Quantum Mechanical Design of Enzyme Active Sites
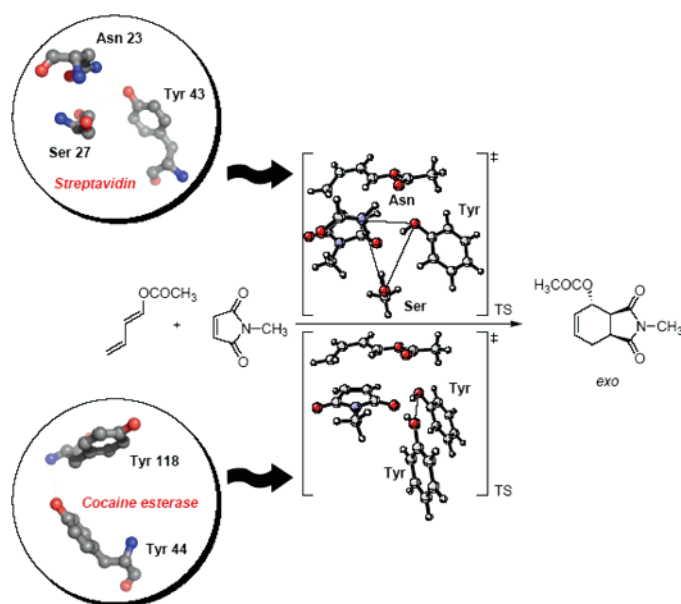
Xiyun Zhang,[†,‡] Jason DeChancie,[§] Hakan Gunaydin,[§] Arnab B. Chowdry,[†]
Fernando R. Clemente,[§] Adam J.T. Smith,[§] T. M. Handel,[*,†] and K. N. Houk[*,§]

*Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California,
San Diego, California 92093 and the Department of Chemistry and Biochemistry,
University of California, Los Angeles, California 90095*

houk@chem.ucla.edu

Received September 15, 2007

The design of active sites has been carried out using quantum mechanical calculations to predict the rate-determining transition state of a desired reaction in presence of the optimal arrangement of catalytic functional groups (theozyme). Eleven versatile reaction targets were chosen, including hydrolysis, dehydration, isomerization, aldol, and Diels−Alder reactions. For each of the targets, the predicted mechanism and the rate-determining transition state (TS) of the uncatalyzed reaction in water is presented. For the rate-determining TS, a catalytic site was designed using naturalistic catalytic units followed by an estimation of the rate acceleration provided by a reoptimization of the catalytic site. Finally, the geometries of the sites were compared to the X-ray structures of related natural enzymes. Recent advances in computational algorithms and power, coupled with successes in computational protein design, have provided a powerful context for undertaking such an endeavor. We propose that theozymes are excellent candidates to serve as the active site models for design processes.

Introduction

Enzymes are powerful catalysts that are capable of increasing reaction rates (k_{cat}/k_{uncat}) by up to 17 orders of magnitude for a variety of reactions.[1,2] Most remarkable, enzymes can perform under ambient conditions and in aqueous solutions.[1,2] Some enzymes display perfect control over stereochemistry and

---

[†] University of California, San Diego.
[‡] Current address: Codexis, Inc. 640 Galveston Drive, Redwood City, CA 94063.
[§] University of California, Los Angeles.

(1) Miller, B. G.; Wolfenden, R. *Annu. Rev. Biochem.* **2002**, *71*, 847−885.
(2) Lad, C.; Williams, N. H.; Wolfenden, R. *Natl. Acad. Sci. U.S.A* **2003**, *100*, 5607−5610.

regioselectivity, whereas others display a breadth of specificity.[3,4] All of these features are attractive for industrial, therapeutic, and synthetic applications.[5−7] Enzyme catalysts for the selective production of normally unfavorable stereochemical pathways in high yields would be of obvious benefit to the synthetic community, in addition to an alternative to the high-temperature and high-pressure systems often employed in industrial synthesis. The design of enzymes remains one of the greatest challenges of protein engineering.[8,9]

Benkovic et al. classified enzyme design efforts into three levels with increasing difficulty.[10] Level 1 includes modifying the substrate or cofactor specificity while retaining native catalytic activity. The engineering of trypsin for non-native substrates[11] and glutathione reductase for another cofactor[12] are the documented success at this level. At Level 2, residues are introduced to proteins that are able to bind the substrates and generate novo activity. Examples include the engineering of a cyclophilin into a proline-specific endopeptidase[13] and converting a PLP-dependent racemase into an aldolase.[14] Level 3 is the most challenging and requires the introduction of both binding and catalysis abilities. This has been attained experimentally through catalytic antibodies raised from eliciting immunization against haptens/transition state analogs of a target reaction. However catalytic antibodies are feeble catalysis with relatively weak binding constants of 10^3-10^5 M^{-1} for transition states.[15] This is due to a number of factors including a general lack of a principle component of catalysis via partially covalent interactions[16] and minimal control of the active site design through the hapten, which is, at best, a rough mimic of the transition state.[17] Computational design could, in principle, tune optimal functionality in the protein to bind the actual transition state or substrate.

Approximate energy functions and minimization algorithms have become more reliable for identifying amino acid sequences compatible with a target tertiary structure.[18,19] This success of inverse folding, as it is known, reached a milestone in the design of a 93-residue α/β protein called Top7, a protein that has a novel sequence and topology by Baker's group.[20] Design efforts

have also been directed toward binding[21,22] and conformational specificity.[23] Enzyme design is a more stringent test of structure prediction methodology and has been met with a few successes to date. Kaplan and DeGrado created an O_2-dependent phenol oxidase within a designed four-helix bundle fold.[24] Bolon and Mayo redesigned thioredoxin to hydrolyze *p*-nitrophenyl acetate using a histidine side-chain yielding a k_{cat}/k_{uncat} of 180.[25] Pokala and Handel have recently published approximate solvation models that produce good correlations with experimental data while providing the speed necessary for design calculations.[26] The most effective design was achieved by Hellinga et al. through the introduction of triose phosphate isomerase activity into a ribose-binding protein with an acceleration of 10^5.[27]

Designing new enzyme catalysts strictly de novo on a routine basis will be extraordinarily difficult. Indeed, the estimated probability of 1 in 10^{77} for randomly generating a functional sequence of a β-lactamase domain presents a staggering obstacle to active site design.[28] However, narrowing the sequence selection to a small area of an existing protein or enzyme scaffold should render enzyme design more feasible.[25,27] This retrofitting strategy[29] can produce hybrid enzymes, a term coined by Benkovic et al. for proteins containing elements from more than one functional protein.[10]

Computational enzyme design methodology has been built into the programs RosettaMatch,[30] ORBIT,[31] DEZYMER,[32] and EGAD.[26] For example, the protocol for enzyme design implemented by Baker and co-workers applies RosttaMatch[30] to place a geometric description of the catalytic site composed of a TS and catalytic residue(s) into existing protein scaffolds, selected based on the ability of the fold to orient the functional groups necessary for proficient catalysis. Then Rosetta design energy functions and minimization methods[33] are subsequently employed to redesign the vicinity of the active site with side-chains that will stabilize the protein and maintain the geometry of the catalytic site. Using rigid protein scaffolds to construct active is geometrically limiting; one can image the probabilistic difficulty of finding a scaffold to geometrically satisfy several catalytic groups simultaneously. Yet, this strategy is the current state-of-the-art. Furthermore, Mayo,[34] Hellinga,[32] and Handel have developed enzyme design algorithms based on a geometric description of an active site followed by redesign methodology to a previously targeted protein scaffold.

(3) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1196−1202.

(4) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186−195.

(5) Bolon, D. N.; Voigt, C. A.; Mayo, S. L. *Curr. Opin. Chem. Biol.* **2002**, *6*, 125−129.

(6) Hilvert, D. *Annu. Rev. Biochem.* **2000**, *69*, 751−793.

(7) Liese, A.; Filho, M. V. *Curr. Opin. Biotechnol.* **1999**, *10*, 595−603.

(8) Tann, C.-M.; Qi, D.; Distefano, M. D. *Curr. Opin. Chem. Biol.* **2001**, *5*, 696−704.

(9) Ballinger, M. D.; Tom, J.; Wells, J. A. *Biochemistry* **1995**, *34*, 13312−13319.

(10) Nixon, A. E.; Ostermeier, M.; Benkovic, S. J. *Trends Biotechnol.* **1998**, *16*, 258−264.

(11) Hedstrom, L.; Szilagyi, L.; Rutter, W. J. *Science* **1992**, *255*, 1249−1253.

(12) Scrutton, N. S.; Berry, A.; Perham, R. N. *Nature* **1990**, *343*, 38−43.

(13) Quemeneur, E.; Moutiez, M.; Charbonnier, J.-B.; Menez, A. *Nature* **1998**, *391*, 301−304.

(14) Seebeck, F. P.; Hilvert, D. *J. Am. Chem. Soc.* **2003**, *125*, 10158−10159.

(15) Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X. *Angew. Chem., Int. Ed.* **2003**, *42*, 4872−4897.

(16) Zhang, X.; Houk, K. N. *Acc. Chem. Res.* **2005**, *38*, 379−385.

(17) Mader, M. M.; Bartlett, P. A. *Chem. Rev.* **1997**, *97*, 1281−1302.

(18) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82−87.

(19) Ponder, J. W.; Richard, F. M. *J. Mol. Biol.* **1987**, *193*, 775−791.

(20) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364−1368.

(21) Shifman, J. M.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13274−13279.

(22) Looger, L. L.; Dwyer, M. A.; Smith, M. B.; Hellinga, H. W. *Nature* **2003**, *423*, 185−190.

(23) Shimaoka, M.; Shifman, J. M.; Jing, H.; Takagi, J.; Mayo, S. L.; Springer, T. A. *Nat. Struct. Biol.* **2000**, *7*, 674−678.

(24) Kaplan, J.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11566−11570.

(25) Bolon, D. N.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14274−14279.

(26) Pokala, N.; Handel, T. M. *J. Mol. Biol.* **2005**, *347*, 203−227.

(27) Dwyer, M. A.; Looger, L. L.; Hellinga, H. W. *Science* **2004**, *304*, 1967−1971.

(28) Axe, D. D. *J. Mol. Biol.* **2004**, *341*, 1295−1315.

(29) Hecht, M. H. In *Protein Engineering and Design*; Carey, P. R., Ed.; Academic Press: New York, 1996; pp 1−50.

(30) Zanghellini, A.; Jiang, L.; Cheng, G.; Althoff, E. A.; Roethlisberger, D.; Wollacott, A. M.; Meiler, J.; Baker, D. *Protein Sci.* **2006**, *15*, 2785−2794.

(31) Dahiyat, B. I.; Mayo, S. L. *Protein Sci.* **1996**, *5*, 256−263.

(32) Hellinga, H. W.; Richard, F. M. *J. Mol. Biol.* **1991**, *222*, 763−785.

(33) Kuhlman, B.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10383−10388.

(34) Lassila, J. K.; Privett, H. K.; Allen, B. D.; Mayo, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16710−16715.

Essential to these design processes is the incorporation of a three-dimensional description of the active site onto a protein scaffold. Quantum mechanics can be used to predict the optimal position of the catalytic functional groups to stabilize a particular transition state. This ideal arrangement calculated for a model system is referred to as a theozyme, short for theoretical enzyme.[35,36] The theozyme model incorporates the transition structure surrounded by the necessary catalytic functional groups that will lower the activation barrier for the reaction. Theozymes serve as the model for the active site structure and predicting activation barriers for the theoretical active site relative to the uncatalyzed reaction in aqueous solution. The geometric and energetic data about an active site, combined with the high probability of the sequence to fold, are the ingredients that should lead to successful enzyme design.

A recent study demonstrates that quantum mechanical theozyme models involving common enzymatic functionalities yield realistic and accurate active site geometries compared to the X-ray crystal structures of known enzymes.[37] We have now applied the theozyme model to a series of eleven reaction targets, including hydrolysis, dehydration, isomerization, aldol, and Diels—Alder reactions, not involving natural substrate, as shown in Table 1. For each of these reactions, we present the uncatalyzed reaction mechanism and reaction coordinate, the designed catalytic site for the rate-determining TS (theozyme), a comparison of the designed site to X-ray crystal structures of related natural enzymes, and the predicted rate accelerations by the designed catalytic site. Our aim is to explore the potential of quantum mechanical calculations as a critical stage in enzyme design and to provide designs for active sites that might be incorporated into full enzyme designs in the future.

Computational Methods

Calculation of Reaction Mechanism and Reaction Coordinate. B3LYP/6-31G(d) geometry optimizations were carried out using Gaussian 03[38] in the gas phase for all the reaction targets except Reaction 2. For Reaction 2, due to the large size of the system, HF/3-21G optimizations were used. B3LYP and HF are accepted as suitable methods to deliver accurate geometries for a variety of previously studied reactions;[39-41] however, reported energies using these methods are known to be less accurate than higher-level

(35) Na, J.; Houk, K. N. *J. Am. Chem. Soc.* **1996**, *118*, 9204−9205.

(36) Tantillo, D.; Chen, J.; Houk, K. N. *Curr. Opin. Chem. Biol.* **1998**, *2*, 743−750.

(37) DeChancie, J.; Clemente, F. R.; Gunaydin, H.; Smith, A. J. T.; Zhang, X.; Houk, K. N. *Protein Sci.* **2007**, *16*, 1851−1866.

(38) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(39) Wiest, O.; Houk, K. N. *Top. Curr. Chem.* **1996**, *183*, 1−24.

(40) Wiest, O.; Montiel, D. C.; Houk, K. N. *J. Phys. Chem. A* **1997**, *101*, 8378−8388.

(41) Guner, V.; Khuong, K. S.; Leach, A. G.; Lee, P. S.; Bartberger, M. D.; Houk, K. N. *J. Phys. Chem. A* **2003**, *107*, 11445−11459.

TABLE 1. List of Reaction Targets

Entry	Reaction Target
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	

electron correlation methods. Frequency calculations were performed to verify the nature of all stationary points as minima or transition states, to compute free energies, and to obtain thermal corrections to the Gibbs free energy at 298 K. To mimic, if only crudely, the interior of a protein, the theozyme energetics were re-evaluated for a value of $\epsilon = 4.33$. Because the uncatalyzed reactions of interest were carried out in aqueous solution, solvation energies were calculated at the HF/6-31+G(d) (Reactions 1−9), or HF/6-31+G(d,p) level (Reactions 10 and 11) and with the CPCM−UAKS model for the geometries optimized in the gas phase in conjugation with the dielectric of water ($\epsilon = 78.39$).[42] To simulate a hydrophobic environment of a protein interior, the catalyzed reactions were treated with the PCM continuum model using a dielectric of 4.0. The calculated overall activation free energies of the uncatalyzed reaction, $\Delta G^{\ddagger}_{uncat}$, are compared with accessible experimental data to validate the calculations. The uncatalyzed reactions in water are

(42) Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, *1*, 70−77.

often assumed to involve water ionization, which generates, albeit slowly, hydroxide and hydronioum catalysts.[43,44]

Design of Catalytic Site and Estimation of Rate Acceleration. A catalytic mechanism for each reaction target was chosen, inspired by natural enzymes for similar reactions. A catalytic site was designed by selecting 1−5 functional groups to carry out the desired chemistry and performing the same geometry optimizations as for the uncatalyzed reaction. Functional groups are those present on natural amino acid side-chains.[35,36] For example, acetamide was used in place of the side chain for asparagine or glutamine. The functional groups were then optimized to maximally stabilize the transition structure with respect to the assumed initial starting structure. Initial geometries for the applied array of functional groups (adapted from naturally occurring enzymatic active sites) to stabilize the transition structure were placed manually using close to ideal distances and angles for hydrogen-bonding and covalent interactions. The same functional groups were also optimized surrounding the substrate. The activation barrier of the catalyzed reaction, $\Delta G^{\ddagger}_{cat}$, is compared to $\Delta G^{\ddagger}_{uncat}$ to estimate the rate acceleration by the catalytic site.

Verification of Catalytic Site Geometries. For catalytic sites composed of 2−5 functional groups, the predicted geometries were compared against the catalytic residues present in related enzymes: serine hydrolases, tyrosyl-DNA phosphodiesterase, scylatone dehydratase, cyclophilin, deoxyribose-phosphate aldolase (DERA), streptavidin, cocaine esterase, and carbonic anhydrase. All of the natural proteins have high-resolution crystal structures available. Similarity between natural and predicted sites was characterized both qualitatively based on orientation and quantitatively based on distances between chemically relevant atoms.

Results and Discussion

Five criteria were applied to select viable reaction targets for this study and for possible future developments of a functional enzyme. Reactions should be slow in water, quantified here as having a half-life of at least one week; substrates must be commercially available or easy to prepare and be soluble in water; reactions are preferably unimolecular and involve no cofactors at this early stage of design studies and have interest for medicinal or synthetic applications. In most cases, other proteins have been developed as catalysts, establishing the feasibility of the reaction target for catalysis. Each of the 11 selected reaction targets, listed in Table 1, meet at least three of the four criteria.

The reactions are versatile, providing a thorough test for how general this protocol can be. Reactions 1−3 are ester and amide hydrolysis reactions. This type of reaction has been extensively investigated, with respect to enzyme models.[45,46] Convenient fluorescent and biological assays have been developed, as well.[47] Reaction 1, the hydrolysis of *p*-nitrophenyl acetate (PNP), has an easy assay to monitor the *p*-nitrophenoxide ion produced during hydrolysis.[48] A binding protein has previously been altered to act as a catalyst.[25] Reaction 2, the hydrolysis of cocaine, is interesting due to its applications in human therapy.[49−52] Reaction 3, the hydrolysis of acylindole, is more challenging than ester hydrolysis and has a catalytic antibody

capable of 750-fold rate acceleration.[53] Reaction 4, hydrolysis of the organophosphorus nerve agent sarin, neutralizes the toxic effects of the nerve agent, and an enzyme that catalyzes this reaction would represent a potential prophylactic treatment.[54−57] Reaction 5, scytalone dehydration, is fundamental in melanin biosynthesis and can be monitored due to the production of a conjugated product.[58] Reaction 6, the *cis−trans* isomerization of a prolyl peptide bond, is of great biological importance in protein folding and stability.[59] Reactions 7 and 11, intramolecular aldol reactions, are examples of a common type of reaction utilized in organic synthesis.[60−62]

Reactions 8−10 are Diels−Alder reactions and are the only bimolecular reactions in our set. While this increases the complexity of the calculations, it potentially allows a designed enzyme to act as an entropy trap, a strategy that has proven successful for Diels−Alder catalytic antibodies.[63] Diels−Alder reactions normally give a mixture of *endo-* and *exo-*products. The *endo-*pathway is generally more favorable, due to secondary orbital electrostatic an steric interactions.[64] The generation of the *exo-*product, therefore, may be of synthetic significance and is the desired target of the catalytic site designs for Reactions 8 and 9. Reaction 10 is a step in a potential synthesis of Vitamin B6.[65]

The mechanism and reaction coordinate for each uncatalyzed reaction in water are detailed below. The calculated overall activation barriers, $\Delta G^{\ddagger}_{calc}$, are summarized and compared with available experimental data, $\Delta G^{\ddagger}_{exp}$, in Table 2. The structures of designed catalytic sites, respective natural counterparts, and predicted rate accelerations are summarized in Table 3. Only the terminal moieties used during geometry optimizations are depicted in Table 3, though the amino acid side chains they represent are indicated.

1. Hydrolysis of *p*-Nitrophenyl Acetate (PNP). The uncatalyzed reaction 1 (see Supporting Information) is postulated to involve three steps: (1) water autoionization generates hydroxide and hydronium, (2) hydroxide and hydronium simultaneously attack the substrate to generate the gem-diol intermediate, and (3) the elimination of the gem-diol intermedi-

(43) Zhang, X.; Houk, K. N. *J. Org. Chem.* **2005**, *70*, 9712−9716.

(44) Gunaydin, H.; Houk, K. N. UCLA: 2007.

(45) Thomas, N. R. *Nat. Prod. Rep.* **1996**, 479−511.

(46) Lerner, R. A.; Benkovic, S. J.; Schultz, P. G. *Science* **1991**, *252*, 659−667.

(47) Tanaka, F. *Chem. Rev.* **2002**, *102*, 4885−4906.

(48) Anderson, J.; Byrne, T.; Woelfel, K. J.; Meany, J. E.; Spyridis, G. T.; Pocker, Y. *J. Chem. Educ.* **1994**, *71*, 715−718.

(49) Kuhar, M. J. *Ciba Found. Symp.* **1992**, *166*, 81−95.

(50) Boelsteri, U. A.; Goldin, C. *Arch. Toxicol.* **1991**, *65*, 351−360.

(51) Hahn, I.-H.; Hoffman, R. S. *Emerg. Med. Clin. North Am.* **2001**, *19*, 493−511.

(52) Landry, D. W.; Zhao, K.; Yang, G. X.; Glickman, M.; Georgiadis, T. M. *Science* **1993**, *259*, 1899−1901.

(53) Benedetti, F.; Berti, F.; Colombatti, A.; Ebert, C.; Linda, P.; Tonizzo, F. *Chem. Commun.* **1996**, 1417−1418.

(54) Doctor, B. P.; Raveh, L.; Wolfe, A. D.; Maxwell, D. M.; Ashani, Y. *Neurosci. Biobehav. Rev.* **1991**, *15*, 123−128.

(55) Raveh, L.; Grunwald, J.; Papier, Y.; Cohen, E.; Ashani, Y. *Biochem. Pharmacol.* **1993**, *45*, 2465−2474.

(56) Broomfield, C. A.; Lockridge, O.; Millard, C. B. *Chem. Biol. Interact.* **1999**, *119−120*, 413−418.

(57) Saxena, A.; Sun, W.; Luo, C.; Myers, T. M.; Koplovitz, I.; Lenz, D. E.; Doctor, B. P. *Neuroscience* **2006**, *30*, 145−148.

(58) Nixon, A. E.; Firestine, S. M.; Salinas, F. G.; Benkovic, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3568−3571.

(59) Schreiber, S. L.; Albers, M. W.; Brown, E. J. *Acc. Chem. Res.* **1993**, *26*, 412−420.

(60) Agami, C. *Bull. Soc. Chim. Fr.* **1988**, 499−507.

(61) Agami, C.; Platzer, N.; Sevestre, H. *Bull. Soc. Chim. Fr.* **1987**, *2*, 358−360.

(62) List, B.; Lerner, R. A.; Barbas, C. F., III. *Org. Lett.* **1999**, *1*, 59−62.

(63) Kim, S. P.; Leach, A. G.; Houk, K. N. *J. Org. Chem.* **2002**, *67*, 4250−4260.

(64) March, J. A.; Smith, M. B. *Advanced Organic Chemistry. Reactions, Mechanisms and Structure*, 5th ed.; John Wiley & Sons: New York, 2001.

(65) Karpeiskii, M. Y.; Florent'ev, V. R. *Russ. Chem. Rev.* **1969**, *38*, 1244−1256.

TABLE 2. Calculated Activation Barriers for Reactions 1−11 and Comparison with Available Experimental Data

reaction	$\Delta G^{\ddagger}_{calc}$ (kcal/mol)	$\Delta G^{\ddagger}_{exp}$ (kcal/mol)	absolute deviation (kcal/mol)
1	24.4	25.9[66]	1.5
2	28.8	26.1[69]	2.7
3	30.8	29.2[53]	1.6
4	28.2	25.1[72]	3.1
5	23.8	24.1[58]	0.3
6	23.1	20.2[76]	2.9
7	34.4	31.9[62]	2.5
8	36.2	N/A	N/A
9	32.1	N/A	N/A
10	40.7	N/A	N/A
11	42.0	N/A	N/A

ate to give the products *p*-nitrophenol and acetic acid. The activation barrier for water autoionization is 23.8 kcal/mol.[43,44] The activation free energy for decomposition of the gem-diol intermediate to give hydroxide and hydronium was taken from a previously computed value of 17.6 kcal/mol;[59] in addition to the activation barrier for elimination of the gem-diol intermediate of 18.0 kcal/mol.[59] These previously computed values integrated with ground state energies relative to the reactants gave the calculated activation free energy for the rate-limiting step of 24.4 kcal/mol. This activation barrier is in reasonable agreement with the experimentally determined value of 25.9 kcal/mol.[66]

The catalytic site design was inspired by the Ser−His−Asp catalytic triad found in serine proteases: serine is the nucleophile that forms a covalent intermediate with the substrate, histidine, and aspartate are general acid/base residues that activate serine, catalyze the formation of the covalent intermediate and its subsequent hydrolysis.[67] The activation barrier of the catalyzed reaction, $\Delta G^{\ddagger}_{cat}$, was calculated to be 17.0 kcal/mol. When compared to the experimentally determined uncatalyzed activation barrier, $\Delta G^{\ddagger}_{uncat}$, the catalytic triad could lower the activation barrier by 7.4 kcal/mol. With the incorporation of an additional oxyanion hole, it may be possible to lower this barrier even more dramatically.

To determine if the geometry of the catalytic site is a reasonable one, we compared it against arylesterase, an enzyme that catalyzes the hydrolysis of phenyl acetate to phenol and acetate (EC 3.1.1.2, PDB code: 1AV4).[68] The geometry of the catalytic triad is well maintained.

2. Hydrolysis of Cocaine. Reaction 2 (see Supporting Information) goes through three steps: (1) water autoionization, (2) formation of the gem-diol intermediate from cocaine, hydroxide and hydronium and (3) the collapse of the gem-diol intermediate into the products. Again, the activation free energies for barriers toward reactants and products originating from the gem-diol intermediate were taken from previous work.[59] The activation free energy was calculated to be 28.8 kcal/mol for the rate-limiting step, which correlates reasonably well with the experimental value of 26.1 kcal/mol.[69]

Like Reaction 1, the catalytic site was based on the Ser−His−Asp catalytic triad borrowed from serine proteases. It

stabilizes the hydroxide association TS. $\Delta G^{\ddagger}_{cat}$ is 21.0 kcal/mol, 7.8 kcal/mol lower than the experimentally determined $\Delta G^{\ddagger}_{uncat}$.

The geometry of the catalytic site is identical to that designed for Reaction 1. We also compared the catalytic site against natural cocaine esterase (EC 3.1.1.1, PDB code: 1AUO).[70] The geometry of the catalytic triad is well maintained and, again, Asp168 is perpendicular to His199 in the apo form of the natural cocaine esterase.

3. Hydrolysis of Acylindole. Reaction 3 (see Supporting Information) again starts with water autoionization followed by the formation of the gem-diol intermediate with the hydroxide and hydronium. Subsequent C−N dissociation gives indole and acetic acid. The activation free energy for decomposition of the amide containing gem-diol intermediate to give hydroxide and hydronium was taken from a previously computed value of 16.6 kcal/mol,[59] in addition to the activation barrier for elimination of the gem-diol intermediate of 17.3 kcal/mol.[59] These previously computed values integrated with ground state energies relative to the reactants gave the calculated activation free energy for the rate-limiting step of 30.8 kcal/mol, which is in good agreement with the experimentally determined value of 29.2 kcal/mol.[53]

The catalytic site was based on the Ser−His−Asp catalytic triad. It stabilizes the hydroxide association TS. $\Delta G^{\ddagger}_{cat}$ is 16.8 kcal/mol, 14 kcal/mol lower than $\Delta G^{\ddagger}_{uncat}$ determined experimentally. It should be noted, however, that the second TS is very close in energy to the first, and it could also become rate-determining.

The geometry of the catalytic site is like those designed for Reactions 1 and 2. It was compared against streptogrisin B (EC 3.4.21.81, PDB code: 1DS2),[71] a natural serine protease. The geometry of the catalytic triad is maintained, except that Asp168 is perpendicular to His199 in the apo form of the natural cocaine esterase. In addition, Asp168 hydrogen bonds with the backbone amide of His199.

4. Hydrolysis of Sarin Nerve Agent. Reaction 4 (see Supporting Information) involves the hydrolysis of the phosphotriester substrate sarin. The free energy of activation for the rate-limiting step, as calculated herein, is 28.2 kcal/mol, in reasonable agreement with the experimental barrier of 25.1 kcal/mol.[72]

The designed catalytic site is composed of three main catalytic units: a nucleophile, an oxyanion hole, and a general base. The nucleophile is a histidine backed up by an aspartate, the oxyanion hole is a lysine/asparagine dyad, and the general base is an aspartate. The origin of sarin's toxicity is its covalent inhibition of acetylcholinesterase, thereby shutting down neuromuscular communication, leading to paralysis and death. The rate-limiting step in this process is the base-catalyzed attack of water on the phosphoenzyme covalent intermediate[73] so this is the step that has been modeled. The calculated free energy barrier is 9.4 kcal/mol, which is 18.8 kcal/mol below the uncatalyzed barrier.

(66) Tantillo, D.; Houk, K. N. *J. Comput. Chem.* **2002**, *23*, 84−95.

(67) Hedstrom, L. *Chem. Rev.* **2002**, *102*, 4501−4524.

(68) Wilce, M. C.; Dooley, D. M.; Freeman, H. C.; Guss, J. M.; Matsunami, H.; McIntire, W. S.; Ruggiero, C. E.; Tanizawa, K.; Yamaguchi, H. *Biochemistry* **1997**, *36*, 16116−16133.

(69) Turner, J. M.; Larsen, N. A.; Basran, A.; Barbas, C. F., III.; Bruce, N. C.; Wilson, I. A.; Lerner, R. A. *Biochemistry* **2002**, *41*, 12297−12307.
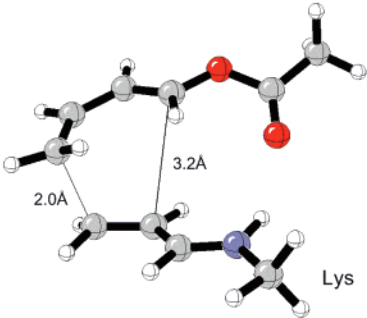
(70) Kim, K. K.; Song, H. K.; Shin, D. H.; Hwang, K. Y.; Choe, S.; Yoo, O. J.; Suh, S. W. *Structure* **1997**, *5*, 1571−1584.

(71) Bateman, K. S.; Huang, K.; Anderson, S.; Lu, W.; Qasim, M. A.; Laskowski, J., M.; James, M. N. *J. Mol. Biol.* **2001**, *305*, 839−849.

(72) Duman, D. P.; Durst, H. D.; Landis, W. G.; Raushel, F. M.; Wild, J. R. *Arch. Biochem. Biophys.* **1990**, *277*, 155−159.

(73) Millard, C. B.; Kryger, G.; Ordentlich, A.; Greenblatt, H. M.; Harel, M.; Raves, M. L.; Segall, Y.; Barak, D.; Shafferman, A.; Silman, I.; Sussman, J. L. *Biochemistry* **1999**, *38*, 7032−7039.
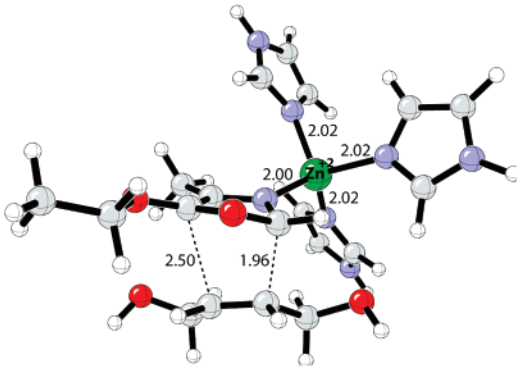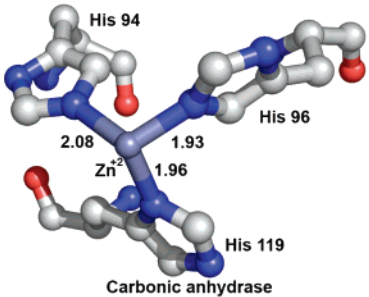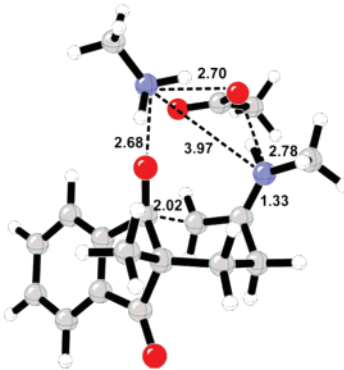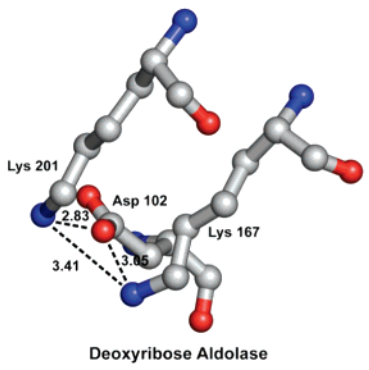
TABLE 3. Structures of Designed Catalytic Sites, Respective Natural Counterparts, and Predicted Rate Accelerations

Reaction	Designed Catalytic Site	Natural Counterpart	$\Delta\Delta G^{\ddagger}$ (kcal/mol)
1		Arylesterase	−7.4
2		Carboxylesterase	−7.8
3		Streptogrisin B	−14.0
4			-18.8

Table 3. (Continued)

Reaction	Designed Catalytic Site	Natural Counterpart	$\Delta\Delta G^{\ddagger}$ (kcal/mol)
5		Scytalone Dehydrase	−12.0
6		Cyclophilin	-3.1
7		Deoxyribose-phosphate Aldolase	−25.5
8		Streptavidin	-3.5

Table 3. (Continued)

Reaction	Designed Catalytic Site	Natural Counterpart	$\Delta\Delta G^{\ddagger}$ (kcal/mol)
9		N/A	−9.1
10		Carbonic anhydrase	-7.4
11		Deoxyribose Aldolase	-33.7

The nucleophile and oxyanion hole units were inspired by tyrosyl-DNA phosphodiesterase (Tdp1, E.C. 3.1.4.-, PDB Code: 1NOP).[74] The functional groups representing the asparagine, lysine, and histidine look very similar to their positioning in the enzyme. The aspartate oxygen that hydrogen bonds with the nucleophilic His is also in the same place, but as with the catalytic triad models it is in plane with the His instead of out of the plane. The general base aspartate is positioned where the analogous His general base is in the Tdp1 active site.

5. Dehydration of Scytalone. Reaction 5 (see Supporting Information) is a three step process: (1) water autoionization generates hydroxide and hydronium ions, (2) hydroxide and hydronium ions rapidly convert scylatone into an enol intermediate, and (3) hydroxide and hydronium ions catalyze the water elimination of enol to give the final product. The water autoionization step determines the activation free energy to be
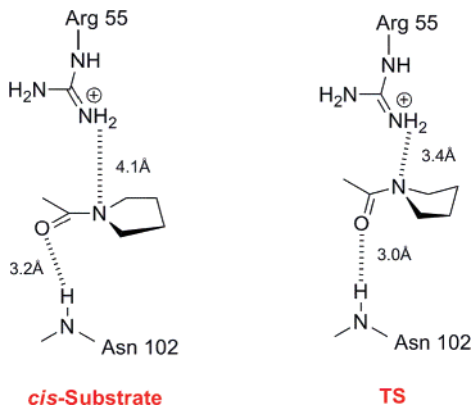
23.8 kcal/mol, comparable to the experimentally observed value of 24.1 kcal/mol.[58]

The catalytic site was designed based on the natural scytalone dehydratase (EC 4.2.1.94, PDB code: 2STD).[75] As shown in Table 3, the catalytic site of natural scytalone dehydratase consists of five residues: Tyr30 and Tyr50 assist the protonation of the substrate, carbonyl, His85 abstracts a proton, Asp31 activates His85, and His110 assists hydroxide elimination.[58] The designed catalytic site is composed of histidine, aspartate (or glutamate), and lysine. Histidine catalyzes hydroxide elimination and aspartate and lysine compose general acid/base residues. $\Delta G^{\ddagger}_{cat}$ is 11.8 kcal/mol, 12.0 kcal/mol lower than $\Delta G^{\ddagger}_{uncat}$ determined experimentally.

6. Isomerization of a Prolyl Peptide Bond. Reaction 6 (see Supporting Information) involves *cis*-proline rotating about its peptide bond to form *trans*-proline, a rotation from 2.9 to 176.3°.

(74) Davies, D. R.; Interthal, H.; Champoux, J. J.; Hol, W. G. J. *Chem. Biol.* **2003**, *10*, 139−147.

(75) Nakasako, M.; Motoyama, T.; Kurahashi, Y.; Yamaguchi, I. *Biochemistry* **1998**, *37*, 9931−9939.

SCHEME 1. Catalytic Mechanism of Cyclophilin Proposed by Hur and Bruice[78]

The calculated transition state involves a rotation of the peptide bond to 56.4° and forms a favorable C−H···O interaction. The calculated activation free energy is 23.1 kcal/mol, close to the experimental value of 20.2 kcal/mol.[76]

The designed catalytic site used cyclophilin (EC 5.2.1.8, PDB code: 1AMH),[77] the natural enzyme for cis-trans isomerization of prolyl peptide bonds, as a reference. As shown in Table 3, cyclophilin has two catalytic residues: Arg55 and Asn102. The proposed catalytic mechanism, shown in Scheme 1, involves the guanidinium of Arg55 stabilizing the lone pair becoming localized on the prolyl amide N in the TS and the backbone of Asn102 interacting with the amide carbonyl O in the *cis*-substrate and TS.[78]

Because positioning side chains is an easier problem than positioning backbone atoms for protein design, the backbone interaction was replaced by the side chain of asparagine. Still, this site yielded an activation barrier only 2.6 kcal/mol lower than that of the uncatalyzed reaction because arginine binds strongly to the carbonyl O of the non-prolyl amide bond instead. Replacing the arginine with a lysine reinforced the stabilization of the prolyl amide N lone pair and gave $\Delta G^{\ddagger}_{cat}$ of 20.0 kcal/mol, 3.1 kcal/mol lower than $\Delta G^{\ddagger}_{uncat}$ determined experimentally. This catalytic site therefore provided more reasonable acceleration and is the one selected for inclusion in Table 3.

7. Intramolecular Aldol Reaction. Reaction 7 (see Supporting Information) proceeds though three steps: (1) water autoionization generates hydroxide and hydronium, (2) the generated ions interact with the substrate to form an enol intermediate, regenerating two water molecules, and (3) a C−C bond forms via the intramolecular aldol addition, with the simultaneous transfer of a proton from the enol oxygen to the carbonyl oxygen. C−C bond formation is the rate determining step. The activation free energy of 34.4 kcal/mol is close to the experimental value of 31.9 kcal/mol.[62]

The catalytic site was designed based on a natural type I aldolase, Deoxyribose-phosphate Aldolase (DERA) (EC 4.1.2.4, PDB code: 1JCL).[79] It is composed of three catalytic residues: one lysine is a nucleophile to form an enamine intermediate with the substrate; one aspartate and another lysine are general acid/base residues that facilitate the formation of enamine and

its subsequent hydrolysis. $\Delta G^{\ddagger}_{cat}$ is 8.9 kcal/mol, 25.5 kcal/mol lower than $\Delta G^{\ddagger}_{uncat}$ determined experimentally.

We compared the geometry of the optimized catalytic site to that of DERA. The geometry of the catalytic site is well maintained. A notable feature is that in DERA, the distance between the ϵ-N atoms of Lys167 and Lys201 is 3.4 Å. This close distance was proposed to be critical for the pK_a shift of Lys167 to neutral.[79] In the designed catalytic site, the corresponding N−N distance is 3.7 Å. Furthermore, in the designed catalytic site, both aspartate O atoms hydrogen bond with lysine N atom, presenting a more self-saturating H-bonding network.

8. Diels−Alder Reaction between 1-Acetoxybutadiene and N-Methylmaleimide. The Diels−Alder reaction can proceed through either an *endo-* or *exo*-pathway. The *endo*-pathway is favored over the *exo*, yielding a calculated product ratio of 720:1. The uncatalyzed *exo*-pathway has a calculated activation free energy of 36.2 kcal/mol. While it may be difficult to design a set of catalytic residues that predictably favor the *exo*-pathway, it should be easier to design a cavity with the appropriate shape complementarily for the *exo*-pathway. This data only presents the design of the catalytic site, not the protein cavity into which it could be grafted.

Diels−Alder reactions can be accelerated by hydrogen-bonding to the dienophile, to lower the LUMO.[64] The designed catalytic site was based on the oxyanion hole from streptavidin,[43] and uses a tyrosine, asparagine, and serine residue to withdraw density from the dienophile. The geometry of the Ser−Tyr−Asn oxyanion hole was compared to that of streptavidin (PDB code: 1STP).[80] The orientation of the catalytic residues is very well maintained albeit the distances between the three catalytic residues are closer in streptavidin. $\Delta G^{\ddagger}_{cat}$ is 32.7 kcal/mol, 3.5 kcal/mol lower than $\Delta G^{\ddagger}_{uncat}$ determined computationally. However, the catalytic contribution from a decrease in activation entropy achieved by the close proximity of the reacting substrates in the active site of the enzyme is not included. Therefore, the predicted activation barrier lowering should be considered a lower bound. A similar situation applied for the following Diels−Alder reactions 9 and 10.

9. Diels−Alder Reaction between 1-Acetoxylbutadiene and Acrolein. The second Diels−Alder reaction uses acrolein as the dienophile, but employs the same diene as Reaction 8. The uncatalyzed *exo*-pathway has a calculated activation barrier of 32.1 kcal/mol. Instead of using an oxyanion hole, the catalytic site is composed of a single lysine residue that forms an iminium intermediate with the dienophile adopted from Ahrendt et al.[82] As shown in Scheme 2, in the presence of a secondary amine catalyst, 1-acetoxybutadine undergoes a Diels−Alder reaction with acrolein to give a mixture of *endo-* and *exo*-products with an 11:1 ratio. Likewise, the designed protein is assumed to use the ϵ-amino group of Lys to attack the dienophile aldehyde and form an iminium intermediate assuming complementary catalytic functionality as present in DERA. For the Diels−Alder reaction between 1-acetoxybutadiene and N-allylidenemethanaminium, the calculated activation barrier is 23.0 kcal/mol, 9.1 kcal/mol lower than that between 1-acetoxybutadiene acetate and acrylaldehyde, determined computationally.

(76) Radzicka, A.; Wolfenden, R. *Science* **1995**, *267*, 90−93.
(77) Perona, J. J.; Craik, C. S. *Protein Sci.* **1995**, *4*, 337−360.
(78) Hur, S.; Bruice, T. C. *J. Am. Chem. Soc.* **2002**, *124*, 7303−7313.
(79) Heine, A.; DeSantis, G.; Luz, J. G.; Mitchell, M.; Wong, C.-H.; Wilson, I. A. *Science* **2001**, *294*, 396−374.

(80) Weber, P. C.; Wendoloski, J. J.; Pantoliano, M. W.; Salemme, F. R. *J. Am. Chem. Soc.* **1992**, *114*, 3197−3200.
(81) Larsen, N. A.; Turner, J. M.; Stevens, J.; Rosser, S. J.; Basran, A.; Lerner, R. A.; Bruce, N. C.; Wilson, I. A. *Nat. Struct. Biol.* **2002**, *9*, 17−21.
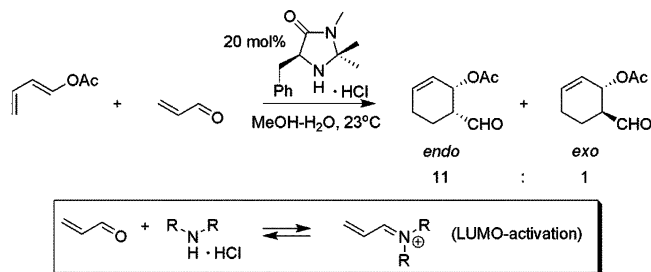(82) Ahrendt, K. A.; Borths, C. J.; MacMillan, D. W. C. *J. Am. Chem. Soc.* **2000**, *122*, 4243−4244.

SCHEME 2. Enantioselective Organocatalytic Diels−Alder Reaction Reported by MacMillan et al.[72]

Conclusion and Outlook

The creation of an effective catalytic active site is fundamental to the success of computational enzyme design. Theozyme calculations are efficient and valuable models to produce potential active site structures and estimations for catalytic efficacy for non-natural reaction targets for enzyme design. Ab initio (HF/3-21G) and density functional (B3LYP/6-31G(d)) calculations were performed to investigate the mechanisms of eleven reactions of design interest, to locate the rate-determining transition states, to design catalytic sites composed of 1−5 catalytic residues, and to predict rate accelerations.

In this study, we have used existing data from crystal structures as an initial validation of our computationally derived active sites, similar to previous theozyme calculations.[35] In the case of the catalytic triad, our comparison revealed a consistent variation between natural active sites and their designed counterparts. Our theozyme optimizations consistently oriented the Asp residue of catalytic triads in a *syn*-in plane hydrogen-bonding fashion with respect to the histidine model (imidazole), whereas natural sites show a *syn*-perpendicular hydrogen-bonding geometry. This discrepancy is an artifact of the gas phase in the QM calculations and exclusion of secondary hydrogen-bonding contacts. In arylesterase, Asp222 makes a hydrogen bond with the backbone of Ile224, a noncatalytic position. The same can be seen for carboxylesterase, where Asp168 hydrogen bonds with the backbone of Val170 and Streptogrisin's Asp102 hydrogen bonds with the backbone of the active histidine. All of these interactions are removed in the QM simulation, where only the terminal moieties of catalytic residues are explicitly considered. However, these structural observations do not disqualify the calculated structures (*syn*-in plane hydrogen-bonding arrangement) as competent catalysts.

Another documented feature of the catalytic triad is its geometric invariance. The relative orientation of the catalytic residues does not change significantly, despite variation in substrate specificity. It is generally accepted that the specificity of enzymes that employ the catalytic triad comes from the shape of the active site cavity, as determined by the protein backbone and other side chains. For specificity, it is vital to consider the possible interactions between protein and substrate, as well as the positioning of catalytic residues. Specificity is not a consideration for the QM calculations presented.

We focus on designing catalytic sites to stabilize the rate-determining transition state for a particular reaction in water. However, enzymes must stabilize all the transition states and maintain stereochemical complementarily along the reaction coordinate. The design of such dynamic catalytic sites is not addressed; instead, we assume that the transition state ensemble can be approximated by a single transition structure. This problem can be approached in the future by calculating theozymes for all of the transition states in the reaction.

The hallmark of the calculated theozymes in this work is the adaptation of nature's highly evolved catalytic groups for the construction of active sites. We contend that this will provide the greatest probability for successfully designing competent catalysts. From entries in the Catalytic Site Atlas,[86] Thornton et al. assembled the catalytic tool kit, a relatively short list of catalytic units which are believed to act synergistically to

10. The Diels−Alder Reaction between But-2-en-1,4-diol and 5-Ethoxy-4-methyloxazole. The Diels−Alder reaction of butendiol and 5-ethoxy-4-methyloxazole is a key step in the synthesis of Vitamin B6 or pyridoxine,[65] which is an important water-soluble member of the B complex vitamin group.[83] The uncatalyzed *endo*-pathway has a calculated activation free energy of 40.7 kcal/mol. The biological functionality utilized to catalyze this reaction is a Zn ion that is bound to three histidine side chains in a tetrahedral fashion. Zinc participates in a number of enzymatic processes and can be associated with a four or five coordinate complex having a distorted-tetrahedral or trigonal-bipyramidal geometry, respectively. Common ligating side chains include histidine, glutamate, aspartate, and cysteine with histidine being the most frequent.[84] The three histidine zinc-binding motif was adopted for the design of the catalytic site, mimicking the catalytic site of carbonic anhydrase (EC 4.2.1.1, PDB code: 1G54).[85] The fourth ligating position of the tetrahedral zinc complex can coordinate to the most basic site of the reacting substrates, namely, the nitrogen of the oxazole diene. The distances for the N-3 atoms of the histidine models and the N of diene to the zinc ion in the catalytic transition structure are all approximately 2 Å and arranged in a tetrahedral fashion, similar to the geometry observed in the carbonic anhydrase active site.[85] The zinc ion provides the capability for Lewis acid catalysis of Diels−Alder reaction by withdrawing electron density from diene-Zn^{+2} bond. The calculated activation barrier is 33.3 kcal/mol, 7.4 kcal/mol lower than that of the uncatalyzed reaction.

11. Hajos-Parrish-Eder-Sauer-Wiechert Reaction. The target reaction in water follows a similar pathway described before, that is, keto−enol tautomerization of the starting triketone catalyzed by hydroxide and hydronium. This step is endergonic by 16 kcal/mol. The intramolecular aldol addition of the enol nucleophile to the cyclic ketone yields the product with an overall activation barrier of 42 kcal/mol. The designed catalytic site based on DERA is predicted to lower the barrier by more than 33 kcal/mol over $\Delta G^{\ddagger}_{uncat}$, exhibiting essentially the same activation barrier than the other intramolecular aldol reaction described before (reaction 6), i.e., $\Delta G^{\ddagger}_{cat}$ 9.2 kcal/mol. This design presents a $N_{Lys}-N_{Lys}$ distance of 3.97 Å, which is compatible with the required pK_a shifting toward neutrality of the nucleophilic Lysine residue.

(83) Pauling, H.; Weimann, B. J. In *Ullmann's Encyclopedia of Industrial Chemistry*; VCH Verlagsgellschaft mbH: Weinheim, Germany, 1996; Vol. A27, pp 530−540.

(84) Auld, D. S. *BioMetals* **2001**, *14*, 271−313.

(85) Kim, C.-Y.; Chang, J. S.; Doyon, J. B.; Baird, T. T., Jr.; Fierke, C. A.; Jain, A.; Christianson, D. W. *J. Am. Chem. Soc.* **2000**, *122*, 12125−12134.

(86) Torrance, J. W.; Bartlett, G. J.; Porter, C. T.; Borkakoti, N.; Thornton, J. *J. Mol. Biol.* **2002**, *324*, 105−121.

catalyze a specific reaction,[87] for example, the ubiquitous Ser-His-Asp catalytic triad. Although the catalytic toolkit is an excellent start, it is not comprehensive. In order to most effectively apply naturalistic active sites to the design problem, it is imperative to assemble a database of catalytic units and corresponding reaction archetypes from all of the available enzymatic structure databases, such as the EzCatDB,[88] Brenda,[89] UniProt,[90] and MACiE.[91] Ongoing efforts in the Houk lab are aimed at collecting naturalistic catalytic units and applying quantum mechanical methods to determine the optimal arrangements in presence of an appropriate prototypical transition structure. The goal is to create a database of quantum mechanical catalytic units to be rapidly utilized to stabilize transition structures of novel reaction targets for enzyme design.

Our calculations based on cyclophilin yielded an interesting and testable prediction. The active arginine was calculated to interact strongly with a non-prolyl amide bond, decreasing the potential transition state stabilization. Replacing the arginine with a lysine reinforced the desired stabilization of the prolyl amide and dramatically increased the stability of the transition state. Experimentally characterizing this arginine to lysine mutation in cyclophilin could produce a more potent prolyl isomerase and serve to further validate the accuracy of this method. However, it must be noted that the predicted stabilization to the transition state in all theozyme models is likely overestimated by gas phase interactions with charged side-chain hydrogen bonds. In enzyme active sites, especially in deeply buried sites, charged residues will be stabilized via hydrogen-bonding interactions by adjacent side-chains to offset the desolvation penalty of buried polars; such interactions will consequently alter the hydrogen-bond donating or accepting ability of the catalytic side-chain accordingly. Our calculations demonstrate the catalytic potential of a specific constellation of catalytic residues applied to a reaction. Ultimately, however, the effectiveness of the designed active sites will only be realized though enzyme design efforts and site-directed mutagenesis experiments to validate the catalytic mechanism.

Supporting Information Available: Cartesian coordinates of all reported structures, as well as the ZPE and free energies, and uncatalyzed reaction coordinates for reactions 1−11. This material is available free of charge via the Internet at http://pubs.acs.org.

JO701974N

(87) Gutteridge, A.; Thornton, J. M. *Trends Biochem. Sci.* **2005**, *30*, 622−629.

(88) Nagano, N. *Nucleic Acid. Res.* **2005**, *33*, D407−D412.

(89) Schomburg, I.; Chang, A.; Edeling, C.; Gremse, M.; Heldt, C.; Huhn, G.; Schomburg, D. *Nucleic Acid. Res.* **2004**, *32*, D431−D433.

(90) Bairoch, A.; Apweiler, R.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L.-S. *Nucleic Acid. Res.* **2005**, *33*, D154−D159.

(91) Holliday, G. L.; Bartlett, G. J.; Almonacid, D. E.; O'Boyle, N. M.; Murray-Rust, P.; Thornton, J. M.; Mitchell, J. B. O. *Bioinformatics* **2005**, *21*, 4315−4316.