates well with the observed enzymatic activities (Figure 12d), with the exception of the 6-fold increase from LovD6 to LovD9. As a consequence of the remote mutations introduced during the DE pathway, the optimal QM geometry was sampled with increasing frequency; this was an indication that the free energy of the catalytically active conformation was lowered gradually along the DE process.

MD simulations performed on the wild-type dimer crystal structure and on a model of the complex formed between LovD and the LovF acyl carrier domain demonstrated that protein−protein interactions stabilize the productive arrangement of the catalytic residues (see LovD and LovD-ACP in Figure 12a). These results indicated that the introduction of mutations via directed evolution freed LovD activity from the natural dependence on protein−protein interactions. As a result, LovD was engineered to function without the allosteric modulation exerted by LovF.

As noted earlier, enzyme active site preorganization plays an important role in the acceleration of enzyme-catalyzed reactions.[22,23] Indeed, overlays of X-ray structures of esterases belonging to different folds evidence very similar dispositions of their catalytic residues; it is known that natural unrelated enzymes with similar activities present similarly arranged catalytic machineries.[41] As observed in the previous section, the importance of precision in biocatalysis has recently been demonstrated for the single step unimolecular Kemp elimination.[38] LovD catalyzes multistep acylation−deacylation reactions; the acyl group is first transferred to Ser76 of LovD in a concerted, rate-limiting, and nearly thermoneutral process (Figure 11d). The subsequent acyl group transfer to MJA occurs via a stepwise mechanism, where a tetrahedral intermediate is formed. Microsecond MD simulations were combined with QM calculations to determine the active site structure and dynamics at different reaction stages.[5] The catalytic machinery of LovD was found to be precisely positioned to efficiently stabilize all of the transition states and intermediates involved in the acylation−deacylation process with minimal rearrangement (Figure 11c).

Microsecond MD simulations, in contrast to crystallography and short nanosecond MD, are capable of elucidating the catalytic proficiency of the DE-engineered LovD variants. These simulations reveal that distal mutations have direct consequences for the enzyme active site geometry. Directed evolution includes remote mutations that induce conformational changes transmitted throughout the protein backbone, which stabilize the catalytically competent arrangement of the active site residues. These observations demonstrate that future enzyme design protocols will need to incorporate sub- or microsecond-scale MD simulations to accurately evaluate the computational predictions, but more importantly, mutations not strictly located at the enzyme active site will need to be predicted and incorporated for enhanced enzymatic activity.

4. MD AIDS CRYSTALLOGRAPHY TO LOCATE THE ORIGINS OF DIVERGENT STEREOSELECTIVITY IN ENGINEERED KETOREDUCTASES

Ketoreductases (KREDs) catalyze the asymmetric reduction of ketones to alcohols, as shown in Figure 13a. These enzymes have

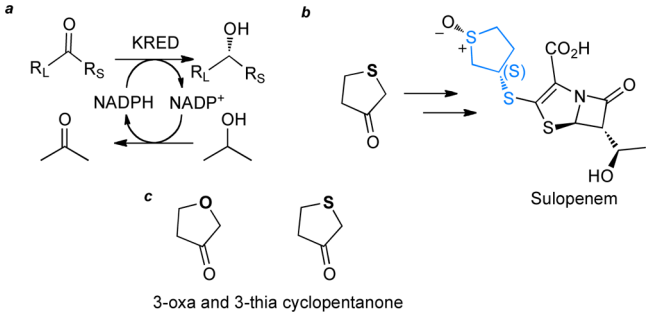


Figure 13. (a) KRED catalytic cycle using isopropanol to regenerate NADPH. (b) 3-Thiacyclopentanone is incorporated into sulopenem. (c) The substrates 3-oxa- and 3-thiacyclopentanone.

been engineered to reduce a wide variety of substrates and are the most commonly used enzymes in industrial pharmaceutical synthesis.[42] Codexis evolved an enzyme for the synthesis of sulopenem,[28] a potent antibiotic whose structure is shown in Figure 13b. The exclusive formation of (*R*)-thiacyclopentanol was achieved by eight rounds of directed evolution starting from the wild-type ketoreductase from *Lactobacillus kefir*, which gives 63% enantiomeric excess (ee) of the *R* enantiomer, to the final deca mutant (G7S/A94T/S96P/R108H/G117S/E145S/N157T/P194N/M206Q/I223 V) Sph, which achieves 99.3% ee of the same *R* enantiomer.[28] During evolution, the single mutation E145S largely increased the *R* selectivity from 63% to 95% ee. The (*R*)-alcohol is converted to the final (*S*)-thioether configuration later in the synthesis.

DOI: 10.1021/ar500452q
Acc. Chem. Res. 2015, 48, 1080−1089

MD simulations, combined with crystallography, and further kinetic data provided an explanation of the observed selectivity, which requires the enzyme to distinguish between sulfur or oxygen and a methylene group, a truly impressive type of selectivity.[8] Figure 14 shows the percent enantiomeric excesses

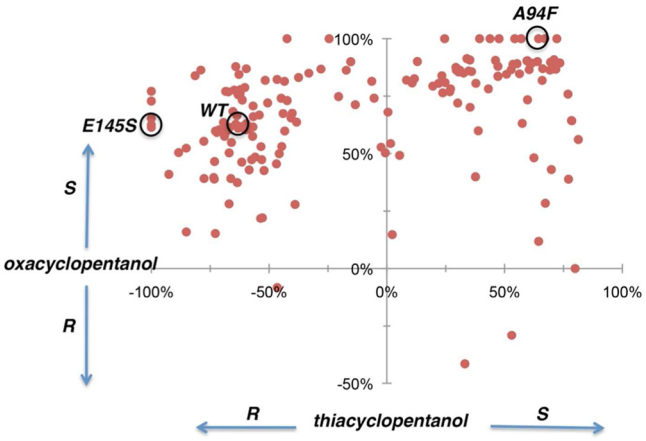


Figure 14. Experimental percent enantiomeric excess (ee) obtained for 3-oxacyclopentanol and 3-thiacyclopentanol by evolved KRED variants. Positive and negative percent ee values indicate excesses of the *S* and *R* enantiomers, respectively. Each point in the graph corresponds to a KRED variant plotted according to the enantioselectivity achieved for each alcohol. The data for the wild-type (WT), E145S, and A94F mutants are circled.

achieved for the reductions of 3-oxacyclopentanone and 3-thiacyclopentanone (Figures 13c and 15b) to their respective alcohols by evolved KRED variants.

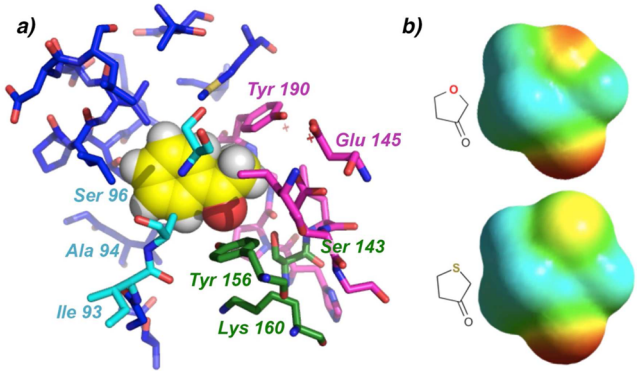


Figure 15. (a) Crystallographic active site in the apo state of wild-type KRED with docked acetophenone (in yellow spheres). The catalytic triad (Ser143, Tyr156, and Lys160) are shown in green, the residues along the small binding pocket in pink, the residues along the large binding pocket in blue, and Ile93, Ala94, and Ser96, which are part of the large binding pocket, in cyan. (b) Electrostatic potential surfaces of 3-oxa- and 3-thiacyclopentanone, where red is negative and blue is positive (isosurface value of ±0.044 hartree), calculated at the B3LYP/3-31G(d) level.

Kinetics studies showed that the wild-type and mutant enzymes bind the substrates very weakly. In fact, the binding is so weak that accurate K_M measurements could not be obtained.[8] The poor binding is due in part to the lack of enough polar groups on the substrates and also to the highly dynamic character of the substrate-binding loop (residues 190−210, the purple/cyan region in Figure 16a), which covers the otherwise solvent-

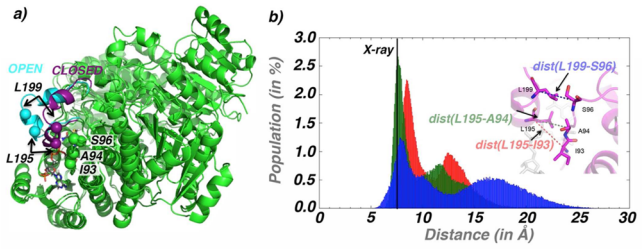


Figure 16. (a) Open (cyan) and closed (purple) conformations of the KRED substrate-binding loop. The Cα atoms of residues Ile93, Ala94, Ser96, Leu195, and Leu199 are shown as spheres. (b) Distributions of distances between Cα atoms of interacting residues derived from 2900 ns MD simulations of wild-type KRED. The X-ray distance (ca. 7 Å) is marked with a vertical line.

exposed active site.[43,44] When MD simulations on the wild-type and the evolved KREDs are performed, the substrate-binding loop opens, and the substrate often moves away from the active site.

Unconstrained MD simulations were performed on selected variants as tetramers, with the cofactor (NADPH) bound to the four active sites and the substrate docked into one of them. These simulations predict the shape and water accessibility of the active site, modulated by the conformational states of the enzyme; these states can be classified according to whether the loop is open, closed, or in an intermediate conformation (see the representative snapshots in Figure 16a and the distributions of characteristic distances in Figure 16b).

The enantioselectivity of the reduction reaction is determined by the relative stabilities of the diastereomeric transition states in the enzyme. To evaluate this, a *theozyme* in which the cofactor and catalytic residues were abbreviated was quantum-mechanically calculated. The *theozyme* is shown in Figure 17a. In the *theozyme*, the side chains of a serine and tyrosine stabilize the oxyanion formed upon hydride transfer from NADPH. The tyrosine is also hydrogen-bonded to the sugar in the cofactor, which also has a hydrogen bond to a protonated lysine. The geometry and charge distribution of the *theozyme* were used to build the substrate−cofactor transition state complexes, an example of which is shown in Figure 17c. These complexes were docked into the active site of the protein, and the protein with the substrate−cofactor transition state complexes served as the starting structures for the MD simulations. In the simulations, the distance between the reactive C atom of the ketone and the H atom of the NADPH and the dihedral angle defining the *Re* or *Si* approach to the ketone were constrained for each of the two diastereomeric *theozymes*. This strategy allowed the rest of the catalytic contacts to fluctuate in response to the protein motions but prevented stereochemical scrambling during the simulations.

The catalytic distances were monitored during the MD simulation. These distances include those from the side-chain oxygen atoms of Tyr156 and Ser143 to the carbonyl oxyanion partially formed in the substrate and also those from the oxygen of the 2′ carbon of NADPH ribose to the side-chain oxygen and nitrogen of Tyr156 and Lys160, respectively (see Figure 17a). These contacts were assumed to facilitate catalysis when less than 3.4 Å (less than the sum of the van der Waals radii of two interacting C atoms). Catalytically competent conformations in the MD simulations were defined as having all four contacts simultaneously below this threshold. The fraction of catalytically competent conformations of transition states for (*R*)- and (*S*)-alcohol formation was computed and compared with the

DOI: 10.1021/ar500452q
Acc. Chem. Res. 2015, 48, 1080−1089

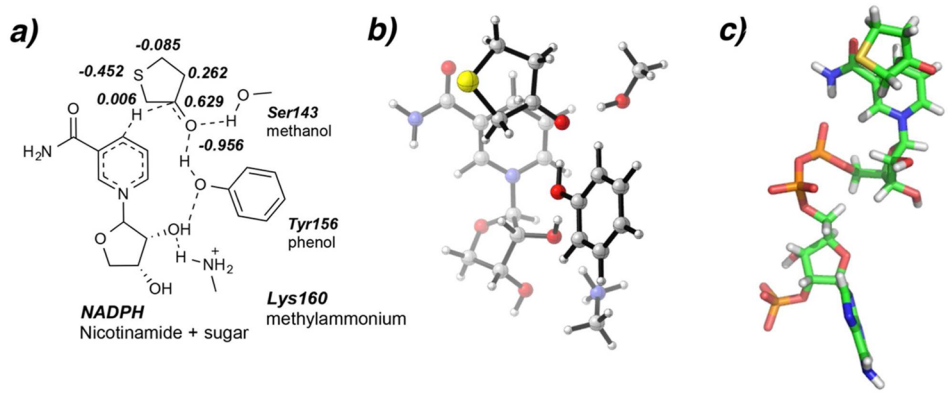


Figure 17. (a) *Theozyme* for 3-thiacyclopentanone reduction in KREDs. The atomic partial charges of the substrate from the *theozyme* are given. (b) DFT-optimized *theozyme* for the formation of (*R*)-thiacyclopentanol. (c) The complete pro-*R* transition structure used in the MD simulations; the geometry of this activation complex was partially restrained to match that of (b).

experimental enantiomeric excess (Figure 18). A conformation of low occurrence in the simulation corresponds to a state with high free energy, whereas a highly sampled conformation possesses a low free energy.

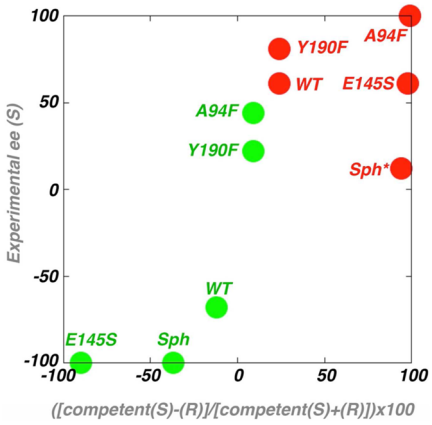


Figure 18. Differences between percentages of catalytically competent pro-*R* and pro-*S* conformations obtained through MD simulations vs the experimental enantiomeric excess of the *S* adducts. Reactions with 3-oxa- and 3-thiacyclopentanone are shown with red and green dots, respectively. Sph* denotes Sph lacking the P194N mutation.

The probability that the enzyme maintains the catalytic contacts to each diastereomeric transition state was found to correlate roughly with the experimental enantiomeric excess. The model predicted the sign of the ee, although the degree of selectivity was not predicted well. The ee values for the highly selective variants (i.e., A94F with oxacyclopentanone, E145S with thiacyclopentanone) were predicted best. These results were used to analyze what types of active site environments favor pro-*R* or pro-*S* reductions for each of the substrates.

The mutations to *L. kefir* modulate the stereoselectivity toward 3-oxa- and 3-thiacyclopentanone. These studies highlight several ways to control enantioselectivity, by creating new hydrophobic contacts or by blocking or allowing water in the active site. Because the relative stabilities of the enzymatic transition states for the formation of the (*R*)- and (*S*)-alcohols determine the selectivity, we developed a novel method of running MD simulations on a restrained *theozyme* that probes how well the enzyme stabilizes these transition structures. Finally, MD shows us what crystallography alone cannot, such as active site hydration, ligand binding poses, and an average ensemble of protein conformations, which may include geometrically quite different states.[8]

5. SUMMARY AND CONCLUSION

The MD simulations described here show how the shapes of active sites of various enzymes depend on subtle changes in the substrate or on mutations remote from the active site. Mutations of noncatalytic residues have profound effects on a variety of factors influencing the reactivity, such as the dynamic motions of catalytic and other active site residues and water penetration. Understanding and controlling these factors will accelerate the development of methods to design highly efficient novel enzymes. Current efforts in this direction are currently ongoing in our group.

AUTHOR INFORMATION

Corresponding Author

*E-mail: houk@chem.ucla.edu.

Notes

The authors declare no competing financial interest.

Biographies

Sílvia Osuna is a Juan de la Cierva postdoctoral researcher at the Institut de Química Computacional i Catàlisi (IQCC) at the University of Girona (Spain). She received her Ph.D. in computational chemistry at the University of Girona in 2010. She was then awarded a Marie Curie IOF Fellowship for a postdoctoral position at the Houk lab at UCLA. She works on the computational exploration of organic and biological reactions. She is an avid artist and painter.

Gonzalo Jiménez-Osés is a postdoctoral scholar at the Houk lab. He received his Ph.D. in asymmetric organic synthesis at the University of La Rioja (Spain) in 2007. He then was first a postdoctoral fellow in computational heterogeneous and green chemistry and then a lecturer in chemistry at the University of Zaragoza (Spain) in 2007−2010. His main research interest is the use of computational methods to explore the interface of chemistry and biology.

Elizabeth L. Noey is a graduate researcher in the Houk lab, computationally studying organic and organometallic reaction mechanisms as well as protein catalysis and enzyme redesign. She is a Foote Senior Fellow and a graduate of the NIH Chemistry−Biology Interface Training Program. She was the USA Triathlon women's collegiate national champion in 2014.

DOI: 10.1021/ar500452q
Acc. Chem. Res. 2015, 48, 1080−1089

K. N. Houk is the Saul Winstein Chair in Organic Chemistry at UCLA. He explores organic and biological reactions with computational methods. He received his Ph.D. as an experimentalist with R. B. Woodward at Harvard in 1968. He is a Fellow of the American Academy of Arts and Sciences, a member of the National Academy of Sciences, and a senior editor of this journal.

■ ACKNOWLEDGMENTS

This research was supported by NIH NIGMS (GM036700, GM097200, and GM075962). ANTON simulations were performed on the National Resource for Biomedical Supercomputing at the Pittsburgh Supercomputing Center with funding from NIH NIGMS (RC2GM093307). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF (OCI-1053575). The authors acknowledge Dr. Jiyong Park for computational support and discussion and Dr. Xiyun Zhang (Codexis) for helpful discussions. S.O. acknowledges the Spanish MINECO for Project CTQ2011-25086/BQU and JdC Grant JCI-2012-14438 and the European Community for CIG Project PCIG14-GA-2013-630978.

■ REFERENCES

(1) Kiss, G.; Çelebi-Ölçüm, N.; Moretti, R.; Baker, D.; Houk, K. N. Computational enzyme design. *Angew. Chem., Int. Ed.* **2013**, *52*, 5700−5725.
(2) Tantillo, D. J.; Jiangang, C.; Houk, K. N. Theozymes and compuzymes: theoretical models for biological catalysis. *Curr. Opin. Chem. Biol.* **1998**, *2*, 743−750.
(3) Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D. De novo enzyme design using Rosetta3. *PLoS One* **2011**, *6*, No. e19230.
(4) Jäckel, C.; Kast, P.; Hilvert, D. Protein design by directed evolution. *Annu. Rev. Biophys.* **2008**, *37*, 153−173.
(5) Jiménez-Osés, G.; Osuna, S.; Gao, X.; Sawaya, M. R.; Gilson, L.; Collier, S. J.; Huisman, G. W.; Yeates, T. O.; Tang, Y.; Houk, K. N. The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat. Chem. Biol.* **2014**, *10*, 431−436.
(6) Jiménez-Osés, G.; Osuna, S.; Noey, E. L.; Houk, K. N. Molecular dynamics explorations of the enhanced activity of evolved enzymes for the Kemp elimination. Submitted for publication.
(7) Kiss, G.; Röthlisberger, D.; Baker, D.; Houk, K. N. Evaluation and ranking of enzyme designs. *Protein Sci.* **2010**, *19*, 1760−1773.
(8) Noey, E. L.; Tibrewal, N.; Park, J.; Bond, C.; Cascio, D.; Jiménez-Osés, G.; Osuna, S.; Liang, J.; Zhang, X.; Huisman, G. W.; Tang, Y.; Houk, K. N.: Origins of stereoselectivity of *Lactobacillus kefir* ketoreductase and evolved mutants. Submitted for publication.
(9) Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Adv. Protein Chem.* **2003**, *66*, 27−85.
(10) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.
(11) Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Kolossváry, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* **2008**, *51*, 91−97.
(12) As described by Lindemann and Rice−Ramsperger−Kassel−Marcus (RRKM) theory. See: Levine, R. D. *Molecular Reaction Dynamics*; Cambridge University Press: Cambridge, U.K., 2005.
(13) Hammes-Schiffer, S.; Benkovic, S. J. Relating protein motion to catalysis. *Annu. Rev. Biochem.* **2006**, *75*, 519−541.
(14) 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.
(15) Black, K.; Liu, P.; Xu, L.; Doubleday, C.; Houk, K. N. Dynamics, transition states, and timing of bond formation in Diels−Alder reactions. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 12860−12865.
(16) Gruebele, M.; Zewail, A. Ultrafast reaction dynamics. *Ber. Bunsen-Ges. Phys. Chem.* **1990**, *94*, 1210−1218.
(17) Laidler, K. J.; King, M. C. Development of transition-state theory. *J. Phys. Chem.* **1983**, *87*, 2657−2664.
(18) Wang, L. H.; Tsai, A. L.; Hsu, P. Y. Substrate binding is the rate-limiting step in thromboxane synthase catalysis. *J. Biol. Chem.* **2001**, *276*, 14737−14743.
(19) Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J. Taking Ockham's razor to enzyme dynamics and catalysis. *Nat. Chem.* **2012**, *4*, 169−176.
(20) Hammes-Schiffer, S. Impact of enzyme motion on activity. *Biochemistry* **2002**, *41*, 13335−13343.
(21) Kamerlin, S. C. L.; Warshel, A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* **2010**, *78*, 1339−1375.
(22) Adamczyk, A. J.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 14115−14120.
(23) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic basis for enzyme catalysis. *Chem. Rev.* **2006**, *106*, 3210−3235.
(24) Pisliakov, A. V.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. Enzyme millisecond conformational dynamics do not catalyze the chemical step. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 17359−17364.
(25) Roca, M.; Oliva, M.; Castillo, R.; Moliner, V.; Tuñón, I. Do dynamic effects play a significant role in enzymatic catalysis? A theoretical analysis of formate dehydrogenase. *Chem.—Eur. J.* **2010**, *16*, 11399−11411.
(26) Hammes-Schiffer, S. Catalytic efficiency of enzymes: a theoretical analysis. *Biochemistry* **2013**, *52*, 2012−2020.
(27) Clarke, A. R.; Wigley, D. B.; Chia, W. N.; Barstow, D.; Atkinson, T.; Holbrook, J. J. Site-directed mutagenesis reveals role of mobile arginine residue in lactate dehydrogenase catalysis. *Nature* **1986**, *324*, 699−702.
(28) Liang, J.; Mundorff, E.; Voladri, R.; Jenne, S.; Gilson, L.; Conway, A.; Krebber, A.; Wong, J.; Huisman, G.; Truesdell, S.; Lalonde, J. Highly enantioselective reduction of a small heterocyclic ketone: biocatalytic reduction of tetrahydrothiophene-3-one to the corresponding (*R*)-alcohol. *Org. Process Res. Dev.* **2010**, *14*, 188−192.
(29) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **2010**, *329*, 305−309.
(30) Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **2013**, *9*, 494−498.
(31) 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. U.S.A.* **2012**, *109*, 3790−3795.
(32) 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. U.S.A.* **2012**, *109*, 10358−10363.
(33) 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.

DOI: 10.1021/ar500452q
Acc. Chem. Res. 2015, 48, 1080−1089

(34) Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **2013**, *339*, 307−310.
(35) Coelho, P. S.; Wang, Z. J.; Ener, M. E.; Baril, S. A.; Kannan, A.; Arnold, F. H.; Brustad, E. M. A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol.* **2013**, *9*, 485−487.
(36) Romero, P. A.; Krause, A.; Arnold, F. H. Navigating the protein fitness landscape with Gaussian processes. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E193−E201.
(37) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *453*, 190−195.
(38) Blomberg, R.; Kries, H.; Pinkas, D. M.; Mittl, P. R. E.; Grutter, 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.
(39) Xie, X.; Meehan, M. J.; Xu, W.; Dorrestein, P. C.; Tang, Y. Acyltransferase mediated polyketide release from a fungal megasynthase. *J. Am. Chem. Soc.* **2009**, *131*, 8388−8389.
(40) Gao, X.; Xie, X.; Pashkov, I.; Sawaya, M. R.; Laidman, J.; Zhang, W.; Cacho, R.; Yeates, T. O.; Tang, Y. Directed evolution and structural characterization of a simvastatin synthase. *Chem. Biol.* **2009**, *16*, 1064−1074.
(41) Smith, A. J. T.; 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. Am. Chem. Soc.* **2008**, *130*, 15361−15373.
(42) Huisman, G. W.; Liang, J.; Krebber, A. Practical chiral alcohol manufacture using ketoreductases. *Curr. Opin. Chem. Biol.* **2010**, *14*, 122−129.
(43) Benach, J.; Atrian, S.; Gonzàlez-Duarte, R.; Ladenstein, R. The catalytic reaction and inhibition mechanism of *Drosophila* alcohol dehydrogenase: observation of an enzyme-bound NAD−ketone adduct at 1.4 Å resolution by X-ray crystallography. *J. Mol. Biol.* **1999**, *289*, 335−355.
(44) Niefind, K.; Müller, J.; Riebel, B.; Hummel, W.; Schomburg, D. The crystal structure of *R*-specific alcohol dehydrogenase from *Lactobacillus brevis* suggests the structural basis of its metal dependency. *J. Mol. Biol.* **2003**, *327*, 317−328.

DOI: 10.1021/ar500452q
Acc. Chem. Res. 2015, 48, 1080−1089