Computational Approaches: Reaction Trajectories, Structures, and Atomic Motions. Enzyme Reactions and Proficiency

Thomas C. Bruice*

Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106-9510

Received November 2, 2005

Contents

1. Introduction

1.1. General Statement

In part, the purpose of this report is to give the experimental enzymologist a view of what computational chemistry can offer. It is the author's opinion that computational procedures will soon become both common knowledge and an invaluable tool of the "experimental scientist". The review focuses on the importance of computational methods in understanding ground and transition geometries, as well as determining the reaction energies from the thermal motions of enzymatic reactions. Because of limited space, the discussion of many other important observations is left to recent reviews of this subject.[1−5] The importance of computational methods to enzymology is illustrated by selected examples emphasizing our own work.

Sole explanations for enzyme activity, such as tighter physical binding of the transition state compared to the substrate[6,7] or reduction in the entropy of activation by freezing the transition state (TS) motion,[8] should no longer be looked upon as the principal driving force in enzyme catalysis. While these features are of importance in certain enzymatic reactions, they are not generally so.[1] What follows is an examination of the latest panacean explanation of enzyme proficiency.

1.2. Covalent Intermediates and Enzyme Proficiency

The formation of covalent enzyme−substrate intermediates has been recently offered as an explanation for large enzyme proficiencies and is suggested to be the most important aspect of enzymatic catalysis—"Beyond the Pauling Paradigm".[9,10]

There are two principal ways to describe the catalytic ability of an enzyme. One is to compare the rate constant of one enzyme to that of another. The other is to compare the proficiency of one enzyme to that of another. By convention, the proficiency of an enzyme is defined as the rate constant (k_{cat}/K_m) for the enzymatic reaction divided by the rate

* Corresponding author e-mail: tcbruice@chem.ucsb.edu.

10.1021/cr050283j CCC: $59.00
Published on Web 07/15/2006

Thomas C. Bruice received his B.A. and Ph.D. degrees at the University of Southern California and served as a National Research Council postdoctoral fellow in chemistry at UCLA. He has served on the faculties of Yale School of Medicine, Johns Hopkins School of Medicine, and Cornell University and is presently Research Professor at the University of California, Santa Barbara. He has published over 575 papers, is a member of the National Academy of Science (USA) and the American Academy of Arts and Science, and is a Fellow of the Royal Society of Chemistry. His major awards include the Chemistry Medal of the National Academy of Science and awards of the American Chemical Society which include the Bader Award in Bioorganic Chemistry, the James Flack Norris Award in Physical Organic Chemistry, and the Repligen Award in Biochemistry.

Table 1. Values of First Order Rate Constants for Enzymatic Catalysis Divided by the Rate Constant in Water at pH 7.0[a]

	k_{cat}/k_o
OMP decarboxylase*	1.4×10^{17}
staphylococcal nuclease	5.6×10^{14}
adenosine deaminase*	2.1×10^{12}
AMP nucleosidase	6.0×10^{12}
cytidine deaminase	1.2×10^{12}
phosphotriesterase	2.8×10^{11}
ketosteroid isomerase*	3.9×10^{11}
triosephosphate isomerase	1.0×10^{9}
cyclophilin, human	4.6×10^{5}

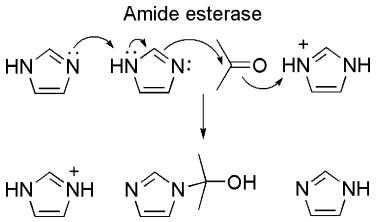
[a] An asterisk indicates an enzymatic reaction that does not involve covalent intermediates.

constant (k_o) for conversion of substrate to product, in water at pH 7.0. Since k_{cat}/K_M is in units of s^{-1} M^{-1} and k_o is in units of s^{-1}, proficiency is in units of molarity (M^{-1}). This is awkward and suggests that k_{cat}/k_o is a better comparison. For simplicity this definition is applied here. Values[11] of k_{cat}/k_o are provided for a series of enzymes in Table 1.

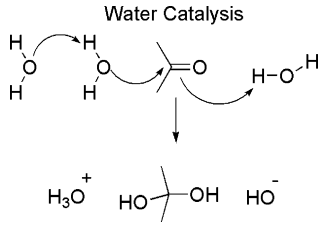
If the covalent intermediate proposal has meaning, the mechanisms of the most reactive enzymes should involve covalent intermediates. In a collection of 24 enzymes,[11] values of k_{cat}/K_M are between 10^5 and 10^9 M^{-1} s^{-1}. Enzymes that do not involve the formation of a covalent intermediate are among the most reactive. A few examples are fumarase[12] ($k_{cat}/K_M > 10^9$ M^{-1} s^{-1}), OMP decarboxylase[13−15] ($\sim 10^8$ M^{-1} s^{-1}), and staphylococcal nuclease[16] ($k_{cat}/K_M > 10^7$ M^{-1} s^{-1}). It is apparent that there are enzymes that catalyze very rapid reactions without forming covalent intermediates.

Because enzyme proficiency is the rate constant for the enzyme divided by the rate constant in water, proficiency becomes greater when k_{cat} becomes greater or when k_o becomes smaller. The values of enzymatic rate constants are 10^4-10^6 s^{-1} for a group of enzymes whose proficiencies range from 10^7 to 10^{20}. The much greater difference in enzyme proficiencies must relate not to the enzymatic reaction but to the reaction in water. Water is a very poor

Scheme 1



Scheme 2



nucleophile, and at pH 7.0 the concentrations of acid (H_3O^+) and base (HO^-) are 10^{-7} M. The more complex the enzymatic reaction, the more roles water must play in the water reaction. As an example, let us compare the hypothetical "amidesterase" of the tri-histidine family and its water counterpart (Schemes 1 and 2). In Scheme 1, nucleophilic addition of an imidazole (HIm) to the amide/ester bond is general base catalyzed by another HIm and general acid catalyzed by an imidizolium ion ($HImH^+$). In this enzyme, the pK_a of the $HImH^+$ is 6.7; therefore, both HIm and $HImH^+$ are readily available at pH 7.0. Replacing both HIm and $HImH^+$ with water molecules (Scheme 2) provides a reaction in which water molecules play the role of both the acid and base catalysts as well as the nucleophile. The water reaction suffers in that water is not as good a base as imidazole (HIm) and not as strong an acid as the imidazolium ion ($HImH^+$). If the Brønsted α and β constants are between 0.5 and 0.7, then k_{cat} will exceed k_o by 10^{10} to 10^{17} (comparable to the average enzyme proficiency). An alternate scheme for the water reaction would have H_3O^+ and HO^- as acid and base catalysts. Since these two species would both be present at $\sim 10^{-7}$ M at pH 7.0, a trimolecular complex of HO^- and H_3O^+ and substrate would, for all practical purposes, never form. In the case of the serine proteases, water molecules would replace Asp-CO_2^-, HIm, and Ser-OH; thus, the enzymatic reaction would be compared to the hydrolysis of a peptide bond in water at pH 7.0 ($k_o \sim 10^{-10}$ s^{-1}).[17]

Complex enzymatic reactions may have proficiencies much greater than simpler enzymatic reactions because water is a poor nucleophile with little acid or base property. The more complex the enzymatic reaction, the more roles water must play badly. Comparison of enzyme proficiencies is not to be confused with comparison of enzymatic rate constants. The numerical rate constants for enzymatic reactions do not depend on whether covalent intermediates are involved. Consider that some enzymatic hydrolysis of esters and amides can involve tetrahedral- and acyl-enzyme intermediates. Much the same can be said about phosphatase enzymes. Such reactions are not known to be exceptionally rapid. These reactions were selected, during the development of the enzymes, as the only way to get to the product, not to provide large rate constants. In short, covalency is a requirement of mechanism, and we must remember that each step in a mechanism has a TS whose formation may require electrostatic assistance and/or acid/base catalysis and is

subject to the stabilizations referred to by Pauling. There are no single explanations for enzyme activity.[1,2,18]

2. Use of Computational Methods in Enzymology

2.1. From Static Crystal Structures to Dynamic Structures Using Molecular Dynamics (MD) Simulation

2.1.1. Explicit Water Solvent Condition

The most common means employed to obtain total structures of enzymes or enzyme complexes is X-ray crystallography. The crystal structure provides time- and crystal lattice-averaged structures that are to some degree distorted by the crystal packing effects. The lattice-averaged structures account for protein conformations that are present for more than 20% of the time at a moderate resolution. The reactive conformation of an enzyme substrate complex (E·S) present at only 15% of the time is invisible in all but an ultrahigh-resolution structure. Coordinates of the hydrogen atoms are usually not provided in crystal coordinates, except at the very highest resolutions. The missing hydrogen atoms are properly placed in MD simulated structures using energy placement protocols. A great advantage of molecular dynamics (MD) simulations is that a crystal coordinate of an enzyme inhibitor complex (E·I) can be manipulated to provide enzyme substrate (E·S), and if the structure of the transition state (TS) is known, the dynamic structure of E·TS can also be acquired.

Modern MD programs (such as CHARMM[19] and AMBER[20]) depend on force fields and can simulate the atomic-level motions of an enzyme solvated in an accurately described water environment (TIP3P)[21] under well-defined conditions of temperature and pressure. A few words follow concerning the computational setup for such MD studies and also for the use of quantum mechanics with molecular mechanics (QM/MM) in determining the E·TS structure and the energies of states. The reader is directed to the Methods section of any number of research papers for details.[22]

MD simulations of enzyme structures and motions are often calculated[23] using the method of Periodic Boundary condition with the Particle Mesh Ewald algorithm (PME). PME treatment is particularly important for accurate description of long-range electrostatic interactions. In this procedure, the enzyme substrate complex is immersed in a pre-equilibrated box of TIP3P water. Since this box infinitely repeats itself in all three dimensions, one can simulate an infinite solvent environment in which the enzymes complex is periodically placed.

MD simulation of enzyme structure and dynamics can also be carried out by use of Stochastic Boundary conditions, where only a small region within a certain sphere of the enzyme is included in the simulation, increasing the computational efficiency. This method is very useful if one is interested in the dynamics of the enzyme only within a limited region, such as an active site. Reactive components in the active site can be treated by quantum mechanics while the rest of the molecule is subject to a molecular mechanics force field (QM/MM). Use of QM/MM allows the calculation of reaction trajectories providing E·TS structures and energies of reaction.[24,25]

Chart 1 shows the computational structure of a dimeric enzyme substrate complex in which one monomer is fully solvated in an equilibrated TIP3P water sphere of radius 42 Å. This sphere is partitioned into the three regions: QM region, a reaction region ($r < 40$ Å using the substrate as the origin), and a buffer region (40 Å $< r <$ 42 Å). The QM region contains the reactive components of enzyme and substrate and is treated by QM methods. The reaction region is treated with Newtonian dynamics, and the buffer region is treated with Langevin dynamics. The region beyond 42 Å of the origin is the reservoir region, which functions as a heat bath to keep the system in equilibrium. A spherical boundary potential centered at the origin is employed to maintain the correct average distribution of water molecules. (The convenient E·S complex in Chart 1 is the dimer of the ortho-quinone PQQ quinoprotein methanol dehydrogenase.)[26]

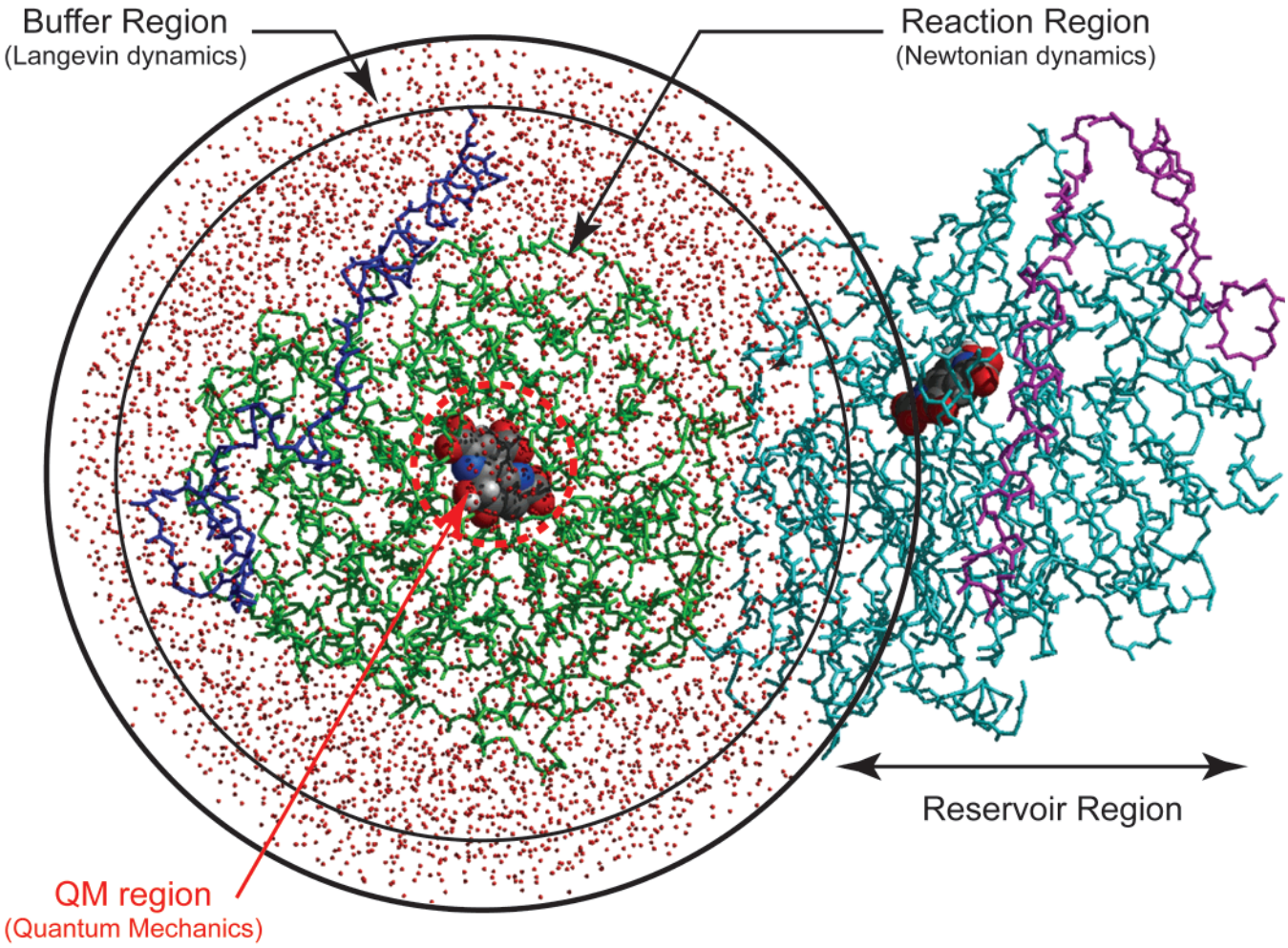
2.1.2. Implicit Water Solvent Condition

MD simulations have traditionally used a large number of explicitly treated solvent molecules. These represent the most detailed methods for the study of the effect of solvent on complex biomolecules. Unfortunately, in explicit solvent simulations, the major portion of the computational time is spent on integrating the trajectories of thousands of solvent molecules, when one is interested only in the motions of the protein atoms. Moreover, statistical convergence is an important issue because the net influence of solvation results from an averaging over a large number of configurations. An alternative approach consists of incorporating the influence of the solvent implicitly. Implicit solvent models are computationally efficient compared to the explicit solvent models, and they are straightforward to interpret, as the water degrees of freedom are absent. The implicit solvent model describes the instantaneous solvent dielectric response and, therefore, eliminates the need for lengthy solvent equilibration steps. Also, the possible artifacts of the replica interactions, observed in periodic boundary explicit solvent simulations, can be avoided, since the solvent is considered as one continuous medium with infinite volume. The implicit solvent models have also been used to model nonaqueous environments, such as membranes.

However, use of implicit solvent is sometimes limited when the properties of the water molecules are substantially different from those of the bulk solvent, for example, when there is an explicit water channel to the active site and when water is one of the reactants (e.g., hydrolysis type reactions), or when there are solvent mediated interactions during protein folding.[27] Also, some studies have shown that the accuracy of an implicit solvent model is not yet comparable to that of experiments or explicit solvent calculations.[28] Recently, a new hybrid explicit/implicit solvent method has been developed which partitions the simulation system into two regions, separated by an elastics boundary.[29] This method models explicitly the hydration of the solute by either a layer or sphere of water molecules, and then generalized Born (GB) theory is used to treat the bulk continuum solvent outside the explicit simulation volume. To make this method computationally efficient, a multigrid method is used to evaluate the pairwise electrostatic and GB terms. It will be interesting to see whether this new method will bridge the gap between the implicit and explicit solvent simulations and capture the salient features of both methods—accuracy and speed.

To set up an MD simulation based on the implicit solvation model, one requires a set of atomic radii, which is an additional set of input parameters compared to the explicit solvent case. A number of implicit solvent models have been

Chart 1



developed and treat the effect of the solvent as an average potential of mean force (PMF) acting on the solute. The implicit solvent model based on a finite difference solution of the Poisson−Boltzmann (PB) theory provides a rigorous theoretical framework and captures the polar component of the free energy of solvation for a given biomolecule quite well.[30] However, it is a computationally expensive method and, therefore, has limited application in MD simulations. This approach has been employed in a practical fashion when calculating the following: various properties of a biomolecule's fixed conformation in order to obtain the electrostatic solvation free energy, shifts in pK_a for specific residues, and protein−protein interactions in solution. Several programs are available for computing the electrostatic potential using this approach, e.g., DelPhi,[31] UHBD,[32] and the PBEQ module[33] incorporated in the simulation program CHARMM.[34]

Another simpler and faster implicit continuum model based on the PB equation is the so-called generalized Born (GB) model.[35,36] In fact, a very good agreement between the GB and the PB models can be achieved if the effective Born radii match those computed exactly using the PB approach. Therefore, one can improve the accuracy of the GB model by improving the way the effective radii are computed. Due to its relative simplicity and computational efficiency, this methodology has become especially popular in molecular dynamics simulations, compared to the more standard numerical solution of the PB equation.[29,37,38]

2.2. Reaction Coordinates and the Structures of States from QM/MM Calculations

2.2.1. Ab Initio and Semiempirical Methods

Due to the number of atoms that comprise the QM region and the necessity for long computational times, high-level ab initio methods are not used in QM/MM computations for large biosystems. Until quite recently, the simplistic semiempirical method of AM1 has been a popular choice (as has PM3). Activation energies of reaction calculated by AM1/MM are generally >10 kcal/mol in excess of the experimental values.[39,40] The AM1 calculations most often have to be corrected for each reaction studied.[41] Alternatively, AM1/MM calculations provide the E·TS structure while the energy of activation is separately determined from the AM1 TS using single point high-level calculations.

A step forward has been the introduction of SCCDFTB (self-consistent-charge-density-function-tight-binding) and SCCDFTB/MM.[42,43] The computational speed of this density function method is comparable to that of AM1, and the results compare with those produced by the B3LYP (Becke-style-3-parameter density-functional theory with the Lee−Yang−Parr correlation functional) quantum mechanics method.[43] Due to its efficiency, the SCCDFTB/MM procedure is often employed to obtain one- or two-dimensional energies or free energy surfaces, from which the E·TS structure can be determined.[44−47]

There are various approaches to the solution of the Schrodinger equation.[48] Hartree−Fock (HF) is the simplest of the wave function based approaches but is to be avoided due to its complete neglect of electron correlation. Electron correlation must be taken into account to treat mechanisms involving proton-transfer processes. A sophisticated means involves the use of MP2 (Møller−Plesset second order of perturbation theory). The use of MP2 is very time-consuming, such that it is applied for single point computation of states (starting, intermediate, and transition states). Examples are the study of class A β-lactamase acylation,[49] in which optimization using HF/6-31+G* was followed by single point calculations with MP2/6-31+G*. A similar approach has been used in the study of the mechanism of acetylcholine esterase.[50] One must be aware of the shortcomings in methods. For example, it has been shown that MP2 calculations overestimate the stability of the transition state for C−S bond formation.[51]

Kinetic isotope effect (KIE) provides an experimental ground to evaluate TS structures. From KIE, one can deduce information on the partial bond lengths of the TS and subsequent atomic charges for the reacting moiety of the TS. Using this key information, an MD simulation can be performed to determine the more precise conformations of the E·TS structure.

For reactions in which values of experimental rate constants are independent of solvent character, the TS structure obtained from the gas phase QM calculation may be directly employed in MD simulations to determine the E·TS conformation. In this case, the forming and breaking bonds, as well as the angles at which these bonds are formed, are fixed, but the rest of the molecule is free to relax according to the enzyme motion. The advantage of this method is that QM calculations are much simpler in the gas phase, under which high-level calculations can be employed to determine the TS structure.

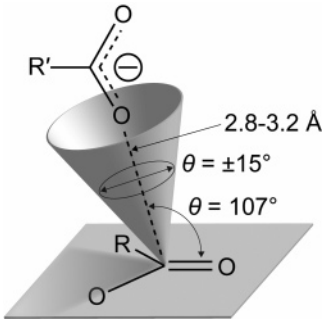
2.2.2. Empirical Valence Bond (EVB) Method

The empirical valence bond (EVB) method[24,52,53] is another QM/MM procedure (See refs 25 and 53 for the origins of this method and for recent versions). In the EVB method, reactions are described by mixing diabatic states that correspond to classical valence-bond (VB) structures, which describe the reactant intermediate(s) and product states. The potential energies of these states are represented by classical MM-like force fields, which are then mixed by empirical off-diagonal terms to reproduce the overall ground state surface.

The EVB method satisfies some of the main requirements for reliable studies of enzymatic reactions. This includes the facilitation of proper configurational sampling and converging free energy calculations as well as the ability to capture correctly the linear relationship between activation free energies and reaction energies (LFER) observed in many important reactions.[52] Calibrating EVB surfaces using *ab initio* calculations and experimental information was found to provide reliable potential surfaces.

The EVB method has been widely adapted for studies of enzyme mechanisms (for a few examples, see refs 54 and 55). Importantly, the method provides the difference between the energy functionals in the enzyme and in the solution reaction rather than only the first principal evaluation of the absolute energies.[52,53] Several very closely related versions have been put forward with basically the same ingredients as those in the EVB method.[25,53]

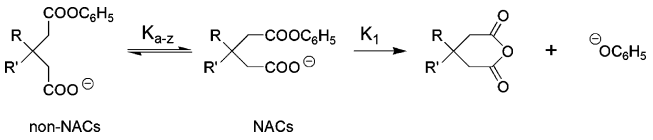
Chart 2



2.3. A Useful Tool but Not a State: Near Attack Conformers (NACs)

In discussions of reaction mechanisms, chemists take into account not only transition state structures and structures along the reaction coordinate but also the conformations of ground state structures. A given ground state conformation is the average of a Boltzmann distribution of conformers. This Boltzmann distribution is routinely available by MD simulations of motions. Within this distribution, there are those conformations which closely resemble the transition state. The transition state can only be reached through these near attack conformers (NACs). The portion of the Boltzmann distribution that is NAC provides the free energy of NAC formation. NACs have been arbitrarily defined as having reacting atoms at van der Waals distance and angles ±15−20° of the angle of the forming bond in the TS.[56−58] The free energy of a conformer has been shown not to change much when the angle of approach deviates by ±15° from the bonding angle in the TS for nucleophilic attack on sp^3 or sp^2 carbons separated by 3.2 Å.[59] Because the NAC does not necessarily imply a well-defined state and it involves several geometrical parameters to define, it is difficult to measure the molarity of NACs by experimental means.

The usefulness of the NAC concept (Chart 2) was first shown in organic models involving cyclic anhydride formation in intramolecular reactions (eq 1) of dicarboxylic acid



monoesters.[56−58] Figure 1 shows a plot of the log of the Boltzmann probabilities for NAC formation (computational) vs the log of the relative rate constants, k_{rel} (experimental). A line of slope 1.0 is drawn connecting glutaric, succinic, and 3,6-endoxo-Δ^4-tetrahydrophthalate monoesters. The slope of 1.0 establishes that the differences in rates of reactions relate quantitatively to the probability for NAC formation and that the rate constant for NAC → anhydride + p-$BrC_6H_4O^-$ is very close to being the same for all dicarboxylic monoesters. Transition state structures were shown [RHF/6-31+G(d)] to be identical for anhydride formation from glutaric, succinic, and 3,6-endoxo-Δ^4-tetrahydrophthalate monoesters. In this fashion, the use of NAC structures to evaluate the importance of ground state conformations in well-characterized chemical reactions was established. This concept can be extended to enzymatic reactions where the E·NAC population among E·S structures

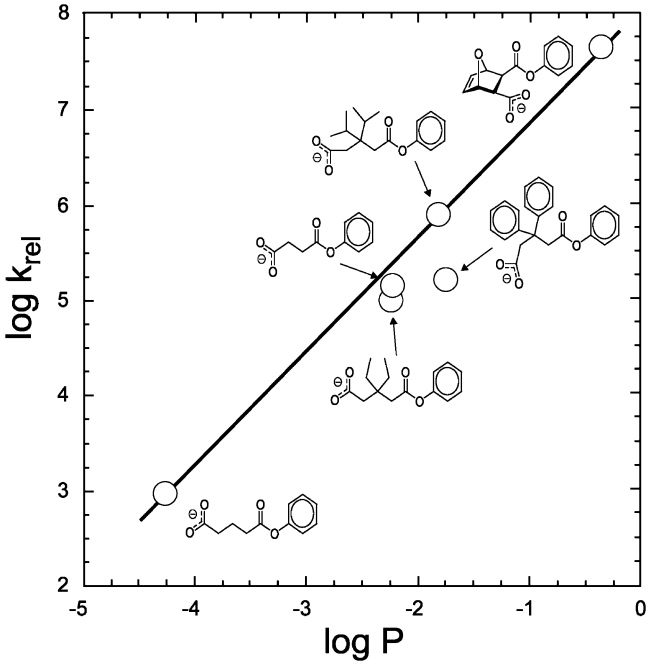


Figure 1. Plot of the log of the relative rate constant vs the log of the Boltzmann probability for formation of NAC structures.

can be related to the rate constant. To date, the free energies for NAC formation in the E·S complexes have been calculated for a number of single substrate enzymatic reactions (chorismate mutase,[60] catechol O-methyltransferase,[61] Hhal methyltransferase,[62] formate dehydrogenase,[63] dihydrofolate reductase,[64] 1,4-β-xylanase,[65] horse liver alcohol dehydrogenase,[66] engineered mutant of papain (a nitrile hydratase),[67] soluble epoxide hydrolase,[68] haloalkane dehalogenase,[69] hyperthermophilic indole-3-glycerol phosphate synthase,[70] and Ser-Ser-Lys catalytic triad peptide amidase[71]). These studies have been carried out either by analysis of the Boltzmann distribution of conformers generated in long-term MD simulations of E·S or by direct calculation of the free energy profile of the E·S complex. From the definition of NAC, the free energy of reaction ($\Delta G^{\ddagger}$) is the sum of the standard free energy for NAC formation (ΔG°_{NAC}) and the free energy for conversion of the NAC to the TS (ΔG_{TS}), as in eq 2. The values of ΔG°_{NAC} for the single substrate

$$(\Delta G^{\ddagger}) = (\Delta G^{\circ}_{NAC}) + (\Delta G_{TS}) \qquad (2)$$

enzymatic reactions studied vary from 0.3 to 1.5 kcal/mol. Thus, for these enzymatic reactions, ΔG°_{NAC} does not contribute much to the free energies of reaction ($\Delta G^{\ddagger}$). ΔG°_{NAC} becomes important in determining $\Delta G^{\ddagger}$ under nonoptimal conditions of temperature and pH and, also, with certain mutants. The value of (ΔG°_{NAC}) for the reaction in water can be sizable and contribute importantly to $\Delta G^{\ddagger}$. The difference in free energies of reaction in water and enzyme ($\Delta\Delta G^{\ddagger} = RT \ln(k_{cat}/k_o)$) can be partitioned into $\Delta\Delta G^{\circ}_{NAC}$ and $\Delta\Delta G_{TS}$. These terms represent the kinetic advantage of the enzymatic reaction over the water reaction by NAC formation and conversion of the NAC to the TS. The kinetic advantage of the enzymatic reaction due to favorable NAC formation is of importance in determining enzyme proficiency.

To this date, only single substrate reactions have been examined for the importance of NAC formation in determining enzyme proficiency. In more complex multisubstrate

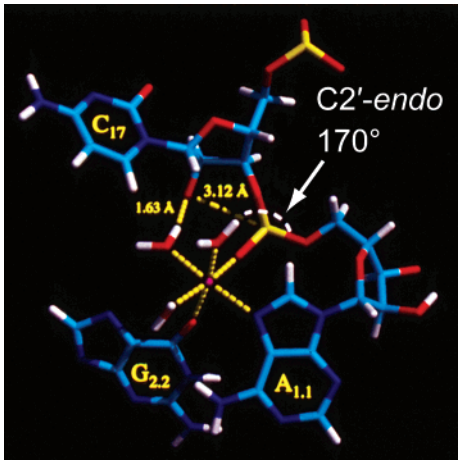


Figure 2. NAC formation in the hammerhead ribozyme. An inner sphere water molecule ligated to Mg^{2+} is hydrogen bonded to the 2′-hydroxyl with a distance of 1.63 Å. The distance and angle for attack of O2 on phosphate are 3.12 Å and 170°.

enzyme reactions, NAC formation may be even more important. Wolfenden and co-workers have determined the proficiency of the ribosome in peptide bond formation. They conclude that the (2×10^7)-fold rate enhancement is due mainly to positioning the substrates within the active site, rather than due to conventional chemical catalysis.[72] Using the Gaussian network model and the anisotropic network model of protein dynamics, Tobi and Bahar showed that a protein ligand samples an ensemble of conformations at protein equilibrium conditions and that the ligand binds selectively to a complimentary conformation.[73] The study is said to "lend support to the concept that proteins, in their native conformation, are predisposed to undergo conformational fluctuations that are relevant, or even required for, their biological functions".

3. A Few Examples of What Can Be Observed by MD Simulations That Are Not Available by Experimental Methods

3.1. Specific Self Cleavage of Hammerhead RNA

Despite the numerous experiments performed on many different hammerhead motifs, a number of questions still remain, such as the number of Mg^{2+} or Mn^{2+} essential in the cleavage reaction and the precise details of the mechanism of Mg^{2+} or Mn^{2+}catalysis of self-cleavage of the hammerhead RNA at a specific site.[74−78] MD simulations were carried out,[79] employing AMBER force fields,[80] using the coordinates of freeze trapped hammerhead RNA[81] with two Mg^2 ions. One Mg^{2+} was associated with the *pro-R*-phosphate oxygen of the leaving nucleotide, as shown to be required by experimental methods. The second Mg^{2+}, placed at site 3, plays a structural role. The main question of the mechanistic problem is how does the RNA structure allow an in-line nucleophilic attack of the ribose 2′-O^- upon phosphate P?

Analysis of the thermally derived conformations at ambient temperature provided the NAC structure of Figure 2, which was present at 18% of the 1 ns simulation time.[79] The MD calculation was repeated after changing the seed value for the random number generator to assign initial velocities. The very same NAC conformation was found to be present at 8% of the 1 ns simulation time. Neither X-ray nor NMR

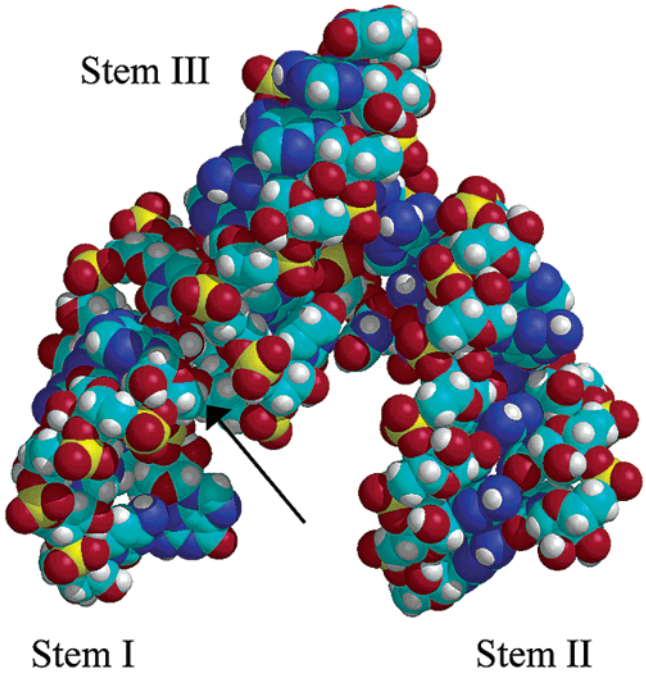
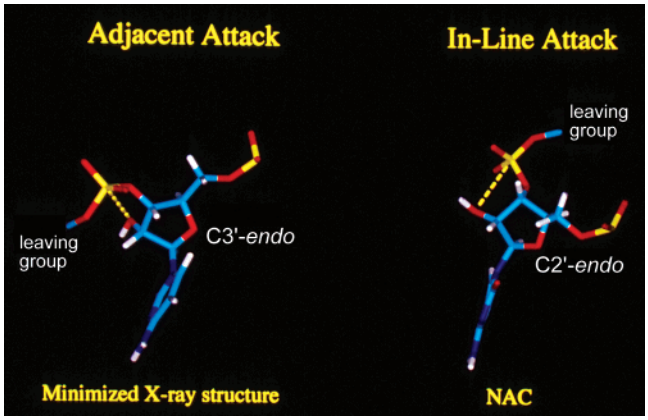


Figure 3. Molecular dynamic presentation of the X-ray coordinates for the ground state of the hammerhead ribozyme. The arrow points to the position of bond breaking. (Reprinted with permission from ref 83. Copyright 2003 American Chemical Society.)

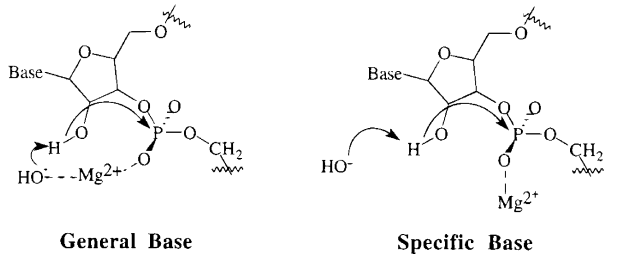
Chart 3



methods would be expected to find this reactive conformation. The motions leading to the catalytically competent conformations are traceable. Stems I and II (Figure 3) move toward each other, prompting torsional rotations in the phosphodiester backbone near the active site, which leads to the unstacking of residues C17 and G5 from A6. A ribose sugar pucker flip from C3′-*endo* to C2′-*endo* (Chart 3) in the nucleotide containing the 2′-hydroxyl nucleophile takes place. Next, C17 restacks with A6 to form the NAC structure of Figure 2. Examination of Chart 3 shows that the ribose sugar pucker flip places the 2′-O in position for the in-line S_N2(P) nucleophilic addition of 2′-O^- onto the phosphate P. It is known that the log k_{rate} vs pH profile is of slope 1.0, which is in accord with the specific or general base catalysis of Chart 4. The general base mechanism is preferred on the basis that at higher pH the profile slope becomes zero.[82]

The dynamics of the Mg^{2+} catalysis of self-cleavage of the hammerhead RNA has been investigated[83] by density functional theory with the B3LYP/6-31G(*d*,*p*) basis set. Solvent was modeled by COSMO[84] using a polarizable continuum to represent water. The results of the computations

Chart 4



are presented in Scheme 3. The reaction is initiated by a general base-catalyzed attack of (what would be on ribose) 2′-OH on phosphate P to form a pentacoordinate intermediate. The rate-limiting step is the breakup of the intermediate to form products. The leaving group is assisted by protonation by magnesium bound water. The reaction coordinate is presented in Figure 4.[83] These calculations provide structures

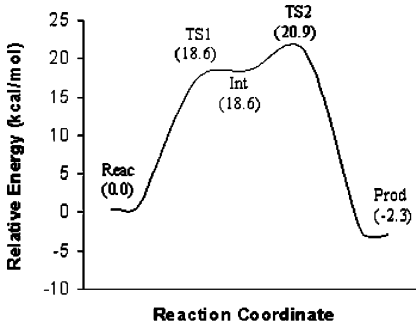


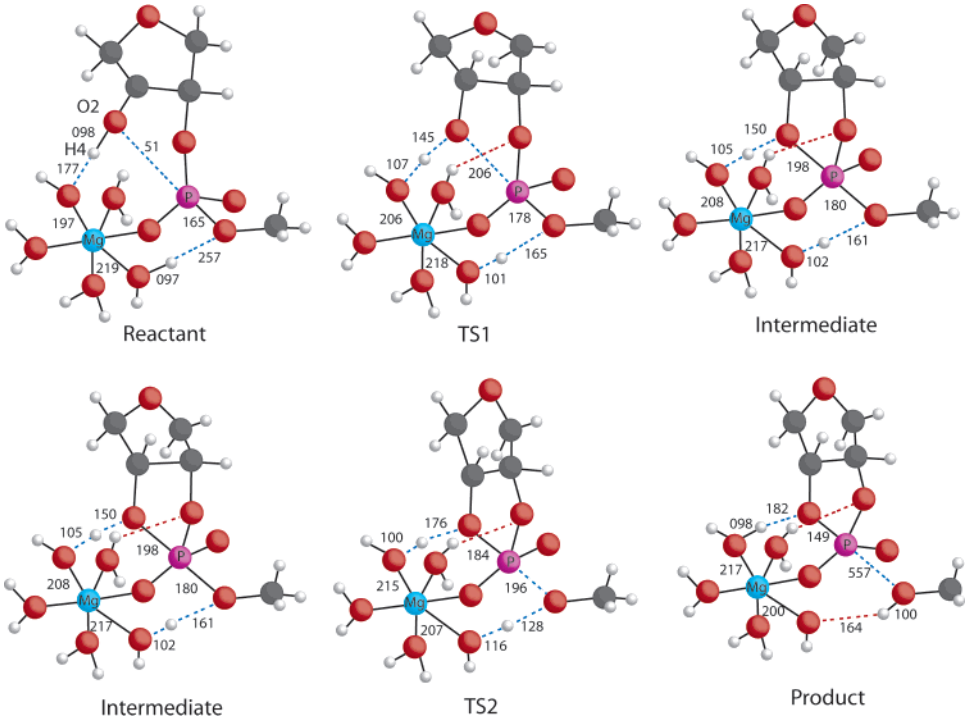
Figure 4. Relative energies, including solvation effects, for the nucleophilic addition of 2′-O to phosphate and conversion of intermediate to product. (Reprinted with permission from ref 83. Copyright 2003 American Chemical Society.)

and final energies consistent with experimental results and indicate that, at the present level of theory, the reaction is stepwise. In the hands of an experimentalist, the reaction would be considered as concerted because experimental methods cannot place the first transition state. The possibility of two metal ions participating in the cleavage reaction has not been eliminated. Explanations of important salt effects are possible, though they are beyond the scope of this review.

3.2. Modeling of Missing Loop and Structural Clarification of Unobservable High B-Factor Portions of a Crystal Lattice

Modeling can be accomplished by use of long-term MD simulations. Often, regions of protein, with high B-factor, are found missing from the crystal structures. Sometimes these missing loops are part of the active site and should not be ignored. For example, with the aid of the MD simulations, the positioning of an important flexible loop at the active site of inosine-adenosine-guanosine nucleoside hydrolase[85] was shown to be important for structuring of the active site. The MD study of orotidine-5′-monophosphate decarboxylase in the ground state and the intermediate state[14] established the 203−218 loop changes from an unstructured form to a ordered β-hairpin, stabilizing the intermediate state and the TS for its formation. In certain enzymatic reactions, E + S ⇔ E·S results in structural strain in the enzyme portion of E or S (or both) that is relieved on E·S → E·TS (see chalcone isomerase, *vide infra*).

Scheme 3



3.3. Determination of Important Protonation States

Association and dissociation of proton acids are invisible in X-ray structures at all but the highest resolution. The protonation state of titratable groups (Asp/Glu-CO_2H, Tyr-OH, His-ImH^+, Lys-NH_3^+) is important to both the structure and function of biochemical systems, and there has been much activity in the design of methods for the determination of pK_a. If the pK_a values of catalytic site functional groups are known, the experimentalist can apply this knowledge to the interpretation of the dependence of k_{cat} on pH. The majority of methods designed for biochemical systems are dependent upon classical electrostatic calculations by solving the Poisson−Boltzmann equation.[85] Important features to define are the protein dielectric constant (ϵ) and boundary. Quantum chemistry with a dielectric continuum,[14] and hybrid QM/MM/dielectric calculations[86] have been applied successfully.[87,88] Antosiewicz et al.[89] calculated 60 pK_a values in 7 proteins and found $\epsilon = 20$ provided values close to the experimental value. Schutz and Warshel[90] have presented arguments in favor of the use of two types of dielectric constants dependent upon electrostatic charges and permanent dipoles. Procedures based upon MD and free energy perturbation are offered.[11,12] Three recent studies[35,91,92] are of particular interest. Case and co-workers[35] computed pK_a shifts, relative to a model in solution using both explicit solvent and the general Born model.[91] Li and Cui[92] have applied the free energy perturbation approach to calculate pK_a with QM/MM potentials and periodic boundary conditions, with and without including long-range electrostatics explicitly in solution simulations. Values of computed pK_a compared favorably with experimental values.

Another means of determining if a substrate, inhibitor, or enzyme functional group is protonated is by testing the stability of the system during the MD simulation of the structures of both protonated and unprotonated species starting with the crystal coordinates. An interesting example

Chart 5

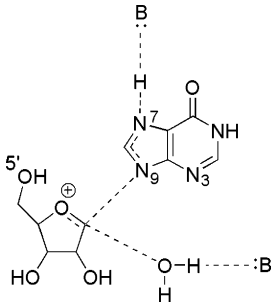
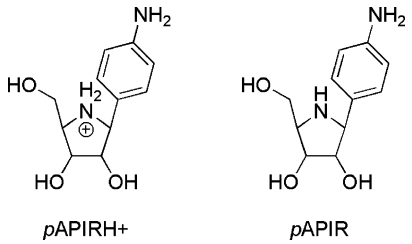


Chart 6

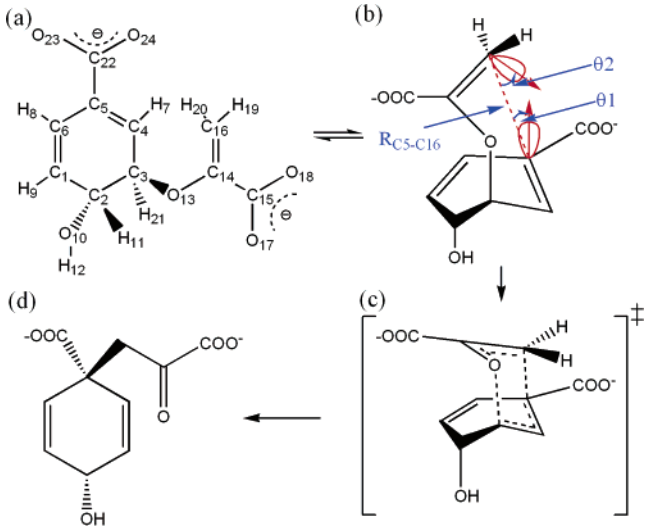


involves a study of trypanosomal inosine-uridine nucleoside hydrolase (IU-NH). Using multiple kinetic isotope effects, Horenstein and Schramm showed the chemical mechanism to be S_N1-like, which proceeds via an exploded ribooxocarbenium cation TS (Chart 5).[93]

They identified *p*-aminophenyliminoribitol (*p*APIR) to be a potent inhibitor of IU-NH from *Crithidia fasciculata*. The cyclic amine function of the iminoribitol ring (pK_a 6.5) can be either protonated (*p*APIRH^+) or unprotonated (*p*APIR) (Chart 6). The *p*APIRH^+ resembles the charge and geometry of the ribooxocarbenium cation transition state and is looked upon as a transition state analogue inhibitor.[94]

It was known that the inhibitor was associated with the enzyme as *p*APIR rather than *p*APIRH^+. Following formation of this complex, there followed a slow reorientation

Scheme 4



reaction providing the greatly inhibited enzyme species. It was not known if this species had the composition of E·*p*APIR or E·*p*APIRH$^+$. The *p*APIRH$^+$ species was established as the final resident of the active site using *ab initio* calculations and molecular dynamics simulations.[95] Nanosecond MD simulations showed that the structure of the complex of neutral E·*p*APIR did not resemble the X-ray structure well. However, the computed structure for the E·*p*APIRH$^+$ was shown to closely resemble the X-ray structure.

4. Important Lessons Learned from Simple Systems

4.1. Contributions of Ground State Structure and "Transition State Stabilization"

According to general use, the term "transition state stabilization" includes a number of features, which add up to decreasing the free energy of activation. These are the following: the complementary recognition of shape and charges of TS, as suggested by Pauling;[6,7] general base and general acid catalysis; certain instances of metal ion ligation to leaving groups; and release of bound water and release of strain when E·S → E·TS.

Aside from "transition state stabilization", enzymes control the conformations of the E·S ground state. This is a very important feature in the determination of the rate of an enzymatic reaction. It has been pointed out[2] that the separation of TS stabilization in E·TS and formation of reactive conformations in E·S is best understood when investigating enzymatic reactions which involve an intramolecular reaction of the substrate without formation of covalent intermediates. Multiple substrate calculations are very costly in computer time, and the question of enzyme stabilization of TS is complicated by covalent intermediate formation.

4.1.1. Chorismate → Prephenate Reaction in Water, Chorismate Mutases, Mutants, and "Catalytic" Antibody

The ratio of k_{cat} for the *E. coli* chorismate mutase (EcCM) catalysis compared to k_o for the reaction in water is 10^6.[96] The intramolecular conversion of chorismate → prephenate

Chart 7

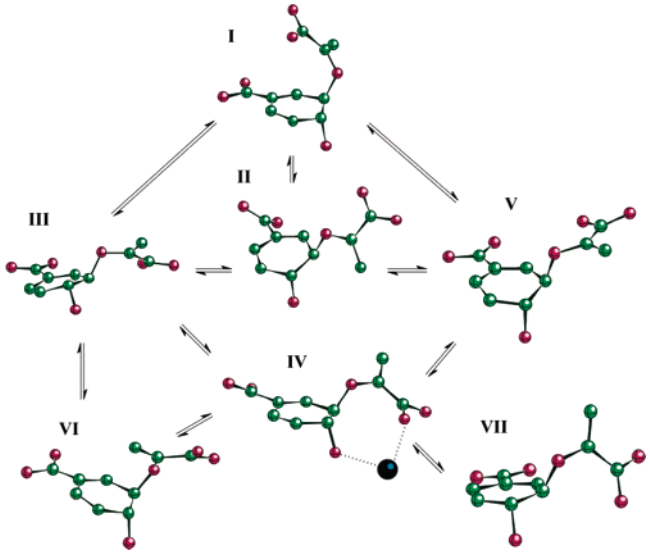
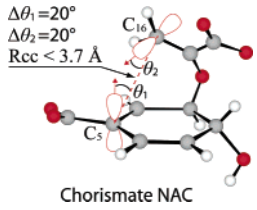


Chart 8



(Scheme 4) takes place only from a diaxial conformation from which the NAC structure is formed (eq 3). The seven

$$\text{diequatorial} \leftrightarrow \text{diaxial conformer I} \leftrightarrow \text{NAC}-[\text{TS}]\rightarrow\text{prephenate} \quad (3)$$

diaxial conformers that exist in water (Chart 7) and the rate constants for their exchange of structures were determined by long term (33 ns) MD simulations in TIP3P water.[69] In water, conformer I was shown to represent 0.0007% of the total diaxial conformers. Starting with conformer I, a turning motion of the $\theta 1$ dihedral angle forms the NAC (Chart 8).

The free energies for NAC formation in water (ΔG°_{NAC}) calculated from results of both MD simulations[60] and free energy simulation[34,97,98] agree: 8.4 and 8.2 kcal/mol, respectively. The values of ΔG°_{NAC} at the active site of the EcCM, determined by these two methods, were also in agreement (~0.5 kcal/mol).

A powerful tool of physical organic chemistry is the linear free energy plot. The values of free energies for preequilibrium NAC formation (ΔG°_{NAC}) and free energies of activation ($\Delta G^\ddagger$) for the chorismate → prephenate reactions were determined in the following six systems: *E. coli* (EcCM) and *B. subtilis* (BsCM) wild-type enzymes, the Glu52Ala mutant of EcCM (E52A), the Arg90Citruline mutant of BsCM (R90Cit), the catalytic antibody IF7, and water.[99] The values of $\Delta G^\ddagger$ were determined from experimental rate constants. The values of ΔG°_{NAC} were determined by use of thermodynamic integration (TI),[69] a method for free energy simulation, as follows: the free energy of chorismate at every 5−10° along the dihedral angle (dih1= C4−C3−O13−C14) was calculated from the 200−300 ps of production MD with the fixed dih1. NAC is formed in the range dih1 = 50−80°. Thus, G°_{NAC} is the summation of free energy cost for rotating dih1 to reach dih = 50−80°

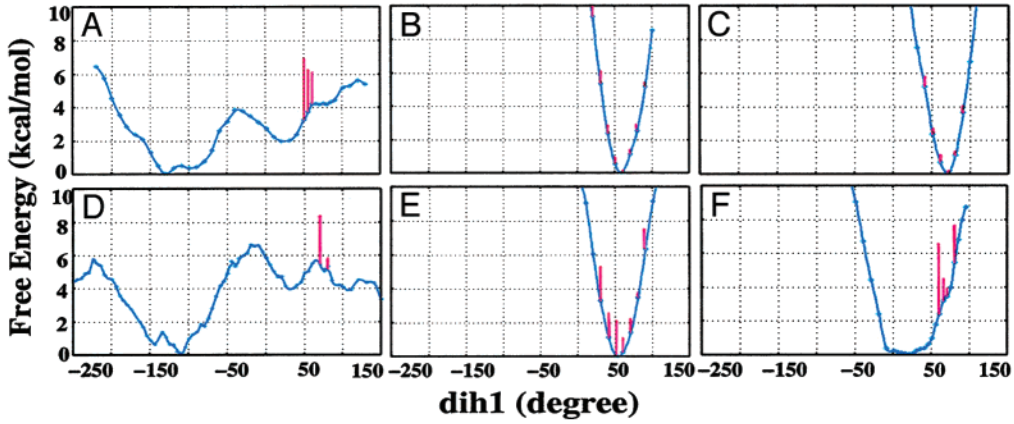


Figure 5. Free energy profiles along the dih1 of conformer I ($\Delta G°_{dih1}$) in water (A), w-EcCM (B), W-BsCM (C), IF7 (D), E52A (E), and R90Cit (F). The red bars represent the free energies for NAC formation at a given dih1 value. The precipitous minima in plots B and C occur at the optimum dih1 for NAC formation. (Reprinted with permission from ref 98. Copyright 2003 National Academy of Sciences.)

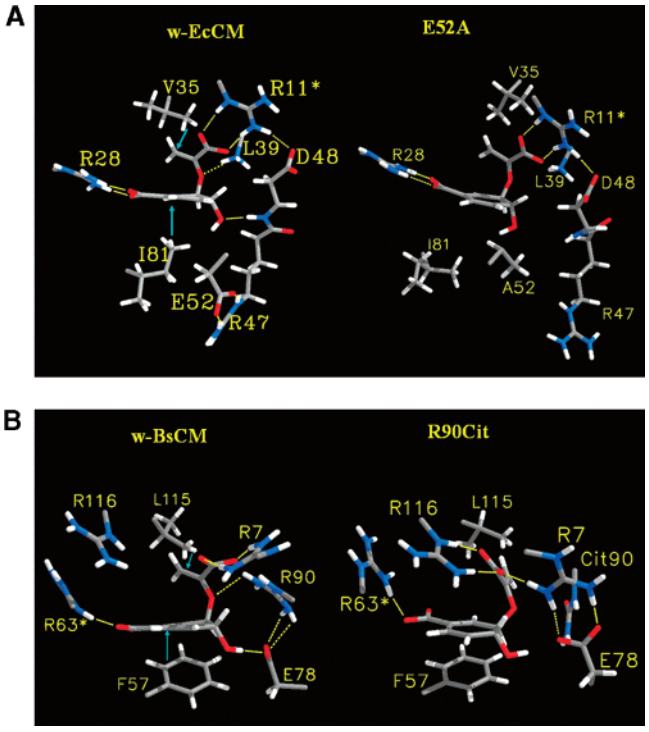


Figure 6. (A and B) Views comparing the structures of chorismate NAC, at the energy minimum, in the active site of w-EcCM and the mutant E52A (A) and in the active site of w-BsCM and mutant R90Cit (B). Notice the changes in structure for NAC in w-BsCM on mutation of Arg90 to citriline. (Reprinted with permission from ref 98. Copyright 2003 National Academy of Sciences.)

from the most stable conformation and the log of the probability of NAC formation from the fixed dih1 between 50 and 80°. The free energies vs dih1 are plotted in Figure 5.

Very much is learned by studying Figure 5. The sharp and narrow minima of the free energy profiles for chorismate in the EcCM and BsCM (Figure 5B and C) show that the active sites have been put together to bind the NAC. The minima for the E52A mutant (Figure 5E) of EcCM has moved and widened compared to the case of the wild-type enzyme. This change is very little, and one would guess that the activity of the mutant should be just a bit less than that of the wild-type enzyme. Examination of the R90Cit mutant of BsCM (Figure 5F) establishes a marked rearrangement of the active site. This alone would be expected to be

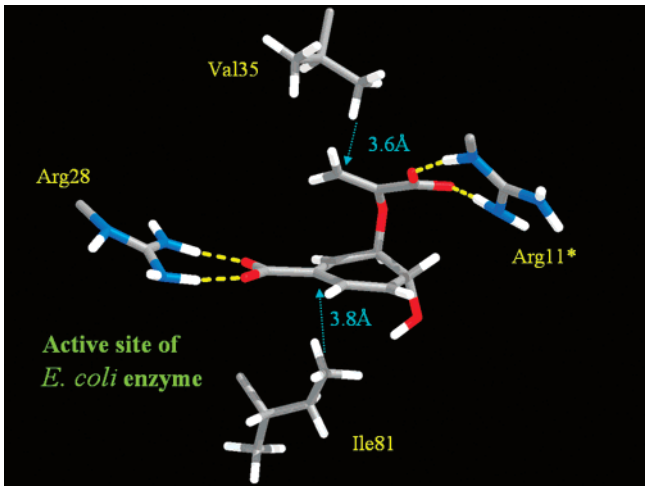


Figure 7. Chorismate at the active site of w-EcCM. Very similar structuring is present at the active sites of w-BsCM and the thermophilus enzyme. In each case, the two carboxylates of chorismate are held in position by arginines to form the NAC and two hydrophobic side chains are within the van der Waals distance of reacting C5 and C16 to prevent them from moving away. (Reprinted with permission from ref 100. Copyright 2005 American Chemical Society.)

accompanied by a significant loss of catalytic activity. The profile for the "catalytic antibody" IF7 (Figure 5D) establishes interactions that resemble those in water (Figure 5A). Views of the most stable geometries of chorismate NAC in EcCM and mutant E52A and in BsCM and mutant R90Cit are shown in parts A and B, respectively, of Figure 6.

Structuring common to the chorismate mutase·NAC complexes is shown in Figure 7. The involved arginines in the EcCM and BsCM enzymes are Arg28 and Arg11*, and Arg7 and Arg63*, respectively [arginines denoted with an asterisk come from the adjacent subunit]. The interaction of the chorismate carboxylates and the arginines creates the NAC. The hydrophobic substituents in the EcCM and BsCM enzymes are Val35 + Ile81 and L115 + F57, respectively. The sideways motions of reacting C5 and C16 are prevented by these pairs of hydrophobic substituents which are in van der Waals contact with substrate C5 and C16—much like book ends. In the R90Cit mutant of BsCM, the neutral citruline does not take the position of positively charged Arg90.

A linear free energy plot of the calculated free energies of NAC formation ($\Delta G°_{NAC}$) vs the experimentally deter-

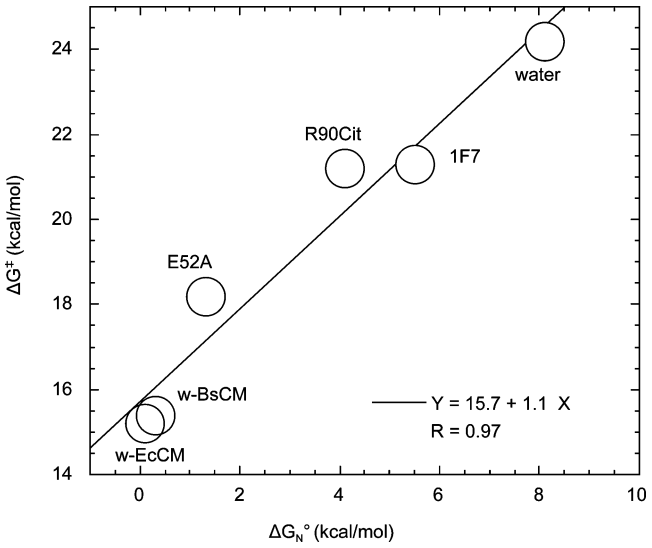


Figure 8. Plot of the computed free energies of NAC formation ($\Delta G°_N$) vs the experimentally determined free energies of activation ($\Delta G^‡$) for the Claisen rearrangement of chorismate. (Reprinted with permission from ref 98. Copyright 2003 National Academy of Sciences.)

Table 2. Free Energies of Experimental $\Delta G^‡$ and Computational $\Delta G°_{NAC}$ and the Resultant Free Energies for NAC → TS (ΔG_{TS})

	$\Delta G^‡$	$\Delta G°_N$	$= \Delta G_{TS}$
water	24.2	8.1	16.1
IF7	21.3	5.5	15.8
R90Cit	21.2	4.1	17.1
E52	18.2	1.3	16.9
W-BsCM	15.4	0.3	15.1
W-EcCM	15.2	0.1	15.1

mined free energies (eq 1) of reactions ($\Delta G^‡$) is presented in Figure 8. The slope of 1.1 (correlation coefficient 0.97) allows only one explanation. The free energy (ΔG_{TS}) of NAC → TS is close to being a constant and the free energies for reaction are a linear function of the free energy of NAC formation (Table 2).

The correlation of free energies of NAC formation and free energies of reactions is analogous to the correlation for dicarboxylic acid monoesters cyclizing to form intramolecular anhydrides (Figure 1).

The Claisen rearrangement of chorismate at the E. *coli* active site, as well as in water, and the gas phase[100] was also explored by use of QM/MM procedures. The SCCD-FTB/MM two-dimensional procedure was employed to identify TS structures followed by B3LYP/6-31+G* single point computations to determine activation energies in enzymatic and water reactions. As found in MD and thermodynamic integration free energy studies, 90% of the kinetic advantage of enzymatic over water reactions is due to a more efficient NAC formation in enzyme. The near identity of TS structures and their charge distributions in the gas phase, in water, and in the enzyme does not support significant electrostatic stabilization of the TS. Transition state stabilization can only account for ~10% of the enzymatic advantage over the water reaction.

4.1.2. Chalcone → (S)-Flavanone in Water and Chalcone Isomerase

The enzymatic and nonenzymatic chalcone → (S)-flavanone (Scheme 5) reaction has been studied extensively.[101−103] *Medicago* chalcone isomerase has much in common with EcCM. Both are single substrate enzymes which catalyze intramolecular reactions without the formation of covalent intermediates. The reactions are kinetically indistinguishable. They share, essentially, the same values of K_M and k_{cat}. The spontaneous rearrangement reactions in water also share the same value of k_o.[102]

Results from the investigation of the mechanism for chalcone → (S)-flavanone catalyzed by *Medicago* chalcone isomerase and in water[104] follow. Ground state conformers of chalcone (CHN) are defined by three dihedral angles (Scheme 5). Cis and trans forms of dih1 are identical due to the symmetry of the chalcone B ring. Thus, two conformations exist for CHN in water. Conformer I (dih2 = ~0°) has the nucleophilic O2′ and the electrophilic C9 in proximity whereas conformer II (DH2 =180°) has C9 turned away from O2. Thus, the reaction can only occur from conformer I. Conformer II is more stable in water (2.5 kcal/mol) than conformer I. Also, conformer II is the initial substrate in the enzymatic reaction. The sequence of events in water and enzyme is shown in eqs 4 and 5.

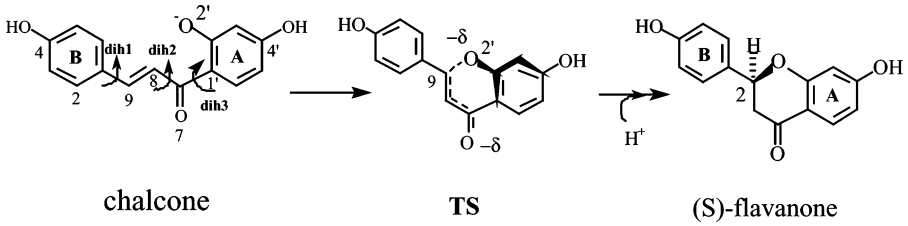
$$\text{II} \rightarrow \text{I} \rightarrow \text{NAC} \rightarrow \text{TS} \quad \text{Water} \tag{4}$$

$$\text{II} + \text{E} \rightarrow \text{E·II} \rightarrow \text{E·I} \rightarrow \text{E·NAC} \rightarrow \text{TS} \quad \text{Enzyme} \tag{5}$$

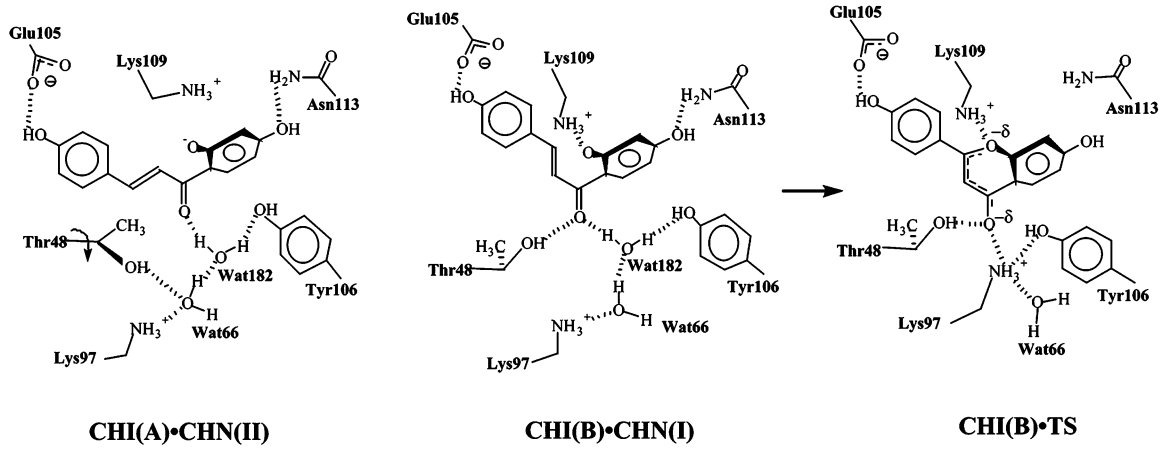
The free energy for NAC formation, $\Delta G°_{NAC}$, was determined in water as well as in both active sites (A and B) of the dimeric enzyme. $\Delta G°_{NAC}$ values are 6.0, 5.7, and 3.5 kcal/mol in the active sites A and B and in water, respectively. Thus, the enzymatic reaction does not gain any kinetic advantage over the water reaction as a result of a favorable NAC formation.

It is of value to note the relationship between TS stabilization and the importance of NAC formation in determining enzyme proficiency. The difference in the free energies of reaction in enzyme and water ($\Delta\Delta G^‡$) equals $\Delta\Delta G°_{NAC}$ plus the difference in free energies of NAC → TS ($\Delta\Delta G_{TS}$). With chorismate mutase,[69] $\Delta\Delta G°_{NAC}$ predominates $\Delta\Delta G^‡$ and interactions between enzyme and ligand are much the same for E·S and E·TS. It has been determined for prolyl isomerase[105] and orotidine decarboxylase[14] that $\Delta\Delta G°_{NAC}$ accounts for a negligible portion of $\Delta\Delta G^‡$, and it has been determined that these enzymes bind their respective TSs more tightly than NACs. Thus, $\Delta\Delta G_{TS}$ is useful in estimating the TS stabilization effect.

Scheme 5



Scheme 6



In the *Medicago* chalcone isomerase (CHI) mechanism (Scheme 6), the important features of catalysis deal with lowering the TS free energy. MD simulations of enzyme complexes of chalcone conformation 1, [CHI·CHN(I)], and the transition state, [CHI·TS], were performed. Three catalytic features were recognized to be important in the conversion CHI·NAC → CHI·TS

: (i) general acid catalysis by Lys97-NH_3^+ of ring closure (Scheme 6); (ii and iii) release of strain and freeing of bound water molecules on CHI·CHN(I) → CHI·TS. These are briefly described. Because Lys97 is deep inside the catalytic cleft, the movement of substrate toward Lys97 in the active site requires entering a hydrophobic cleft. The cleft is too narrow for the bulky substrate due to repulsion between C9 and O2′, but partial formation of the C9−O2′ bond in the TS allows movement into the cleft and, therefore, accessibility to Lys97-NH_3^+ (Figure 9).

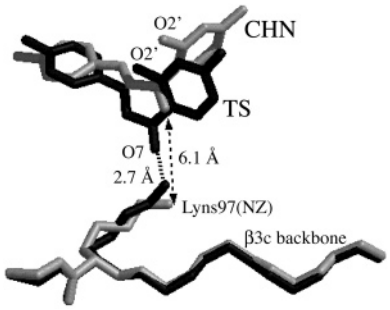


Figure 9. Superposition of enzyme-bound chalcone and the transition state when matching the backbone of the β3c strand. Only heavy atoms are shown. The E·S complex is in gray, and the E·TS complex is black. (Reprinted with permission from ref 104. Copyright 2004 National Academy of Sciences.)

In CHI·CHN(I), the A ring of CHN lies parallel to the protein surface and caps the binding site, sequestering seven water molecules in the binding site. Analysis of the mean squared displacement reveals that four particular water molecules are tightly bound in the active site of CHI·CHN-(I). Upon formation of the CHI·TS, due to the smaller size of the TS, most of the active site water molecules become relatively mobile and exchange with bulk solvent. Only one water molecule remains tightly bound to Lys97-NH_3^+ in the CHI·TS complex (Figure 10). Thus, on going from CHI·CHN(I) to CHI·TS, at least three tightly bound water molecules are released. This release is an entropic advantage of the enzyme reaction relative to the water reaction.

The apo-enzyme structure is distorted on formation of the enzyme·substrate complex and returns to the apo-enzyme structure on formation of the enzyme·TS complex. This release of strain lowers the free energy content of the enzyme transition state complex. The βH3 hairpin and the L6 loop of the apo-enzyme move closer to the ligand binding pocket on CHI binding and move back again on formation of the TS (Figure 11).

4.1.3. From Simple Systems a Profound Statement

E. coli chorismate mutase and *Medicago* chalcone isomerase have comparable values of k_{cat} and K_M. The values of k_o for

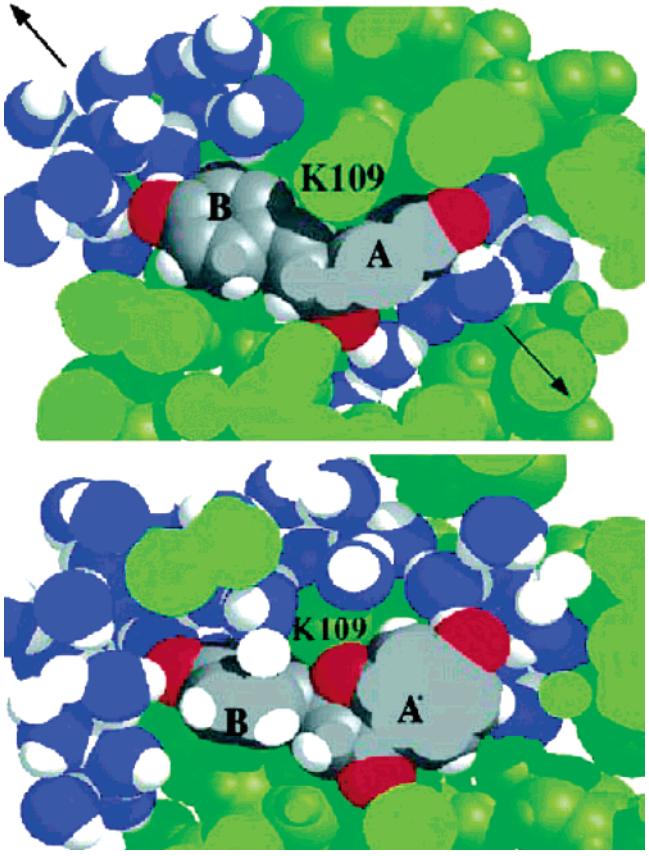


Figure 10. Dissected view of enzyme-bound chalcone (A) and the enzyme bound TS (B), cutting from the solvent toward the interior of the binding pocket. The water oxygens are blue, and the active site residues are green. (Reprinted with permission from ref 104. Copyright 2004 National Academy of Sciences.)

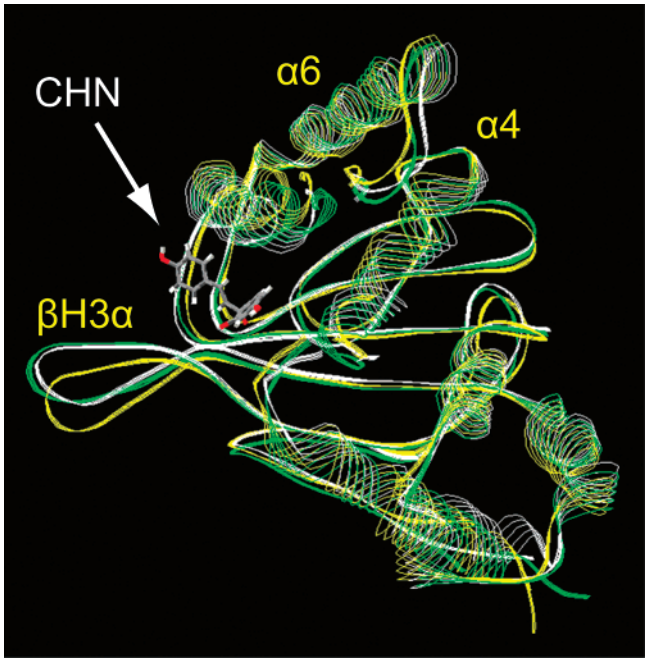


Figure 11. Changes of conformation of chalcone isomerase going from apo-enzyme (white) to enzyme substrate complex (green) and then to enzyme·transition state complex (yellow). Note the overlap of apo-enzyme and enzyme·transition state structures.

chorismate → prephenate and chalcone → (*S*)-flavanone in the water reactions are also essentially the same. Thus, the catalytic proficiencies of the two systems are identical. The choice of these two kinetically identical systems has provided a test of the long standing belief that the general feature of enzymatic catalysis is stabilization of the transition states.[6,7] Enzymatic proficiencies have been believed to depend on the difference in TS stabilization by enzyme and by water.[11] In the chorismate → prephenate reaction, the advantage of the mutase enzymatic reaction over the water reaction relies on the enhancement of the NAC formation in the ground state with only ~10% contribution of TS stabilization. In the chalcone → (*S*)-flavanone reaction, the advantage of the *Medicago* chalcone enzymatic reaction compared to the reaction in water is ultimately in TS stabilization. The important lesson from this comparison is that catalysis occurs where it is most needed, as noted by Jencks.[106] Generalities are to be avoided, as noted by Bruice.[107]

4.1.4. Thermophilic Indole-3-glycerol Phosphate Synthase

A common property of thermophilic enzymes is their stability at temperatures greater than ambient. Various suggested mechanisms for the thermostability of twenty-six thermophilic enzymes are provided in Table 5 of the valuable review[108] by Vieille and Zeikus. The question commonly asked is how thermophilic enzymes achieve the balance between the need for stability at high temperatures and the need for sufficient flexibility in order to function properly.

The hyperthermophilic indole-3-glycerol phosphate synthatase from *Sulfolobus solfataricus* (IGPS) catalyzes the terminal ring closure step in tryptophan biosynthesis (CdRP → IGP, Scheme 7). Useful insight into the catalytic mechanism of IGPS is provided in a publication[109] by Kirschner and co-workers.

Chart 9

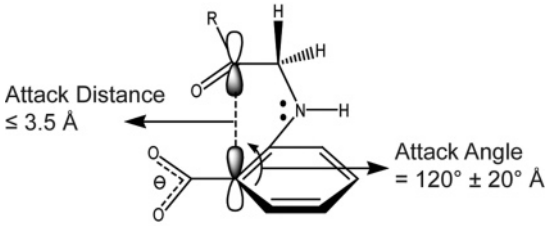


Table 3. Relationship of Specific Activity of IGPS vs NAC Population at Given Temperatures

temp (K)	relative activity	population of NACs
298	1	1
333	48	25
363	740	478
385	4200	1103

IGPS is another single substrate enzyme which catalyzes a ring closure reaction without formation of an enzyme covalent intermediate. The object of a study by Mazumder-Shivakumar[70] was to compare MD simulations of the enzyme·substrate complex (IGPS·CdRP) structures at ambient and at higher temperatures in order to determine ground state structural differences that contribute to the rate enhancement of the reaction at higher temperatures. The population of near attack conformations (Chart 9) of IGPS·CdRP should be proportional to the rate constant for reaction at different temperatures. From the MD studies at the various temperatures, this was found to be the case (Table 3). In this study, NACs were defined as conformers with the distance between the two carbons of the forming bond within 3.5 Å and the defined attack angle within 20° deviation from the angle of 120° in the TS (Chart 9). At 385 K the attack angle in the enzyme bound substrate is on average (134°) that required for ring closure while at 298 K the angle (165°) is not favorable for the ring closure reaction. A critical role in positioning of the substrate is the hydrogen bonding of Lys52-NH_3^+, as shown in Chart 10. The positioning of the general acid Lys110-NH_3^+ to the electrophilic carbonyl oxygen is also favorable at 385 K ($-H^+$- - -O═ of 2.8 Å) but not at 278 K ($-H^+$- - -O═ of 3.4 Å).

From these results it may be concluded that, in the evolution of this thermophilic enzyme, electrostatic forces were optimized to provide reactive conformations at 385 K. As the temperature is lowered, there is an increase in the strength of the electrostatic interactions, accompanied by a distortion of the substrate and catalytic apparatus of the active site. Could this be general for thermophilic enzymes?

4.1.5. Cytosine Methylation Catalyzed by M.HhaI

The bacterial enzyme M.HhaI catalyzes the reactions of certain DNA-cytosine residues with *S*-adenosyl-L-methionine

Scheme 7

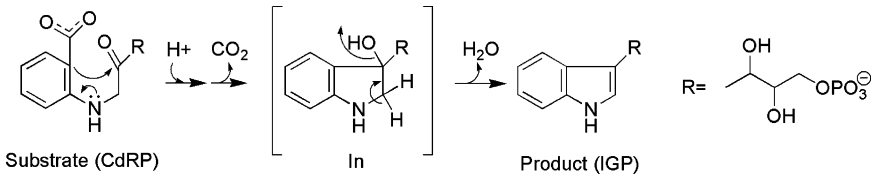


Chart 10

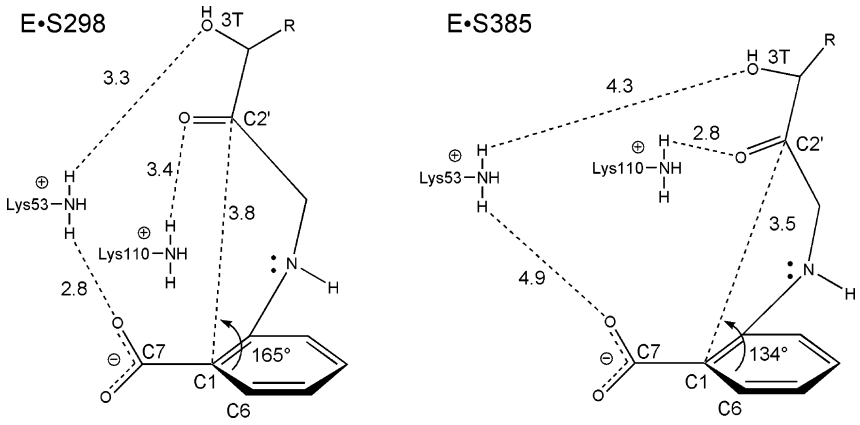
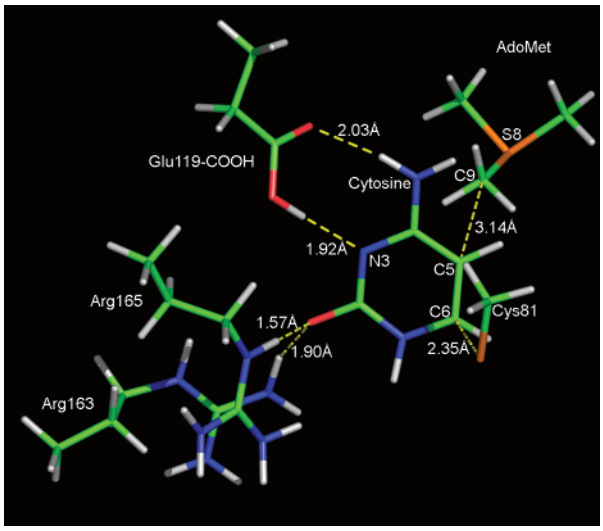


Chart 11



(AdoMet) to provide a DNA-[5-methyl] cytosine and *S*-adenosyl-L-homocysteine (AdoHcy).

The mechanism of the M.HhaI catalysis has been investigated in this laboratory by QM/MM computations.[110] The QM region depicting the reactant ground state is found in Chart 11. From past studies, the mechanism has been considered to be stepwise (Scheme 8).[111] The stepwise mechanism has been suggested[112] to involve general acid (by Glu119-CO_2H) catalysis of Cys81-S^- addition to cytosine followed by general base (by Glu119-CO_2^-) catalysis of methylation. Alternately, the stepwise reaction has been proposed to be catalyzed by stabilization of the 6-CyS-cytosine anion through the electrostatic interactions with Arg163 and Arg165.[113]

Following SCCDFTB/MM generation of ground, intermediate, and transition states, single point calculations at the level of B3LYP/6-31+G*//MM were carried out to determine activation energies. Cys81-S^- addition and methylation were found to be concerted (Scheme 9) with the transition state

Scheme 9

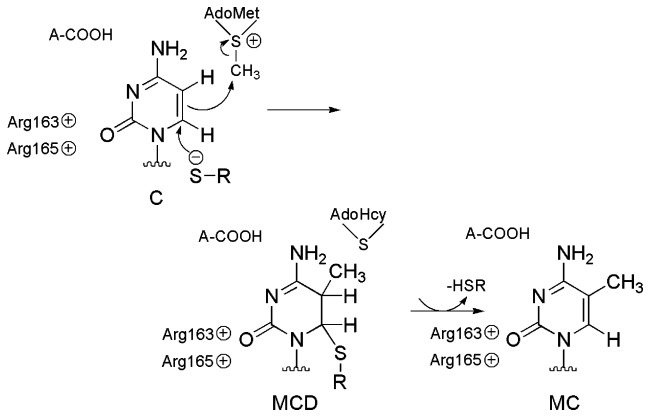
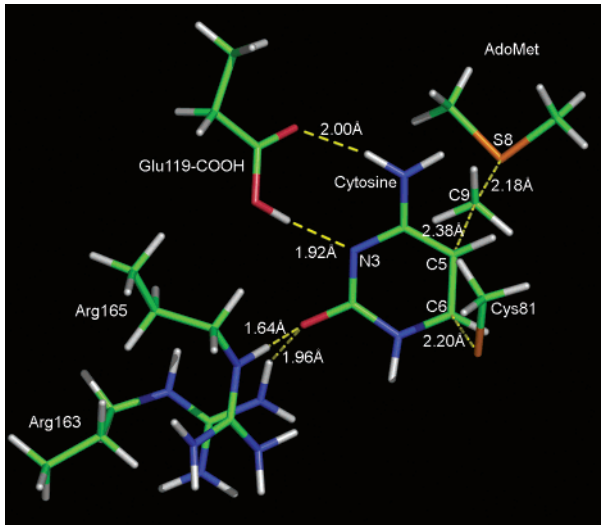
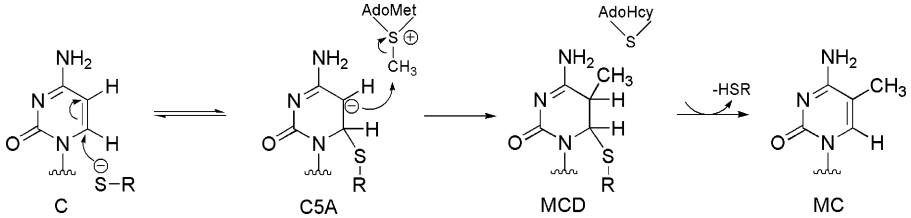


Chart 12



(Chart 12) at ~40% reaction. Changes in the separation of electrostatic interactions of Glu119-COOH, Arg163, and Arg165 on going from the ground state to the transition state

Scheme 8



Scheme 10

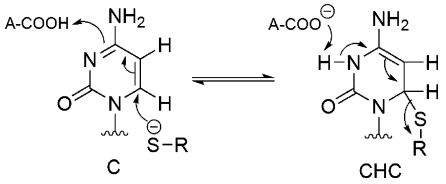
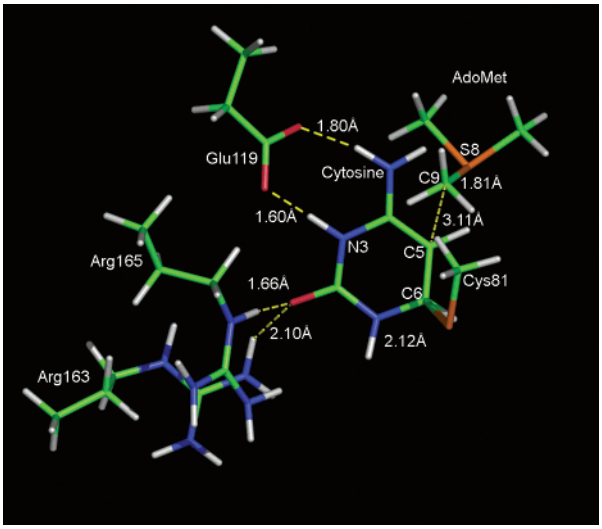


Chart 13



for the concerted reaction are 0.02, 0.06, and 0.07 Å, respectively. Thus, the role of Glu119-COOH is not catalytic but is to place the cytosine in position to create the active ground state conformation. Arg163 and Arg165 are also important in creating the reactive conformer and may play a small role in catalysis.

A side reaction involves Glu119-COOH general acid-catalyzed addition of Cys81-S$^-$ to cytosine (Scheme 10). The structure of enzyme•CHC is given in Chart 13. The formation of CHC is associated with a very small activation energy; CHC has stability comparable to that of the reactants and does not undergo methylation. The reversible reformation of CHC serves as a means for proton exchange in cytosine which is a side reaction during methylation.

The reaction coordinates for these reactions at the level of SCCDFTB/MM are present in Figure 12. The direct

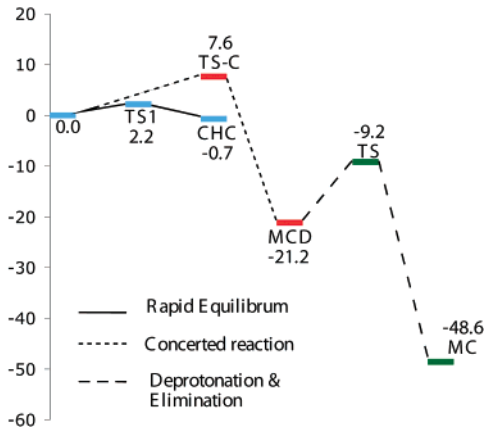


Figure 12.

product of the methylation step (5-methyl-6-Cys81S-5.6-dihydrocytosine) undergoes deprotonation with expulsion of Cys-S$^-$. It would appear that the only base present is HO$^-$, which is at a suitable concentration at pH 7, as shown in Figure 12.

5. Dynamic Motion in Catalysis—What Is It Worth?

5.1. Cross-correlation and Normal Mode Analysis

Long-term MD simulations of X-ray derived coordinates of an enzyme system in an explicit or implicit water pool provide dynamic coordinates much the same as those of the protein in water solution. (There are implicit solvent systems that approach being able to provide membrane-like media.[114]) It is a simple procedure to determine the time dependence of the relative motions of any two atoms from their MD trajectories. These thermal motions occur in picosecond time scales and can be followed into the nanosecond time range. Motions of any related pair of atoms are either in the same direction (correlated motion) or in the opposite directions (anticorrelated motion). A protein structure is composed of loops connecting rather rigid structures such that correlated and anticorrelated motions are not random. Using the data generated by MD simulation, the cross-correlated coefficient $C_{i,j}$ can be calculated from the cross-correlated coefficient equation (eq 6),

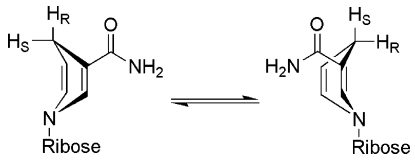
$$C_{i,j} = \langle \Delta\vec{r}_i \cdot \Delta\vec{r}_j \rangle / (|\Delta\vec{r}_i| \cdot |\Delta\vec{r}_j|) \qquad (6)$$

where Δ_i and Δ_j are the atomic displacements of atoms i and j, respectively. Here, $C_{i,j} = 1$ means 100% correlation, and $C_{i,j} = -1$ means 100% anticorrelation between the motion of atoms i and j.

A very common question concerning the importance of such dynamic structures to the rate constants of enzymatic reactions is: since enzymatic reactions take place in milliseconds to seconds, how could motions at the picosecond range, which can be followed to nanosecond times, be of kinetic significance? The picosecond motions can only be followed to the nanosecond times because of the speed limits of present computers. These short distance motions continue for the life of the ETS complex. A reactive conformer (NAC) forms 10^n times before crossing the TS energy maximum. Anticorrelated motions may be important in forming the NAC by pushing reactants together.

Motions of the enzyme that occur in the time space for enzymatic reactions are usually domain motions, which can be followed by NMR. The normal mode analysis (NMA) methodology provides a means to study these motions computationally.[115−117] Experimentally, the motions of a molecule can be detected by the vibrational spectrum (i.e. infrared spectrum). The vibrations are given by the normal modes of the molecule, which correspond to the absorption in the vibrational spectrum. In the procedure of normal mode analysis,[118] the potential energy function of the system is expanded around the minimum of the energy surface. A harmonic approximation is applied to truncate the high order displacements so that the potential energy is expressed in a summation of quadratic terms. A Hessian matrix is built based on the truncated potential energy function. The matrix is diagonalized in order to obtain the eigenvalues and eigenvectors. The eigenvalue corresponds to the frequency of a motion, and its associated eigenvector gives the direction and relative amplitude of this motion. The nature of atom i

Chart 14



coordinate displacements ($\Delta\vec{r}_i$), characterized by NMA, is given by the expression of eq 7,[119]

$$\Delta\vec{r}_i = \vec{r}_i - \vec{r}_i^{\circ} = \sqrt{\frac{k_B T}{m_i}} \sum_{j=7}^{3N} \frac{\vec{a}_j^{(i)}}{\omega_j} \qquad (7)$$

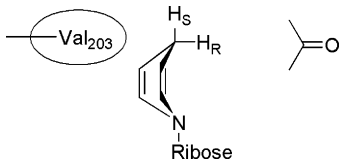
where k_B is the Boltzmann constant, m_i is the atomic mass of atom i, T is the absolute temperature, ω_j is the frequency for the jth normal mode, and $\vec{a}_j^{(i)}$ is the corresponding eigenvector for the atom i.

The summation starts from mode 7 to 3N, while the first 6 modes related to the translational and rotational motions are excluded. From this expression it can be seen that the lowest frequency modes generally provide the largest contributions to the atomic displacement of any atom. Thus, the low-frequency modes of proteins are particularly interesting because they make major contributions to the conformational motions at thermal equilibrium. Although the ability of NMA is limited due to the harmonic approximation of the potential energy, it is adequate to acquire the essential features of the collective motions for the biological system. It has been shown by many studies that the results from NMA not only are qualitatively reasonable in most cases, compared to those from MD, but also can provide a proper description of the functionally important large scale motions.[115−122] Cui et al.'s study of structural plasticity in F1-ATPase using NMA comprises the entire $\alpha_3\beta_3\gamma$ complex, which is difficult to employ with the MD simulation.[123] Mazumder-Shivakumar et al. showed the critical motion of a flexible loop in the IAG-nucleoside hydrolase[85] by the NMA based on a simplified elastic network model.[124,125] Brink et al. demonstrates the potential of the combination of normal mode analysis with multiple model refinement to elucidate the multiple conformations of flexible proteins in their study of human fatty acid synthase.[126] The time span of NMA calculated motions is generally not provided but can be estimated from the frequency of the motions. The motions can be in the time span of enzymatic reactions.

5.2. Hydride Equivalent Transfer Reactions

The reactive ground state conformers are probably the same regardless of whether H^- transfer involves a TS over the reaction barrier or involves tunneling. Almarsson[127] carried out a MD study of dogfish lactate dehydrogenase (DLD) containing pyruvate and NAD(P)H at the active site. In the X-ray crystal structure, the transferring H_R of the planar NAD(P)H is at a distance of 2.5 Å from the pyruvate carbonyl carbon. (At this time (1993) the dihydronicotinamide portion of NADH was planar in all crystal structures due to the wrong force field parameters employed.) 1,4-Dihydronicotinamides possess two quasi-boat conformations. Using molecular dynamics, it was shown that the two quasi-boat conformations are interconvertible by a wagging motion of C4 such that H_R is axial in one conformation and H_S is axial in the other conformation (Chart 14). When H_S is transferred to substrate, H_S is axial, and when H_R is transferred, it is axial. In DLD·NAD(P)H·pyruvate, the axial H_R is at a 1.5 Å distance from the pyruvate carbonyl carbon. The motion of this anisotropic system is essential to the hydrogen transferring mechanism. Further, Almarsson[127] found that the bulky isobutyl group of Ile249 and the isopropyl side chain of Val136 cover the B-face of the nicotinamide ring. The normal wagging motion of the 1,4-dihydropyridine is directed away from the bulky amino acid side chains such that the anisotropic motion of the enzyme pushes NAD(P)H toward the substrate (Chart 15).

Chart 15



The feature of bulky groups present at the face of the cofactor distal to the transferred hydride equivalent was found to be general in many NAD(P)H dependent dehydrogenases: lactate dehydrogenase, Ile249 and Val36; dihydrofolate reductase, Phe102; malate dehydrogenase, Ala245 and Leu157; glyceraldehydes-3-P dehydrogenase, Ile12 and Try317; horse liver alcohol dehydrogenase, Val203.[127]

5.2.1. Horse Liver Alcohol Dehydrogenase (HLADH)

HLADH is functional as a homodimer. Each subunit binds a molecule of NAD^+ and a Zn^{2+} at the active site. MD simulations have defined the structures involved in the stepwise reaction and the path for H^+ transfer away from the active site (eq 8).[66]

$$\text{HLADH·NAD}^+\text{·PhCH}_2\text{OH} \rightarrow \text{HLADH·NAD}^+\text{·PhCH}_2\text{O}^- + \text{H}^+$$

$$\text{HLADH·NAD}^+\text{·PhCH}_2\text{O}^- \rightarrow \text{HLADH·NADH·PhCHO} \qquad (8)$$

Experimental investigations of HLADH by Plapp[128−130] and Klinman[130,131] are notable. The creation of mutants of the bulky amino acid side chain (Val203) and determination of their kinetic properties established[131] a decrease in the H^- transfer rate with a decrease in the bulkiness of the amino acid 203. The log of the experimental k_{cat}/K_M decreases in a linear fashion (eq 9) with an increase in the MD-derived closest contact distance (CCD) between C4 of NAD^+ and C7 of $PhCH_2O^-$ (Figure 13).[132]

$$\log(k_{cat}/K_M) = 1.63\cdot\text{CCD} + 6.65$$

$$\log(k_{cat}/K_M) = \log k_5 - \log K_d(\text{ROH}) \qquad (9)$$

Dissociation of $PhCH_2OH$ (where K_d(ROH) is the dissociation constant of $PhCH_2OH$) is not likely related to the amino acid at position 203 because this amino acid is not in the alcohol binding pocket but on the other side of the nicotinamide ring. Thus, the linear dependence of k_{cat}/K_m probably is due to the linear dependence of k_{cat} on CCD. In this series of mutants, the relative rate is determined by the mole fraction of NACs. The bulky substituents promote NAC formation.

There is much interest in determining if the bulky residue not only limits the active site space, thereby bringing

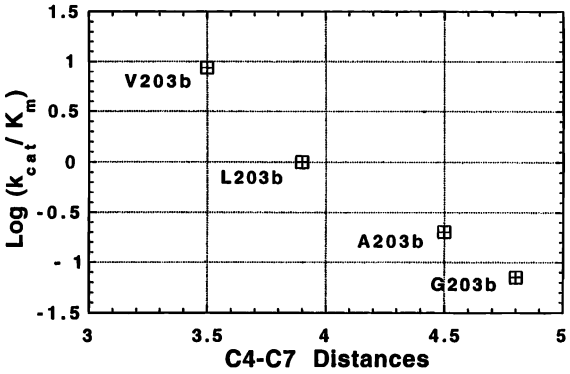


Figure 13. Log of the second-order rate constants for the reactions of LADH (Val203Leu, L203b) and three mutants (Val203Leu, L203b; Val203Ala, A203b; Val203Gly, G203b) with $PhCH_2OH$ plotted vs the calculated close contact distances of C4 of $NADH^+$ and C7 of the substrate. Rate constants are from the experimental literature of refs 112 and 113. (Reprinted with permission from ref 132. Copyright 1999 Elsevier.)

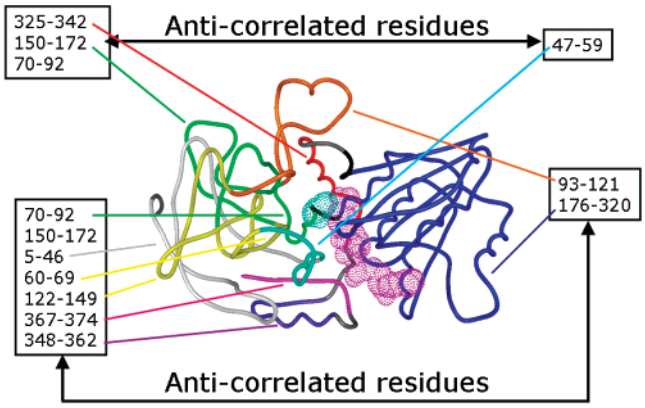


Figure 14. Anticorrelated motions mapped onto the HLADH enzyme monomer structure with NAD^+ (pink) and substrate (cyan) in place. The regions of the anticorrelated motions are assigned and labeled. (Reprinted with permission from ref 134. Copyright 2004 National Academy of Sciences.)

reactants together, but also actually physically pushes reactants together. Long term (10 ns) MD simulation of the HLADH·NAD^+· $PhCH_2O^-$ species followed by cross-correlation analysis provided the mapping of Figure 14.[133]

Figure 14 shows the amino acid sequences of the cofactor binding domain (blue) are in correlative motion moving together, and this motion is anticorrelated with the substrate and essentially the remaining residues. These back and forth anticorrelated motions of the NAD^+ binding domain and the pushing together of the reactants were determined by examination of the cross-correlation coefficients and MD distance histograms. The four pairs of anticorrelated interactions at the active site and in the vicinity of the NAD^+ nicotinamide are (i and ii) the cofactor binding domain Cα of Val292 and the CG1 of Val203, both with C7 of $PhCH_2O^-$; (iii) the cofactor binding domain amide carbonyl oxygen of 1318 with the amide N of H67; and (iv) the cofactor binding domain Cα of T178 with the carbonyl oxygen of L141 (Figure 15).[134] The motions of (i) and (ii) are most important in the 0.5 Å motion of C4 of NAD^+ toward C7 of $PhCH_2O^-$ to form near attack conformations. Long-term nanosecond MD simulations of an E·S provide a Boltzmann distribution of thousands of conformations. Excluding the rapid equilibrium between these conformations, they must vary from those with no reactivity at all to those with maximum activity. Thus, the free energy of activation is dependent upon conformation.

Recent studies with acetylcholine esterase[135] and dihydrofolate reductase[136] have involved QM/MM calculations starting with each of a number of E·S species differing in the conformation of S as well as E. The E·S species were chosen from snapshots collected from MD simulations. Values of the calculated reaction energies differed significantly, dependent on the individual E·S structure. The average reaction energy, however, compared favorably with the experimental value. A comparable procedure has been carried out with the HLADH·NAD^+·$PhCH_2O^-$ complex.[134] Seventy-one snapshots were randomly chosen from the 10 ns MD simulation trajectory, and the reaction energy was computed for each. Of these, the twenty-one conformers associated with the lowest reaction energies possessed NAD^+·$PhCH_2O^-$ reactant conformations characterized as NACs.[134] A portion of the E·S structures, associated with greater reaction energies, also possessed NAD^+·$PhCH_2O^-$ conformations characterized as NACs. This confirms that not only the substrate conformation but also the enzyme structure require the optimal configuration to lower the reaction barrier. As in the previous studies,[135,136] the reaction barriers were averaged.[134] When the averaged value was corrected to the vibration energy,[137] the resultant value compared favorably with the experimental tunneling barrier height for hydride transfer (15.6 vs 16 kcal/mol, respectively).

MD simulations and cross-correlation analysis along with QM/MM calculations have provided strong evidence that in the HLADH·NAD^+·$PhCH_2O^-$ → HLADH·NADH·PhCHO reaction the rate-limiting step of hydride equivalent transfer involves a motion of the NAD^+ binding domain which brings together C4 of NAD^+ and H−C7 of $PhCH_2O^-$ structured as an NAC. The formation of the reactive conformation is, at times, associated with the *lowest* energies for E·TS formation. Presumably, when NAC formation is not associated with the

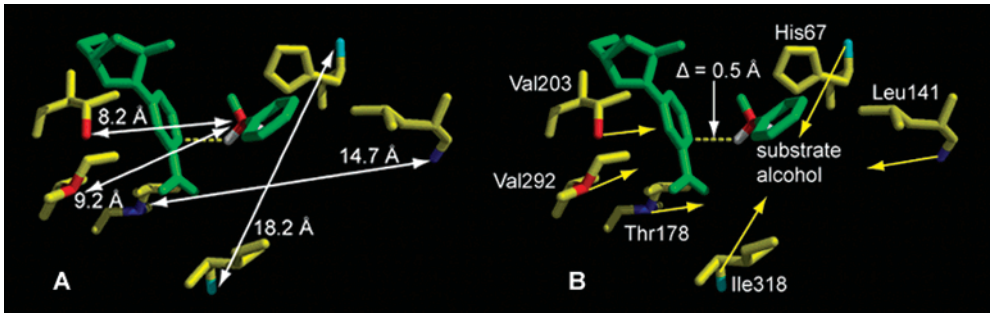


Figure 15. (A and B) Anticorrelated motions at the active site of HLADH·NAD^+·$PhCH_2OH$. In part A, the arrows provide the average distance between interacting pairs. In part B, the pushing of cofactor toward substrate by Val203, principally, and Val292 is shown to decrease the distance between cofactor Cα and substrate C7 by 0.5 Å.

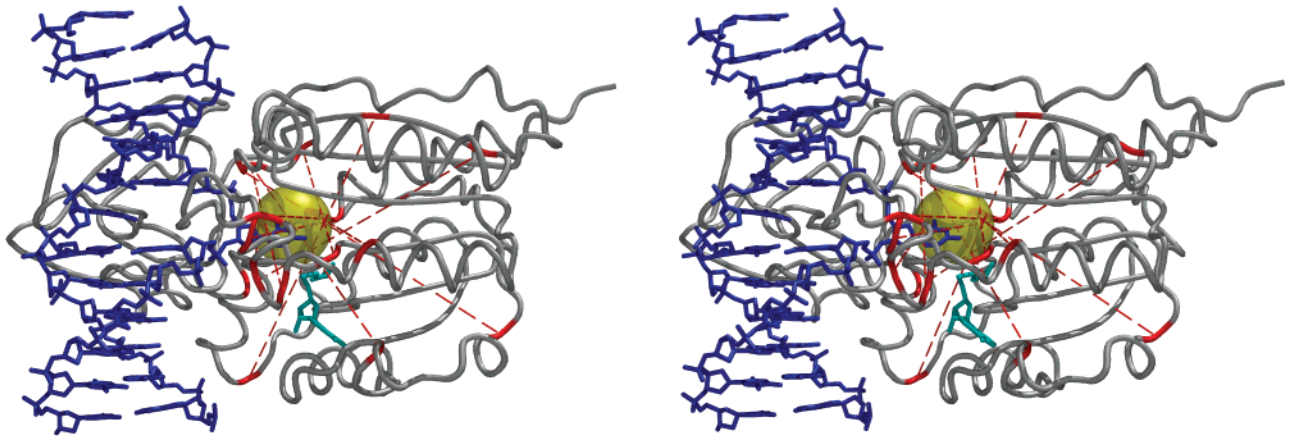


Figure 16. Stereoview of the interaction of the combined SCA-anticorrelation pairs of amino acid residues at the active site of Mhal methyltransferase.

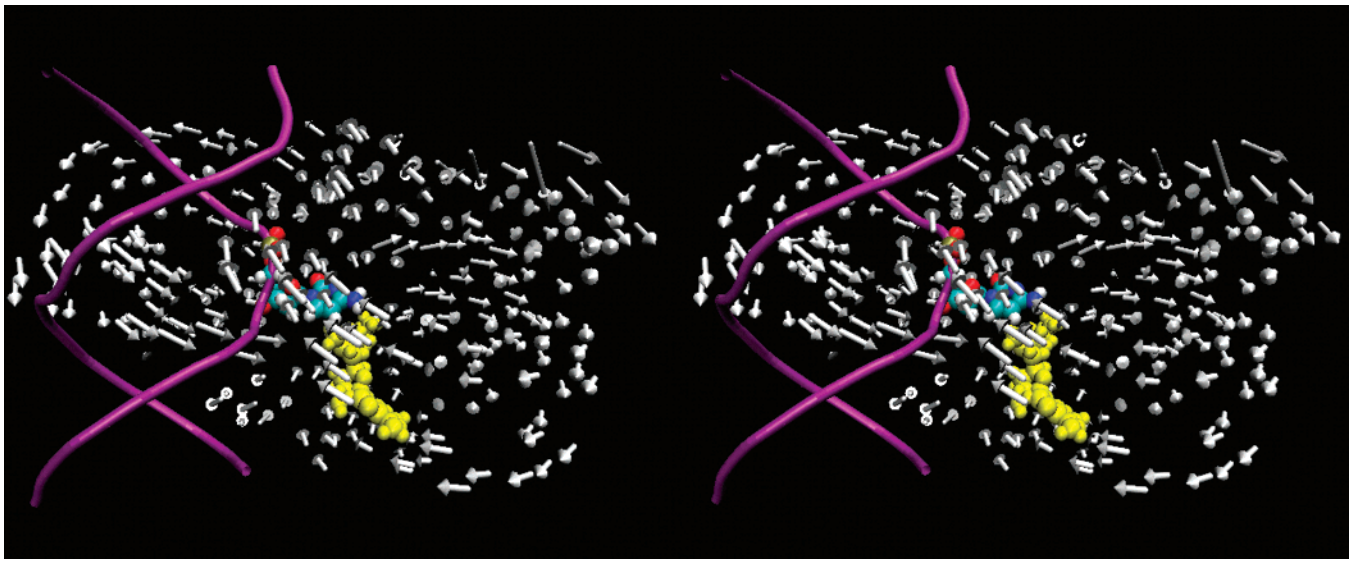


Figure 17. An overall picture of the low-frequency mode of the enzyme interaction. Gray arrows represent the direction of motion of Cα atoms. The enzyme residues are omitted for easier viewing. The DNA backbone structure, the flipped out base, and the cofactor SAM are shown in pink, yellow, and to-atom color, respectively. (Reprinted with permission from ref 145. Copyright 2005 National Academy of Sciences.)

lowest energies, the protein conformation is not optimal for reaction. It has been shown that a single enzyme molecule (β-galactosidase) exhibits large temporal fluctuations of the turnover rate constant from 1 ms to 100 s. Further, the rate constant fluctuations are associated with fluctuations of the protein conformations observed on the same time scale.[138]

5.3. M.HhaI Methyltransferase Reaction

M.HhaI is one of the many AdoMet dependent DNA modifying enzymes. It catalyzes the AdoMet methylation of the first cytosine residue in the specific sequences 5′-GCGC-3′ of double stranded DNA. The structure of M.HhaI has been characterized by X-ray crystallography in complexes with various forms of DNA substrate.[139,140] From the X-ray crystal structure of a ternary complex of M.HhaI, its DNA substrate, and the reaction product, *S*-adenosyl-L-homocysteine (SAM), it was found that the target cytosine swings 180° completely out of the DNA helix and into the active site pocket.[139] The DNA binding induces a large movement of the "catalytic" loop (residues 80−89) closure. The mechanism of the M.HhaI catalyzed methylation of a DNA-cytosine is provided in section 4.1.5. NMR studies have shown that addition of the cofactor analogue SAM enhances the "base-flipping motion" of the target base.[141] Further, the crystal structures of enzyme containing DNA with mismatches (G:A, G:U, or G:AP [AP = abasic]) at the target base established that the enzyme does not require a flipped-out cytosine residue as part of its initial recognition mechanism.[142] Do thermal motions play a role in these enzymatic processes?

5.3.1. Statistical Coevolution Analysis and Anticorrelated Motions[143]

Cross-correlation analysis from the MD simulations of the M.HhaI methyl transferase complex with DNA and *S*-adenosyl-L-methionine (AdoMet) provided knowledge of the pairs of anticorrelated amino acid residues. Separately, a coevolution statistical coupling analysis (SCA)[144] was carried out to determine the coupling of pairs of residues in coevolution. The anticorrelation data of MD and the SCA data where combined by determining the product of the two sets in order to weigh anticorrelated motion by evolutionary conservation. In this fashion, both evolutionary important interactions that do not manifest as dynamic couplings and MD correlations that are not conserved are eliminated. In essence, the SCA·MD reveals the subset of residues that display conserved

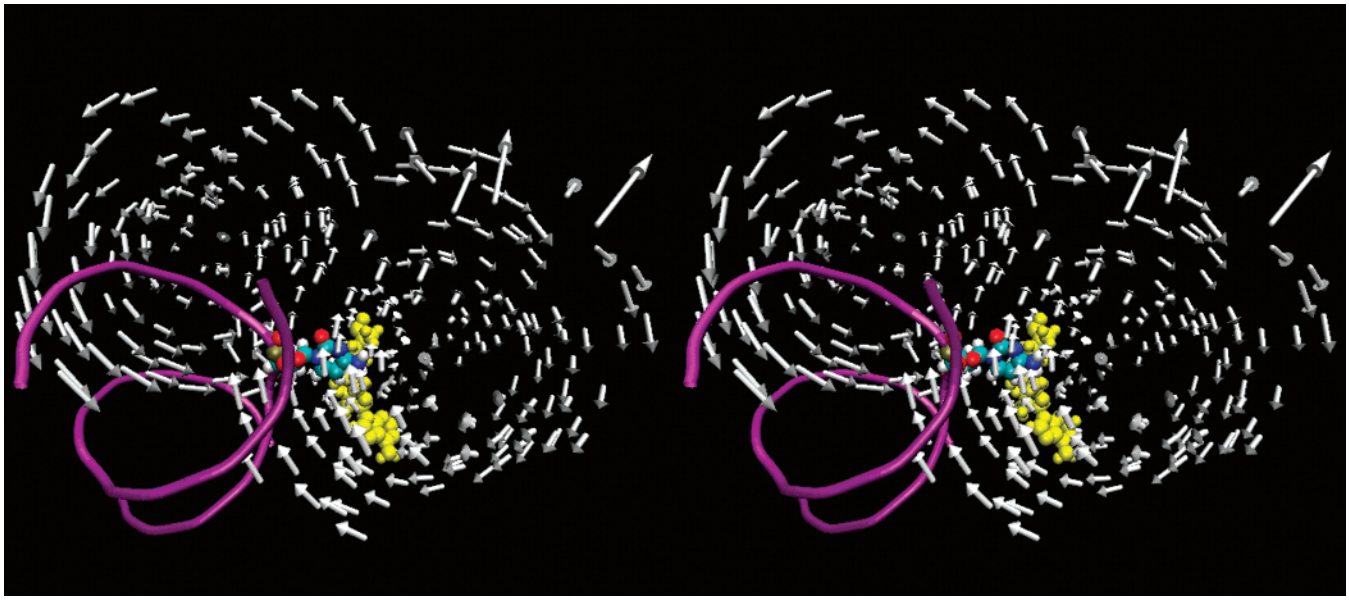


Figure 18. Top view of Figure 17. The arrows representing the large (lower right) and small (upper left) domains qualitatively form two circles. The group of arrows entering the minor groove (from the right of the helix) points directly toward the phosphodiester and sugar backbone around the flipped base. The arrows entering the major groove from the back (left) of the helix point directly to the complementary orphan guanine. (Reprinted with permission from ref 145. Copyright 2005 National Academy of Sciences.)

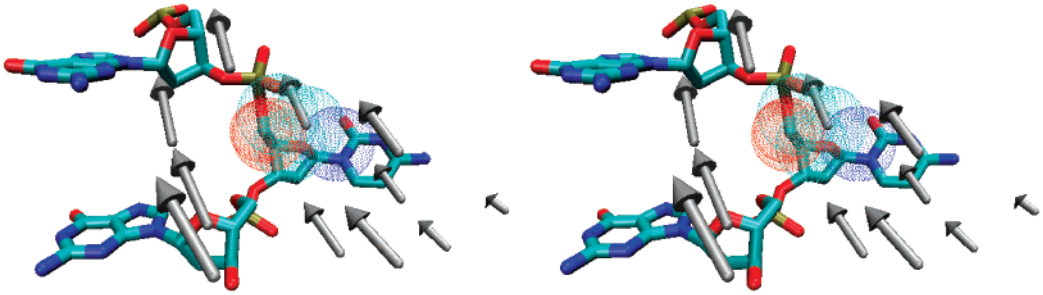


Figure 19. Enlarged local view of the vectors for resides 79−89 (Phe79, Pro80, Cys81, Gln82, Ala83, Phe84, Ser85, Ile86, Ser 87, Gly88, Lys89). The flipped out base and the adjacent bases of the same strand are shown explicitly. Ser 85 is shown in a dotted surface fashion. The vector of this residue is located on the lower right within the top trio of vectors. (Reprinted with permission from ref 145. Copyright 2005 National Academy of Sciences.)

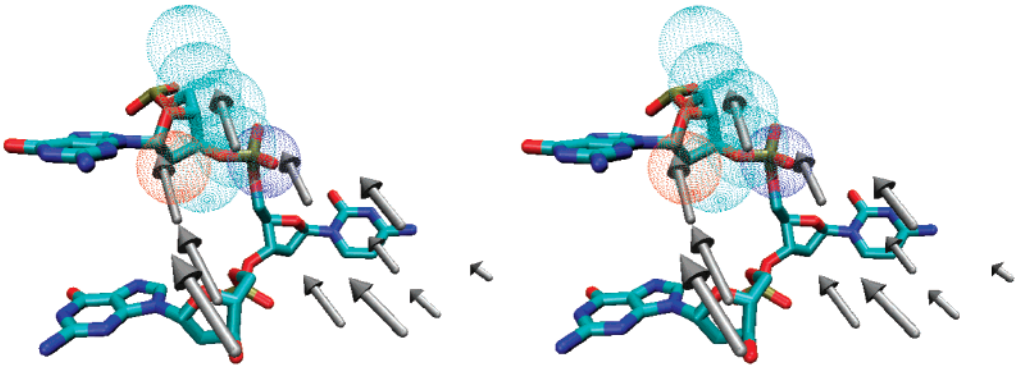


Figure 20. The vector of the Ile86, shown as a dotted surface, is located at the center and on the top of the trio of vectors. (Reprinted with permission from ref 145. Copyright 2005 National Academy of Sciences.)

and coupled anticorrelated motions (Figure 16). Observation of Figure 16 shows that the cross-correlated interactions form a clear focal point centered at the active site.

Will the application of this combined SCA-MD approach to other proteins provide interesting results? Will this approach lead to a means of identifying amino acid residues for experimental mutation studies? How does this approach help in understanding the importance of dynamics in mechanisms? These are some of the questions that can be of potential interest to enzymologists.

5.3.2. Normal Mode Analysis (NMA) of the Mechanism of Turning out of the DNA Target Nucleotide

The elastic network model of NMA has been used in the investigation of the long-time protein dynamics in the M.HhaI reaction.[145] It was found that the large amplitude, low-frequency mode of the DNA Hhal methyltrasferase can be related to the base flipping mechanism. Figure 17 shows an overall picture of the low-frequency mode of the enzyme interaction. Each arrow in gray represents the direction of

motion of the Cα atom. In Figure 17, the DNA backbone structure, the flipped out base, and the cofactor SAM are shown in pink, yellow, and atom color, respectively.

Figure 18 is a top view of Figure 17, and the arrows representing the large (lower right) and small (upper left) domains qualitatively form two circles. In front of SAM and the flipped out base of the large domain (amino acids 1−193) that appear on the right side of Figures 17 and 18, the vectors of the low-frequency normal mode of residues Ser85 and Ile86 (see enlarged local views in Figures 19 and 20) point directly to the phosphate and sugar groups of the DNA backbone near the target base in position to rotate the dihedral angles and flip the base out of the DNA duplex. For the small domain (amino acid residues 194−275) at the back of the DNA backbone (in pink), there are a group of arrows pointing toward the orphan base guanine of the complementary strand from the major groove. The vector of Gln237 is situated on the major groove side in front of this group of vectors. The motions of these vectors are in the proper orientation to assist base separation.

Figures 19 and 20 show the vectors for residues 79−89 (Phe79, Pro80, Cys81, Gln82, Ala83, Phe84, Ser85, Ile86, Ser87, Gly88, Lys89). These vectors are located in front of the large domain and point up toward the phosphodiester group of the flipped out base (Figures 18 and 19). Figures 19 and 20 provide an enlarged local view of the picture. The flipped out base and the adjacent bases of the same strand are shown explicitly to display their relative positions. In Figure 19, residue Ser85 is shown in a dotted surface fashion. The vector of this residue is located on the lower right within the top trio of vectors. It would appear that Ser85 has direct influence on the phosphodiester and sugar group of the flipped out nucleotide.

In Figure 20, the vector of residue Ile86, shown as a dotted surface, is located at the center and on top of the trio of vectors. Ile86 has direct influence on the phosphodiester group of the flipped out nucleotide. Thus, low-frequency motions of the M.HhaI enzyme are likely responsible for base flipping in the major grove pathway.

6. Summary

This review has been put together with the intention of providing the experimental enzymologist interested in structure, dynamics, and function in catalysis with a talking knowledge of the possibilities present in modern computational chemistry. Molecular dynamic (MD) simulations can convert the somewhat distorted coordinates of crystal lattice heavy atoms into coordinates, complete with covalently bound hydrogens, and place them in the context of an aqueous (or other) environment. Portions of enzyme structures with B-factors too large to be "seen" in the X-ray coordinate become "visible" in the MD derived structure. From the MD simulations, one can calculate the dynamic motions of atoms in the picosecond to nanosecond time scale (cross-correlation analysis) and also in the time period for reaction (normal mode analysis). Combining semiempirical/quantum mechanics and MD provides QM/MM methods that allow the computation of the transition state (TS) structure and the energies of activation for individual steps in a multistep process. To assist in the understanding of pH vs log k_{cat} profiles, it is important to know the pK_a values of ionizable functional groups in the protein structure. There are computational means for doing so. The author is of the opinion that these computational approaches will become tools of the X-ray crystallographer and experimentalist.

7. Acknowledgment

The author would like to thank the members of his laboratory, especially Istvan Szabo, for help in the preparation of this manuscript. I thank Professor Arieh Warshel and Dr. Devleena Mazumder-Shivakumar for their helpful communications and Drs. Sun Hur and Devleena-Shivakumar for reading and critical comments. Also, the author expresses gratitude to the National Institutes of Health (Grant MCB-0129568) for continual support and the National Science Foundation for the support they have provided.

8. References

(1) Bruice, T. C.; Benkovic, S. J. *Biochemistry* **2000**, *39*, 6267.
(2) Bruice, T. C. *Acc. Chem. Res.* **2002**, *35*, 139.
(3) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301*, 1196.
(4) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186.
(5) Bruice, T. C.; Kahn, K. *Curr. Opin. Chem. Biol.* **2000**, *4*, 540.
(6) Pauling, L. *Chem. Eng. News* **1946**, *24*, 1375.
(7) Pauling, L. *Nature* **1948**, *161*, 707.
(8) Page, M. I.; Jencks, W. P. *Proc. Natl. Acad. Sci., U.S.A.* **1971**, *68*, 1678.
(9) Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X. Y. *Angew. Chem., Int. Ed.* **2003**, *42*, 4872.
(10) Zhang, X. Y.; Houk, K. N. *Acc. Chem. Res.* **2005**, *38*, 379.
(11) Wolfenden, R.; Snider, M. J. *Acc. Chem. Res.* **2001**, *34*, 938.
(12) Bearne, S. L.; Wolfenden, R. *J. Am. Chem. Soc.* **1995**, *117*, 9588.
(13) Radzicka, A.; Wolfenden, R. *Science* **1995**, *267*, 90.
(14) Hur, S.; Bruice, T. C. *Proc. Natl. Acad. Sci., U.S.A.* **2002**, *99*, 9668.
(15) Gao, J. L. *Curr. Opin. Struct. Biol.* **2003**, *13*, 184.
(16) Wolfenden, R.; Lu, X. D.; Young, G. *J. Am. Chem. Soc.* **1998**, *120*, 6814.
(17) Radzicka, A.; Wolfenden, R. *J. Am. Chem. Soc.* **1996**, *118*, 6105.
(18) Bruice, T. C.; Bruice, P. Y. *J. Am. Chem. Soc.* **2005**, *127*, 12478.
(19) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(20) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(21) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(22) Monard, G.; Merz, K. M. *Acc. Chem. Res.* **1999**, *32*, 904.
(23) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(24) Aqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523.
(25) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(26) Reddy, S. Y.; Bruice, T. C. *J. Am. Chem. Soc.* **2003**, *125*, 8141.
(27) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J. Comput. Chem.* **2002**, *23*, 128.
(28) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(29) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489.
(30) Shea, J. E.; Onuchic, J. N.; Brooks, C. L. *Proc. Natl. Acad. Sci., U.S.A.* **2002**, *99*, 16064.
(31) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1988**, *9*, 327.
(32) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187.
(33) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239.
(34) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(35) Simonson, T.; Carlsson, J.; Case, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 4167.
(36) Lee, M. S.; Salsbury, F. R.; Olson, M. A. *J. Comput. Chem.* **2004**, *25*, 1967.
(37) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.
(38) Dominy, B. N.; Brooks, C. L. *J. Phys. Chem. B* **1999**, *103*, 3765.
(39) Ranaghan, K. E.; Ridder, L.; Szefczyk, B.; Sokalski, W. A.; Hermann, J. C.; Mulholland, A. J. *Org. Biomol. Chem.* **2004**, *2*, 968.
(40) Marti, S.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J.; Field, M. J. *J. Am. Chem. Soc.* **2001**, *123*, 1709.

(41) Bash, P. A.; Ho, L. L.; MacKerell, A. D.; Levine, D.; Hallstrom, P. *Proc. Natl. Acad. Sci., U.S.A.* **1996**, *93*, 3698.
(42) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.
(43) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.
(44) Cerjan, C. J.; Miller, W. H. *J. Chem. Phys.* **1981**, *75*, 2800.
(45) Lazaridis, T.; Tobias, D. J.; Brooks, C. L.; Paulaitis, M. E. *J. Chem. Phys.* **1991**, *95*, 7612.
(46) Chandrasekhar, J.; Shariffskul, S.; Jorgensen, W. L. *J. Phys. Chem. B* **2002**, *106*, 8078.
(47) Rajamani, R.; Naidoo, K. J.; Gao, J. L. *J. Comput. Chem.* **2003**, *24*, 1775.
(48) Friesner, R. A. *Proc. Natl. Acad. Sci., U.S.A.* **2005**, *102*, 6648.
(49) Meroueh, S. O.; Fisher, J. F.; Schlegel, H. B.; Mobashery, S. *J. Am. Chem. Soc.* **2005**, *127*, 15397.
(50) Zhang, Y. K.; Kua, J.; McCammon, J. A. *J. Am. Chem. Soc.* **2002**, *124*, 10572.
(51) Zheng, Y. J.; Ornstein, R. L. *J. Am. Chem. Soc.* **1997**, *119*, 648.
(52) Warshel, A. *Simulations of Chemical Reactions in Enzymes and Solutions*; John Wiley and Sons: New York, 1991.
(53) Shurki, A.; Warshel, A. In *Advances in Protein Chemistry: Protein Simulations*; Elsevier Academic Press: London, 2003; p 249.
(54) Watney, J. B.; Agarwal, P. K.; Hammes-Schiffer, S. *J. Am. Chem. Soc.* **2003**, *125*, 3745.
(55) Villa, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 7887.
(56) Lightstone, F. C.; Bruice, T. C. *J. Am. Chem. Soc.* **1996**, *118*, 2595.
(57) Lightstone, F. C.; Bruice, T. C. *J. Am. Chem. Soc.* **1997**, *119*, 9103.
(58) Bruice, T. C.; Lightstone, F. C. *Acc. Chem. Res.* **1999**, *32*, 127.
(59) Scheiner, S.; Lipscomb, W. N.; Kleier, D. A. *J. Am. Chem. Soc.* **1976**, *98*, 4770.
(60) Hur, S.; Bruice, T. C. *Proc. Natl. Acad. Sci., U.S.A.* **2002**, *99*, 1176.
(61) Lan, E. Y.; Bruice, T. C. *J. Am. Chem. Soc.* **1998**, *120*, 12387.
(62) Lau, E. Y.; Bruice, T. C. *J. Mol. Biol.* **1999**, *293*, 9.
(63) Torres, R. A.; Schiott, B.; Bruice, T. C. *J. Am. Chem. Soc.* **1999**, *121*, 8164.
(64) Radkiewicz, J. L.; Brooks, C. L. *J. Am. Chem. Soc.* **2000**, *122*, 225.
(65) Laitinen, T.; Rouvinen, J.; Perakyla, M. *Protein Eng.* **2000**, *13*, 247.
(66) Luo, J.; Bruice, T. C. *J. Am. Chem. Soc.* **2001**, *123*, 11952.
(67) Reddy, S. Y.; Kahn, K.; Zheng, Y. J.; Bruice, T. C. *J. Am. Chem. Soc.* **2002**, *124*, 12979.
(68) Schiott, B.; Bruice, T. C. *J. Am. Chem. Soc.* **2002**, *124*, 14558.
(69) Hur, S.; Bruice, T. C. *J. Am. Chem. Soc.* **2003**, *125*, 5964.
(70) Mazumder-Shivakumar, D.; Bruice, T. C. *Proc. Natl. Acad. Sci., U.S.A.* **2004**, *101*, 14379.
(71) Valina, A. L. B.; Mazumder-Shivakumar, D.; Bruice, T. C. *Biochemistry* **2004**, *43*, 15657.
(72) Sievers, A.; Beringer, M.; Rodnina, M. V.; Wolfenden, R. *Proc. Natl. Acad. Sci., U.S.A.* **2004**, *101*, 7897.
(73) Tobi, D.; Bahar, I. *Proc. Natl. Acad. Sci., U.S.A.* **2005**, *102*, 18908.
(74) Fedor, M. J.; Uhlenbeck, O. C. *Biochemistry* **1992**, *31*, 12042.
(75) Kuimelis, R. G.; McLaughlin, L. W. *Nucleic Acids Res.* **1995**, *23*, 4753.
(76) Slim, G.; Gait, M. J. *Nucleic Acids Res.* **1991**, *19*, 1183.
(77) Koizumi, M.; Ohtsuka, E. *Biochemistry* **1991**, *30*, 5145.
(78) Dahm, S. C.; Uhlenbeck, O. C. *Biochemistry* **1991**, *30*, 9464.
(79) Torres, R. A.; Bruice, T. C. *J. Am. Chem. Soc.* **2000**, *122*, 781.
(80) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.
(81) Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A. *Science* **1996**, *274*, 2065.
(82) Dahm, S. C.; Derrick, W. B.; Uhlenbeck, O. C. *Biochemistry* **1993**, *32*, 13040.
(83) Torres, R. A.; Himo, F.; Bruice, T. C.; Noodleman, L.; Lovell, T. *J. Am. Chem. Soc.* **2003**, *125*, 9861.
(84) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.
(85) Mazumder-Shivakumar, D.; Bruice, T. C. *Biochemistry* **2005**, *44*, 7805.
(86) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293.
(87) Li, H.; Hains, A. W.; Everts, J. E.; Robertson, A. D.; Jensen, J. H. *J. Phys. Chem. B* **2002**, *106*, 3486.
(88) Konecny, R.; Li, J.; Fisher, C. L.; Dillet, V.; Bashford, D.; Noodleman, L. *Inorg. Chem.* **1999**, *38*, 940.
(89) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, *238*, 415.
(90) Schutz, C. N.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 400.
(91) Feig, M.; Brooks, C. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217.
(92) Li, G. H.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 14521.
(93) Horenstein, B. A.; Schramm, V. L. *Biochemistry* **1993**, *32*, 9917.
(94) Horenstein, B. A.; Parkin, D. W.; Estupinan, B.; Schramm, V. L. *Biochemistry* **1991**, *30*, 10788.
(95) Mazumder, D.; Kahn, K.; Bruice, T. C. *J. Am. Chem. Soc.* **2002**, *124*, 8825.
(96) Ganem, B. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 936.
(97) Kuczera, K. *J. Comput. Chem.* **1996**, *17*, 1726.
(98) Hur, S.; Bruice, T. C. *Proc. Natl. Acad. Sci., U.S.A.* **2003**, *100*, 12015.
(99) Haynes, M. R.; Stura, E. A.; Hilvert, D.; Wilson, I. A. *Science* **1994**, *263*, 646.
(100) Zhang, X. D.; Zhang, X. H.; Bruice, T. C. *Biochemistry* **2005**, *44*, 10443.
(101) Jez, J. M.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. *Nat. Strut. Biol.* **2000**, *7*, 786.
(102) Jez, J. M.; Bowman, M. E.; Noel, J. P. *Biochemistry* **2002**, *41*, 5168.
(103) Jez, J. M.; Noel, J. P. *J. Biol. Chem.* **2002**, *277*, 1361.
(104) Hur, S.; Newby, Z. E. R.; Bruice, T. C. *Proc. Natl. Acad. Sci., U.S.A.* **2004**, *101*, 2730.
(105) Hur, S.; Bruice, T. C. *J. Am. Chem. Soc.* **2002**, *124*, 7303.
(106) Jencks, W. P. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1975**, *43*, 219.
(107) Bruice, T. C. *Annu. Rev. Biochem.* **1976**, *45*, 331.
(108) Vieille, C.; Zeikus, G. J. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 1.
(109) Hennig, M.; Darimont, B. D.; Jansonius, J. N.; Kirschner, K. *J. Mol. Biol.* **2002**, *319*, 757.
(110) Zhang, X.; Bruice, T. C. *Proc. Natl. Acad. Sci., U.S.A.* **2006**, *103*, 6148.
(111) Wu, J. C.; Santi, D. V. *J. Biol. Chem.* **1987**, *262*, 4778.
(112) Chen, L.; Macmillan, A. M.; Verdine, G. L. *J. Am. Chem. Soc.* **1993**, *115*, 5318.
(113) Gabbara, S.; Sheluho, D.; Bhagwat, A. S. *Biochemistry* **1995**, *34*, 8914.
(114) Roux, B. *Curr. Opin. Struct. Biol.* **2002**, *12*, 182.
(115) Go, N.; Noguti, T.; Nishikawa, T. *Proc. Natl. Acad. Sci., U.S.A.* **1983**, *80*, 3696.
(116) Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci., U.S.A.* **1983**, *80*, 6571.
(117) Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.* **1985**, *181*, 423.
(118) Brooks, B. R.; Janezic, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522.
(119) Tama, F.; Valle, M.; Frank, J.; Brooks, C. L. *Proc. Natl. Acad. Sci., U.S.A.* **2003**, *100*, 9319.
(120) Berendsen, H. J. C.; Hayward, S. *Curr. Opin. Struct. Biol.* **2000**, *10*, 165.
(121) Hinsen, K. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 417.
(122) Thomas, A.; Field, M. J.; Mouawad, L.; Perahia, D. *J. Mol. Biol.* **1996**, *257*, 1070.
(123) Cui, Q.; Li, G. H.; Ma, J. P.; Karplus, M. *J. Mol. Biol.* **2004**, *340*, 345.
(124) Tama, F.; Sanejouand, Y. H. *Protein Eng.* **2001**, *14*, 1.
(125) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905.
(126) Brink, J.; Ludtke, S. J.; Kong, Y. F.; Wakil, S. J.; Ma, J. P.; Chiu, W. *Structure* **2004**, *12*, 185.
(127) Almarsson, O.; Bruice, T. C. *J. Am. Chem. Soc.* **1993**, *115*, 2125.
(128) Rubach, J. K.; Plapp, B. V. *Biochemistry* **2003**, *42*, 2907.
(129) Sekhar, V. C.; Plapp, B. V. *Biochemistry* **1990**, *29*, 4289.
(130) Bahnson, B. J.; Park, D. H.; Kim, K.; Plapp, B. V.; Klinman, J. P. *Biochemistry* **1993**, *32*, 5503.
(131) Bahnson, B. J.; Colby, T. D.; Chin, J. K.; Goldstein, B. M.; Klinman, J. P. *Proc. Natl. Acad. Sci., U.S.A.* **1997**, *94*, 12797.
(132) Luo, J.; Kahn, K.; Bruice, T. C. *Bioorg. Chem.* **1999**, *27*, 289.
(133) Luo, J.; Bruice, T. C. *Proc. Natl. Acad. Sci., U.S.A.* **2002**, *99*, 16597.
(134) Luo, J.; Bruice, T. C. *Proc. Natl. Acad. Sci., U.S.A.* **2004**, *101*, 13152.
(135) Zhang, Y. K.; Kua, J.; McCammon, J. A. *J. Phys. Chem. B* **2003**, *107*, 4459.
(136) Thorpe, I. F.; Brooks, C. L. *J. Phys. Chem. B* **2003**, *107*, 14042.
(137) Garcia-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2001**, *114*, 9953.
(138) Vamvaca, K.; Vogeli, B.; Kast, P.; Pervushin, K.; Hilvert, D. *Proc. Natl. Acad. Sci., U.S.A.* **2004**, *101*, 12860.
(139) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. D. *Cell* **1994**, *76*, 357.
(140) Kumar, S.; Horton, J. R.; Jones, G. D.; Walker, R. T.; Roberts, R. J.; Cheng, X. *Nucleic Acids Res.* **1997**, *25*, 2773.
(141) Klimasauskas, S.; Szyperski, T.; Serva, S.; Wuthrich, K. *EMBO J.* **1998**, *17*, 317.
(142) O'Gara, M.; Horton, J. R.; Roberts, R. J.; Cheng, X. D. *Nat. Struct. Biol.* **1998**, *5*, 872.
(143) Estabrook, R. A.; Luo, J.; Purdy, M. M.; Sharma, V.; Weakliem, P.; Bruice, T. C.; Reich, N. O. *Proc. Natl. Acad. Sci., U.S.A.* **2005**, *102*, 994.
(144) Lockless, S. W.; Ranganathan, R. *Science* **1999**, *286*, 295.
(145) Luo, J.; Bruice, T. C. *Proc. Natl. Acad. Sci., U.S.A.* **2005**, *102*, 16194.

CR050283J