# Computational Studies of Enzyme−Catalyzed Reactions: Where Are We in Predicting Mechanisms and in Understanding the Nature of Enzyme Catalysis?

**ARTICLE** *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JANUARY 2002

Impact Factor: 3.3 · DOI: 10.1021/jp012017p



**3 AUTHORS**, INCLUDING:

Bernd Kuhn
Roche
**68** PUBLICATIONS **5,258** CITATIONS

SEE PROFILE

FEATURE ARTICLE

# Computational Studies of Enzyme-Catalyzed Reactions: Where Are We in Predicting Mechanisms and in Understanding the Nature of Enzyme Catalysis?

**Peter A. Kollman, Bernd Kuhn,*,† and Mikael Peräkylä‡**

*Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143*

*Received: May 24, 2001; In Final Form: October 31, 2001*

We review the state of the art of combined quantum mechanical/molecular mechanical calculations on enzyme systems and carry out a further analysis of the role of preorganization effects on enzyme catalysis. On the basis of our calculations, the free energy cost of aligning the functional groups for catalysis (the preorganization free energy) is responsible for a substantial fraction of the $10^{11}-10^{13}$ faster rates of the enzyme-catalyzed reactions for trypsin and catechol *O*-methyltransferase compared to analogous reference reactions in aqueous solution. Although the preorganization free energy contains a large entropic contribution, entropy effects on going from the enzyme−substrate noncovalent complex to the transition state for the reaction are expected to be generally small and not very different in enzyme and in solution. We attempt to give a critical analysis of both our approach and that of Warshel to elucidating the reason for large enzymatic rate enhancements.

## Introduction

The impressive advancements in computational quantum chemistry, in both calculations and the development of density functional theory, have been recognized by a Nobel Prize to Pople[1] and Kohn.[2] Important advances in the inclusion of solvent effects by continuum methods combined with both ab initio[3] and semiempirical[4] quantum chemical methods have also had tremendous impact on modeling phenomena of importance to chemistry.

Among the most important phenomena to which quantum chemical approaches can be applied are chemical reactions.[5] Quantum chemical calculations have played an essential role in understanding gas-phase reaction energetics and dynamics.[6] In combination with either explicit molecular mechanical (MM) models[7] or continuum models,[8] quantum mechanical (QM) approaches have also been used to simulate chemical reactions in condensed phases. Even more challenging has been the effort to find appropriate computational methods to accurately simulate chemical reactions in complex heterogeneous environments such as in enzymes. Such simulations should become ever more important as we continue into the genome era.

This article has two goals: first, to describe the current state of the art, both in terms of methodology and applications, of applying computational methods to enzyme-catalyzed reactions. The second goal of this article is to present our opinion of what makes an enzyme-catalyzed reaction so much more facile than the corresponding solution reaction. We discuss differences to the views of Warshel concerning preorganization effects in catalysis, which have recently appeared in the literature.[9]

**Simulating Chemical Reactions in Enzymes.** Simulations of chemical reactions in enzymes are usually directed at two goals: First, one would like computational methods to determine the free energy profiles of various proposed mechanisms accurately enough to predict the correct mechanism. Second, one would like to compare an enzyme-catalyzed reaction with the analogous solution reaction to elucidate the physical reasons for the catalytic effect. Being able to reproduce the free energy of activation, $\Delta G^{\ddagger}$, of the reaction in both environments is a useful validation of the accuracy of the simulation.

Why is it so difficult to reach the goals noted above? An enzyme-catalyzed reaction typically involves thousands of atoms. Although divide-and-conquer methods are used on systems of such size,[10] they are currently not accurate or efficient enough to meet the two aforementioned goals. Thus, one turns to methods that divide the system into those atoms that are affecting the electronic structure of the reaction center, which are treated quantum mechanically, and the remainder of the atoms, which are represented by classical molecular mechanical methods.

Such methods have their origin in the work by Warshel and Levitt,[11] who used combined semiempirical and molecular mechanical methods to simulate aspects of the reaction catalyzed by lysozyme and in the work of Jorgensen,[12] who combined gas-phase quantum mechanical methods and Monte Carlo/free energy (FE) calculations to study the solvation effect on the reaction

$$Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$$

To reach the goals noted above, one would like to be able to calculate $\Delta G^{\ddagger}$ to an accuracy of 5 kcal/mol or better. Given the number of quantum mechanical atoms (∼30−50) typically required to model enzymatic reactions, it has only been recently that computational power has allowed a sufficiently high level of theory to be applied to converge the quantum mechanical activation energies for such systems to the 5 kcal/mol required.

* To whom correspondence should be addressed. E-mail: bernd.kuhn@roche.com

† F. Hoffmann-La Roche Ltd., Basel, Switzerland.

‡ Department of Chemistry, University of Kuopio, Kuopio, Finland.

10.1021/jp012017p CCC: $22.00 
Published on Web 01/19/2002

Warshel's EVB (Empirical Valence Bond) approach[13] avoids this problem by calibrating an empirical valence bond model to reproduce the experimental $\Delta G^{\ddagger}$ found in the reference solution reaction and using the same parameters in modeling the enzymatic reaction.

In the case of simulating enzymatic reactions, the second challenge is to handle the "link atom" problem at the interface between the QM and MM regions. This is not a problem in simple organic reactions where there are no covalent bonds between the molecules that are reacting, which are represented quantum mechanically, and the solvent, which is represented with molecular mechanics. Recently, improved methods to handle the link atom problem have been described,[14−16] and various approaches to deal with this have been reviewed.[17] At this point, it is not clear how much the requirement for link atoms affects the accuracy of $\Delta G^{\ddagger}$, but it is probably reasonable to assume that as one moves the link atoms sufficiently far from the reacting atoms, the errors introduced should be small.

In addition to the quantum mechanical energies of the reacting atoms, a second major contribution to the calculated $\Delta G^{\ddagger}$ comes from the interaction of the classical mechanical atoms with the quantum mechanical ones. These interactions can be calculated with free energy perturbation (FEP) methods as done by Jorgensen[7,12] and involve of the order of $10^5-10^6$ energy evaluations (and, if performed with molecular dynamics rather than Monte Carlo, $10^5-10^6$ energy gradient evaluations). It is feasible to calculate this number of QM energies and gradients, however, only if one uses semiempirical quantum chemical calculations as done by Gao.[18] If one wants to apply more accurate ab initio methods, one is currently limited to a set of geometries along the quantum mechanically calculated reaction pathway and the classical atoms move in an unrestrained manner, as done by Jorgensen. An alternative is to only employ energy minimization to calculate the QM-MM interaction energies, which lead, in our opinion, to a poor approximation to the free energies. The use of continuum methods to estimate the free energy of interaction of the environment with the quantum mechanical atoms is not as straightforward in enzyme-catalyzed as in solution reactions, but should allow many fewer molecular mechanical atoms to be explicitly included in the calculations.

Because most of the existing QM/MM methods only sample the MM subsystem extensively, which allows for a calculation of the classical free energy contribution to the activation barrier, $\Delta G^{\ddagger}_{MM}$, it is important to further include entropic and vibrational enthalpy corrections to the calculated energy barrier for the QM system, $\Delta E_{QM}$.[19,20] To calculate these rigorously in the enzyme environment is difficult, but in our view, likely to lead to similar contributions for different mechanisms or in comparing enzyme and solution reactions, and thus to be less essential to reaching the two goals mentioned above. Finally, the question remains, given the multitude of lower energy conformations for the molecular groups in the neighborhood of reacting atoms in solution or in the enzyme, whether one can really speak of a single transition state structure, as often exists in gas-phase reactions.

In summary, the three main bottlenecks in calculating $\Delta G^{\ddagger}$ to $\pm 5$ kcal/mol—the accuracy of the quantum mechanical model, the errors inherent in the link atom approach, and the need for sufficient sampling to accurately calculate the free energy of the QM-MM interactions—are all problems that can be solved with sufficient computer power. Although we are some distance from solving them in a general way, there are a number of exciting applications to real enzyme systems, which suggest that we are getting closer to the goals noted above. We next present a selected and far from exhaustive set of examples from recent work from various labs to support this statement.

**Recent Applications of QM/MM Approaches to Enzymatic Reactions.** Because Warshel's EVB−FEP approach does not rely on computationally demanding ab initio calculations, given that one can use experimental data to calibrate the $\Delta G^{\ddagger}$ for the corresponding solution reaction, this method has been applied to enzyme systems for ~15 years. Among the authors' favorite applications are those to staph nuclease,[21] to mechanisms of phosphate hydrolysis,[22] and to the enzyme-catalyzed reaction of orotidine 5′-phosphate decarboxylase (ODCase),[23] which has been described as the most efficient enzyme known, with a difference in activation barrier between enzymatic and nonenzymatic reaction, $\Delta\Delta G^{\ddagger}$, of 24 kcal/mol. A general message from Warshel's work is that the predominant contribution to the $\Delta\Delta G^{\ddagger}$ in various enzyme systems is the electrostatic interaction free energy, which is more favorable in the enzyme than in water because the charges and dipoles in the enzyme are preorganized prior to substrate binding.

Gao has used fully coupled QM/MM and potential of mean force calculations to calculate the free energies in tyrosine phosphatase[18] and ODCase,[24] in both cases achieving impressive agreement with experiment. He has suggested "ground-state destabilization" between the substrate and an enzyme carboxyl group as the basis for the large $\Delta\Delta G^{\ddagger}$, an interpretation that Warshel does not agree with;[23] in his calculations, the greater contribution to $\Delta\Delta G^{\ddagger}$ came from transition state stabilization. Field[25] has analyzed the factors responsible for enzyme catalysis in chorismate mutase using QM/MM approaches and he and Gao have developed a new link atom method.

Aqvist has used EVB−FEP methods to study the free energies along the hydrolysis catalyzed by tyrosine phosphatases.[26] There is some difference in what he and Gao[18] find as the optimum charge state of the phosphate along the hydrolysis pathway, but both calculate a $\Delta G^{\ddagger}$ in impressive agreement with experiment.

In the case of citrate synthase, Mulholland and Karplus[27] and Donini et al.[28] have found an enediolate intermediate to be more stable than an enediol, which addresses a controversial proposal about the importance of matched $pK_a$s[29] and low-barrier hydrogen bonds[30] in enzyme catalysis.

Cui and Karplus[31] have recently studied the second step in the reaction of triosephosphate isomerase (TIM) using QM/MM methods and have concluded that, in the enediolate intermediate, intermolecular proton transfer is more favorable than the intramolecular pathway suggested by Alagona et al.[32] However, only energy minimization was used to determine the QM-MM interaction energies. A useful development in that study was the use of reduced charges on the surrounding atoms, which were derived by evaluating the electrostatic potentials at these atoms with a continuum model. This alleviates the difficulties of calculating large long-range electrostatic effects[28] from the many charged groups in typical proteins.

Recent studies by Yang and co-workers[33,34] have suggested a new approach to the link atom problem and a new approach to interactive QM/MM optimization that allows a DFT/6-31G* model to be used for the QM atoms, followed by free energy calculations. This method has led to useful insights into catalysis by enolase.

Recently, we have shown how to derive quantitative estimates for the free energy of enzyme vs aqueous solution stabilization using a QM-FE approach that builds on the earlier work of Jorgensen.[35] In Table 1, we summarize the calculated and experimental free energies of activation in both enzyme and in

**TABLE 1: QM-FE Calculated and Experimental Free Energy of Activation $\Delta G^{\ddagger}$ [a,b]**

| system | $\Delta G^{\ddagger}$ (enzyme)[c] | $\Delta\Delta G^{\ddagger}$ (solution-enzyme)[d] |
|---|---|---|
| trypsin | 16 (15) | 17 (18) |
| catechol *O*-methyltransferase | 21−24 (18) | 14−18 (15) |

[a] Energies in kcal/mol. [b] Experimental values in parentheses. [c] Free energy of activation in enzyme. [d] Difference in free energy of activation between solution and enzyme.

solution for trypsin[36] and catechol *O*-methyltransferase (COMT),[20] demonstrating the ability of the QM-FE methodology to calculate activation free energies in reasonable agreement with experiment and without the use of any empirical parameters.

**Factors that Determine the Efficiency of Enzyme Catalysis.** What are the factors that lead to a catalyzed reaction having a significantly lower $\Delta G^{\ddagger}$ than the corresponding solution reaction? Although there have been special features of enzymes suggested to be important in speeding up the enzyme-catalyzed reaction such as orbital steering,[37] low-barrier hydrogen bonds,[30] or matched p$K_a$s,[29] the validity and usefulness of such special concepts are unproven.[38,39] On the basis of the fact that many of the studies mentioned in the previous section reasonably reproduce $\Delta G^{\ddagger}$ for the enzyme-catalyzed reaction without including these concepts, they are unlikely to be major factors in lowering the free energy of activation for enzyme-catalyzed reactions.

Two major factors leading to a large $\Delta\Delta G^{\ddagger}$ are first, the alignment of the enzyme groups that can interact with the transition state for the rate-limiting reaction step more favorably than found in water, and second, the preorganization free energy of the reacting groups that is significantly smaller in the enzyme than found in the solution reaction because it is absorbed in the binding free energy for formation of the Michaelis complex.

Warshel and Levitt noted that for the transition state stabilization the electrostatic term is predominant and much more important than the van der Waals contribution, something that has been further supported in many subsequent studies by Warshel and co-workers. This factor is likely to be a major contributor to $\Delta\Delta G^{\ddagger}$ for those enzymatic reactions where the transition state is more highly charged or more polar than the ground state. However, as suggested by Kuhn and Kollman,[20] when the enzymatic reaction involves a less polar transition state than ground state, the enzyme is able to increase $\Delta\Delta G^{\ddagger}$ by interacting less strongly electrostatically with the ground (and transition state) than water does. Presumably, the enzyme uses its ability to interact strongly with the substrate using van der Waals and hydrophobic effects, something that water cannot do, to compensate for weaker electrostatic interaction with the substrate than found in water.

The key role of preorganization of the substrate in enabling a large $\Delta\Delta G^{\ddagger}$ was first analyzed by Page and Jencks.[40] A qualitative picture of this factor is simply that the enzyme binds the substrate such that its reactive groups are well aligned to react without significant further conformational change, whereas in the corresponding solution reaction, it costs significant free energy to align the reactive groups for catalysis. Bruice and co-workers have suggested the concept of near attack conformation (NAC) as a way to estimate how efficiently the enzyme actually achieves optimal alignment.[41] The enzymes analyzed by Bruice typically achieve a NAC 10% of the time in unrestrained molecular dynamics simulations, suggesting that it costs only ~1.4 kcal/mol of preorganization free energy in the enzyme-catalyzed reaction.

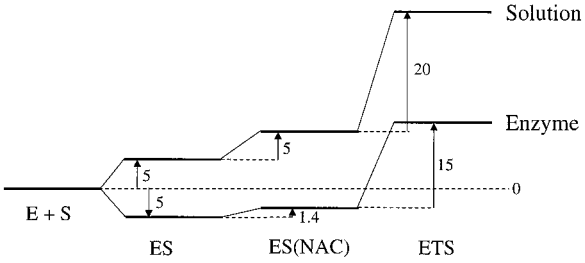


**Figure 1.** Schematic diagram of a representative enzyme-catalyzed and reference solution reaction in which the preorganization free energy cost is a total of 10 kcal/mol in solution (5 kcal/mol to form ES and a further 5 kcal/mol to form ES(NAC)) and 1.4 kcal/mol in the enzyme. The value from the enzyme is an average of the two values (0.4 and 2.4 kcal/mol) noted in the text. The sum of the quantum mechanical and free energy perturbation difference between the correctly oriented noncovalent complex, ES(NAC), and the transition state, ETS, is 15 kcal/mol in the enzyme and 20 kcal/mol in solution. E = enzyme, S = substrate, ES = noncovalent enzyme−substrate complex, ES(NAC) = noncovalent enzyme−substrate complex with groups oriented for catalysis, and ETS = transition state for reaction.

**Preorganization Free Energy in Enzyme Catalysis.** The Michaelis−Menten kinetic model for an enzymatic reaction involves Scheme 1, where E refers to the enzyme, S to the

**SCHEME 1**

$$E + S \underset{}{\overset{K_M}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} ETS \longrightarrow E + P$$

substrate, ES to the enzyme−substrate noncovalent complex, ETS to the transition state for the reaction, and P to the product. This model is useful for comparison with a noncatalyzed reaction in aqueous solution to investigate the basis for the enzymatic rate enhancement. Clearly, one must not forget the possibility that the mechanism for the reaction may be different in water and in the enzyme, and, hence, the computed rate constant for the solution reaction might not be directly comparable to experiment.

What does it mean to compare the free energies of the process E + S → ES in an enzyme-catalyzed and solution reaction? In the enzyme-catalyzed reaction, this corresponds to the noncovalent binding of substrate into the enzyme active site. In the solution reaction, we believe that the appropriate model for ES is to put the fragments in which chemical bonding will be changing in the orientation they have in the enzyme and carry out the same reaction as in the enzyme-catalyzed reaction, with the environment water molecules rather than the enzyme. Warshel's definition[9,42,43] of ES in the solution reaction is different in that he creates a cage of reactant groups where the reactants are only close to each other, but not in an enzyme-like configuration. Moreover, in our opinion, it is conceptually useful to distinguish, as we do in Figure 1, the structure ES, which corresponds to the enzyme substrate noncovalent complex, or, in the case of the solution reaction, a structure with the reactive groups in noncovalent contact, from the structure ES(NAC), in which the reactive groups are in an appropriate geometry for a facile chemical reaction to take place.

The main divergence of opinion between Warshel and us is whether it is needed to include the preorganization free energy of assembling the reactants in an enzyme-like configuration into ES of the solution reaction. We believe that it is essential if one wants to compare one's calculated energy profiles with experimental data for both solution and enzymatic environments. Warshel believes that the free energy cost of bringing the reactants to ES is RT *ln* 55 = 2.4 kcal/mol at room temperature,[9] which is the free energy cost, in a solution that is 55M in water,

of bringing two specific molecules next to each other. The use of this "cratic" approach to estimating the translational free energy of molecular association has been criticized by Holtzer[44] and a more rigorous statistical thermodynamical analysis of the preorganization free energy is provided by Gilson et al.[45] We have used the word "cratic" to mean "preorganization" in references[20,35,36], whereas the standard use of "cratic", as noted by Holtzer,[44] is to refer to relative concentration dependent free energies. In his work, Warshel uses the "cratic" free energy as described by Holtzer for the process E + S → ES in solution. As discussed below, this "cratic" free energy of 2.4 kcal/mol is probably not a bad estimate for forming ES for typical polar neutral molecules in solution, but a very poor estimate for the free energy of forming ES(NAC) in solution.

We believe that Warshel's approach leads to a significant underestimate of the free energy cost of prealigning the groups for catalysis in solution. There are a number of reasons for this. First, the simple volume analysis by Warshel assumes that bringing the reactants together to form ES is an entirely entropic term, which may not be too bad an assumption for the three groups in trypsin (imidazole, methanol, and *N*-methylacetamide),[36] but a poor one for the cation *S*-adenosylmethionine (SAM) and the catechol anion in the COMT-catalyzed reaction.[20] This is because the free energy for bringing an ion pair together in water implies losing considerable solvation free energy, which can be costly. One can evaluate this free energy cost by carrying out potential of mean force calculations as a function of distance and find that about 5 kcal/mol are needed to bring these two groups, SAM and catecholate, together in solution.

Second, the simple volume estimate assumes all relative orientations of the two molecules that have been brought together are equally favorable for the chemical reaction to proceed. Consider hydrogen bonding between imidazole and methanol. Thermal motion in solution would lead to hydrogen bonding between the methanol O−H and the imidazole N that would range from 2.8 to 4.0 Å, with the shorter distances being more probable. However, proton transfer from methanol to imidazole is required for the reaction to proceed and this is exquisitely sensitive to the distance between both moieties. As an example of this exquisite sensitivity, we showed in models of triosephosphate isomerase that increasing the distance between proton donor and acceptor by only 0.3 Å could lead to a 4 kcal/mol higher barrier for proton transfer.[46] Thus, when one carries out calculations on the process of proton transfer, using an optimum distance for the H-bond length, one must evaluate the free energy cost of constraining the H-bond distance to a rather narrow range. By determining the force constants required to achieve a certain standard deviation for the H-bond distance one can use the analytical formulas of Hermans and Wang[47] to calculate the free energy cost for such a restraint. H-bond energies are quite sensitive to both distance and angle and we have used a standard deviation of 0.2 Å for the distance and 20 degrees for the bond angle as our criteria for a "reactive" orientation of the two groups. A similar analysis was also done for the relative orientation of *N*-methylacetamide to the methanol-imidazole complex.

There is no simple way to choose which standard deviations to use and Warshel has criticized[9] our use of 0.2 Å and 20 degrees as criteria for a reactive configuration. However, given the fact that chemical bond changes, in contrast to weak noncovalent interactions, are very sensitive to geometry, we feel that our assumptions of these values are reasonable. We emphasize that those values are standard deviations and the molecular dynamics can of course go outside these ranges in the Hermans/Wang molecular dynamical analysis.[47] In our study of trypsin,[36] we calculated that the free energy of aligning the three reacting groups was 11.4 kcal/mol using the Hermans/Wang method and 8.9 kcal/mol just using a simple geometrical criteria, with both distance and angular orientations taken into account. Thus, the important point is that both give the same order of magnitude for the free energy of preorganization.

**TABLE 2: Average Geometrical Parameters for the Trypsin-Substrate ES Complex during Molecular Dynamics**

| geometrical parameters [a] | average | standard deviation |
|---|---|---|
| R (N...H)(Å)[b] | 1.98 | 0.25 |
| $\theta$ (N...H−O) (degrees)[c] | 154.3 | 13.9 |
| R (O...C) (Å)[d] | 2.99 | 0.14 |
| $\theta$ (O...C−O) (degrees)[e] | 89.5 | 7.1 |

[a] See refs 36, 48 for simulation details. Averages for a simulation time of 50 ps. [b] Distance between imidazole N and serine O−H. [c] H-bond angle involving imidazole N and serine O−H. [d] Distance between serine oxygen and carbonyl carbon of scissile amide bond. [e] Angle between serine oxygen and carbonyl vector of scissile amide bond.

Warshel has suggested[9] that if a reaction is not concerted, one only has to consider the preorganization free energy one molecule at a time. We think that this is not inconsistent with our approach using the sum of the free energies of preorganization of imidazole/methanol and methanol/*N*-methylacetamide as the total preorganization free energy. Whether the proton transfer from methanol to imidazole and attack by methoxide on *N*-methylacetamide is fully concerted, if the *N*-methylacetamide is not in the correct position to be attacked by the methoxide, the proton will simply return to the methanol from the imidazole. Thus, one must consider the free energy for all three groups to be properly aligned for a reaction, which is achieved in the enzyme by its sequence and structure.

One of the assumptions we have made so far in our studies is that the preorganization free energy in the enzyme is compensated by favorable protein−ligand interactions. This assumes that the enzyme is maintaining the reactive groups in the correct position for reaction to occur. How reasonable is this assumption? That it is not unreasonable is supported by molecular dynamics simulations of Bruice on various enzymes.[41] He finds that one achieves a near attack configuration in the range of 10% of the time during a typical molecular dynamics simulation. In addition, as shown in Figure 1, which is a representative free energy diagram for the two systems we have studied, for a typical enzyme−substrate interaction, the free energy has to decrease for the reactant to be favorably bound in the active site. This is because hydrophobic and other nonbonded effects strongly stabilize the ES complex over the separate species and this stabilization overcomes the increase in preorganization free energy in bringing E and S together. If the process of going from a typical ES complex structure to ES(NAC) occurs 10% of the time, as suggested by Bruice's various studies, the free energy of ES would be raised by only RT ln 10, or approximately 1.4 kcal/mol at room temperature, to put it in position to react.

To support the assumption that ES(NAC) has a similar free energy to ES for a typical specific enzyme−substrate noncovalent complex, we have carried out further molecular dynamics simulations on the substrate-trypsin complex, using the same substrate and protocol as in references.[36,48] The geometrical properties of the reactive groups in this ES complex are summarized in Table 2. The average structural properties of the trypsin−substrate complex suggest that it is in a near ideal position to react, with no constraints to maintain the substrate

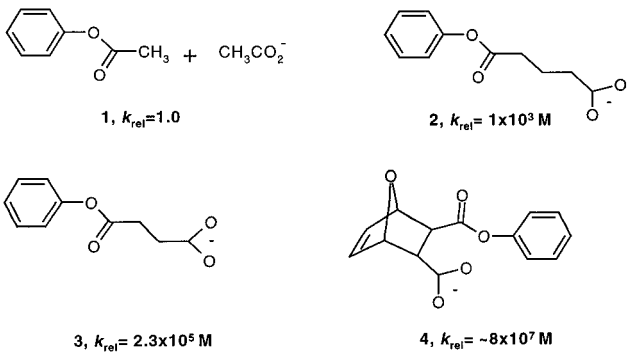


**Figure 2.** Molecules used in reactions of anhydride formation. See ref 49.

in place. The serine oxygen—amide carbonyl carbon distance is 3.0 Å and the angle of attack is 90 degrees, both nearly ideal and with small standard deviations. The serine O—H—imidazole N hydrogen-bond distance is (1.98 ± 0.25) Å, a little larger than ideal and the angle for this hydrogen bond is (154 ± 14) degrees, which is close to ideal, given the fact that a value of 180 degrees for a hydrogen bond is made less probable by its small intrinsic geometrical probability. Forcing the hydrogen-bond distance to be (1.9 ± 0.2) Å, as we did in our methanol-imidazole complex in solution[36] costs 2.4 kcal/mol of free energy, but using the unconstrained trajectory and determining what percentage of the time the hydrogen bond distance is <1.9 Å reveals that it is so 46% of the time. Thus, the free energy to force the system into this NAC conformation (H-bond distance <1.9 Å) is only 0.4 kcal/mol. These data further support the assumption that the preorganization free energy for the reaction, which is ~10 kcal/mol for the solution reaction, can be assumed to be much smaller for an enzyme that binds its substrate in a NAC.

One final result of relevance is the analysis of Peräkylä and Kollman[49] on the intra- and intermolecular anhydride reactions studied also by Bruice.[50] In these molecules, the rate of reaction varies by approximately $10^8$, with the slowest reaction the intermolecular one and the most rapid reaction the intramolecular one involving the molecule with the geometry most constrained toward a NAC. By carrying out both high level ab initio and free energy calculations, we were able to quite accurately reproduce the relative free energies of reactions 1−4 (Figure 2). However, the relative free energies of activation of reactions 1 and 2 were correctly reproduced only by including the preorganization free energy contribution of 5 kcal/mol for the intermolecular reaction 1. The absolute free energies of activation were systematically underestimated, and we do not know the reason for this.

The main point of the above analyses is that the preorganization free energy for forming a reactive configuration, ES(NAC), for a typical solution reaction is significantly larger than that in the enzyme-catalyzed reaction. It is approximately 10 kcal/mol for the two reactions we have studied. The remaining 5 kcal/mol for the $\Delta\Delta G^{\ddagger}$ for trypsin and COMT comes from the first factor noted above, the relative interaction free energies of the enzyme groups with ETS. The preorganization free energy in the enzyme-catalyzed reaction is significantly smaller than in solution, due to the fact that the enzyme uses its various hydrophobic and polar groups to bind the reactive molecules in geometries favorable for the reaction to proceed. As is clear in Figure 1, as long as the geometry for facile reaction in the enzyme is achieved a reasonable fraction of the time, most of the preorganization free energy price is more than compensated by these favorable noncovalent interactions.

We now turn to the free energies for the process ES(NAC) → ETS, which we have estimated using a combination of high level ab initio and free energy perturbation calculations, the latter in both solution and enzyme. By doing the free energy perturbation, entropic effects due to solvent and nonreactive enzyme group reorganizations are included in the free energies. What about other enthalpic and entropic contributions from the reactive groups, whose quantum mechanical energies along the pathway have been evaluated? If one has unconstrained reactants, one can use normal-mode analysis to estimate the entropic, thermal, and zero-point energy changes in going from the reactant complex to the transition state. We have done this for the COMT-catalyzed reaction, using Bruice's geometry[41] for the reactant and transition state, since, in our study, we had used a constrained geometry that fits into the enzyme active site.[20] In our view, it is a reasonable assumption that the unconstrained reactant and transition state would be representative of these structures. The effect of including the above-mentioned free energy corrections is to increase the barrier to reaction by only about 1 kcal/mol.

Warshel has carried out some analyses of the activation entropy as well. In the first paper by Strajbl et al.,[43] an entropic ($-T\Delta S$) contribution for the process ES → ETS of 11.7 kcal/mol was calculated for the solution reaction. In another paper using calculations on a model for general base-catalyzed methanolysis of formamide, they argued that "it seems the neglect of $\Delta S^{\ddagger}$ coincidentally cancels the overestimate of the activation energies because the calculated $\Delta G^{\ddagger}$ agrees well with the corresponding observed value".[42] Finally, in a more recent paper on a model for the subtilisin (serine protease)-catalyzed reaction, Villa et al. found $-T\Delta S$ values for the ES → ETS process in water of about 5 and 2.5 kcal/mol in the enzyme, noting that the differences between these values is "much smaller than previously thought".[9] We disagree with the idea that people thought that this process, ES → ETS, involved a large differential entropy change. In our view, what Page and Jencks[40,51] were focusing on was the process E+S → ES(NAC) that we discussed above. Once one gets to a configuration where the reactants are in place, as they are in ES(NAC), the difference in free energy due to solute entropy effects is very likely to be similar in enzyme and in solution. One must applaud Villa et al.[9] and Strajbl et al.'s[43] bravery in attempting to separately calculate $\Delta S$. Given the larger inherent errors and challenge of computing $\Delta S$, it seems reasonable and more expedient to calculate environmental free energy effects directly using free energy calculations and to estimate solute entropy and vibrational enthalpy effects from the model for ES(NAC) and ETS.

In summary, our calculations have found that the preorganization free energy contribution that differentiates enzyme and solution reactions is in the range of 10 kcal/mol, which allows this term to rationalize approximately $10^7$ of the enzymatic rate enhancement over the corresponding solution reactions. We do not know whether this preorganization free energy will be so large for all reactions, but it can clearly help explain why enzyme-catalyzed reactions which involve radical processes and, thus, are not likely to involve differential electrostatic stabilization are faster than the corresponding solution reactions. It also nicely puts in context and is consistent with the large range in rates ($10^8$) found between the various anhydride reactions[49,50] again emphasizing the close relationship between intramolecular cyclizations and enzymatic catalysis. We stress that this preorganization free energy is not always contained in either the enthalpy or the entropy. This is because, as has been noted by Bruice,[50] the free energy cost of restricting a molecule to a NAC

can be either enthalpic or entropic. Thus, it is often misleading to focus on one of these two components.

## Conclusions

In this review, we have focused on the exciting results that theoretical calculations by others and ourselves, using combined quantum mechanical and molecular mechanical models, have been able to achieve for enzyme-catalyzed reactions and the reference solution reactions. Our studies in simulating enzyme-catalyzed reactions for trypsin and catechol *O*-methyltransferase have emphasized the importance of the preorganization free energy in achieving this agreement for the $\Delta\Delta G^{\ddagger}$ to the solution reaction and in correctly reproducing the relative activation free energies of intra- and intermolecular anhydride formation. The concept of the preorganization free energies has been noted many times before,[50−52] but the advance has been to be able to more quantitatively calculate such terms and, thus, make more meaningful contact with experiment (Table 1). Although it is likely that the relative reactant entropy part of the free energies is small in going from ES(NAC) to ETS, the place where solution and enzyme-catalyzed processes differ is that solution reactions will typically have a significant free energy cost in going from E + S to ES(NAC) (Figure 1), whereas in the enzyme reaction, noncovalent interactions compensate for these terms, leading to a net lowering of free energy and little effect on the calculated free energies of activation in going from ES to ES(NAC). There is disagreement in the literature on these points (ref 61 in 42; footnote on p 11 899, 2nd column in 9), which hopefully will be sorted out in the future.

We conclude by commenting on the tremendous enzymatic rate enhancement observed in orotidine 5′-phosphate decarboxylase, where the enzyme catalyzes decarboxylation $10^{17}$ faster than occurs in solution. This is a reaction that is intramolecular, so the free energy cost of preorganization is likely to be small. Although there is some support for "ground-state destabilization" in this reaction,[24] we and Warshel[23] find this unlikely. In our view,[53] it is more likely that chemical catalysis by the enzyme is involved, which makes the mechanism fundamentally different than found in solution decarboxylation. This is, of course, another way in which enzymes can achieve impressive rate enhancements.

**Acknowledgment.** We acknowledge useful discussions with Jed Pitera and Irwin D. Kuntz and are grateful for research support from the NIH through Grant No. GM-29072.

## References and Notes

(1) Pople, J. A. *Rev. Mod. Phys.* **1999**, *71*, 1267.
(2) Kohn, W. *Rev. Mod. Phys.* **1999**, *71*, 1253.
(3) Tomasi, J.; Cammi, R.; Mennucci, B. *Int. J. Quantum Chem.* **1999**, *75*, 783.
(4) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.
(5) Timoshkin, A. Y.; Bettinger, H. F.; Schaefer, H. F. *J. Phys. Chem. A* **2001**, *105*, 3240.
(6) Corchado, J. C.; Truhlar, D. G.; Espinosa-Garcia, J. *J. Chem. Phys.* **2000**, *112*, 9375.
(7) Severance, D. L.; Jorgensen, W. L. *ACS Symposium Ser.* **1994**, *568*, 243.
(8) Patterson, E. V.; Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **2001**, *123*, 2025.
(9) Villa, J.; Strajbl, M.; Glennon, T. M.; Sham, Y. Y.; Chu, Z. T.; Warshel, A. *Proc. Nat. Acad. Sci.* **2000**, *97*, 11 899.
(10) van der Vaart, A.; Suarez, D.; Merz, K. M. *J. Chem. Phys.* **2000**, *113*, 10 512.
(11) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(12) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 3049.
(13) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; Wiley: New York, 1991.
(14) Assfeld, X.; Rivail, J. L. *Chem. Phys. Lett.* **1996**, *263*, 100.
(15) Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290.
(16) Amara, P.; Field, M. J.; Alhambra, C.; Gao, J. L. *Theo. Chem. Acc.* **2000**, *104*, 336.
(17) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.
(18) Alhambra, C.; Wu, L.; Zhang, Z. Y.; Gao, J. L. *J. Am. Chem. Soc.* **1998**, *120*, 3858.
(19) Moliner, V.; Williams, I. H. *Chem. Commun.* **2000**, *19*, 1843.
(20) Kuhn, B.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 2586.
(21) Aqvist, J.; Warshel, A. *J. Am. Chem. Soc.* **1990**, *112*, 2860.
(22) Aqvist, J.; Kolmodin, K.; Florian, J.; Warshel, A. *Chem. Biol.* **1999**, *6*, R71.
(23) Warshel, A.; Strajbl, M.; Villa, J.; Florian, J. *Biochemistry* **2000**, *39*, 14 728.
(24) Wu, N.; Mo, Y. R.; Gao, J. L.; Pai, E. F. *Proc. Nat. Acad. Sci.* **2000**, *97*, 2017.
(25) Marti, S.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J.; Field, M. J. *J. Am. Chem. Soc.* **2001**, *123*, 1709.
(26) Kolmodin, K.; Aqvist, J. *Febs Lett.* **1999**, *456*, 301.
(27) Mulholland, A. J.; Lyne, P. D.; Karplus, M. *J. Am. Chem. Soc.* **2000**, *122*, 534.
(28) Donini, O.; Darden, T.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 12 270.
(29) Gerlt, J. A.; Gassman, P. G. *Biochemistry* **1993**, *32*, 11 943.
(30) Frey, P. A.; Whitt, S. A.; Tobin, J. B. *Science* **1994**, *264*, 1927.
(31) Cui, Q.; Karplus, M. *J. Am. Chem. Soc.* **2001**, *123*, 2284.
(32) Alagona, G.; Ghio, C.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 9855.
(33) Liu, H. Y.; Zhang, Y. K.; Yang, W. T. *J. Am. Chem. Soc.* **2000**, *122*, 6560.
(34) Zhang, Y. K.; Liu, H. Y.; Yang, W. T. *J. Chem. Phys.* **2000**, *112*, 3483.
(35) Kollman, P. A.; Kuhn, B.; Donini, O.; Peräkylä, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72.
(36) Stanton, R. V.; Perakyla, M.; Bakowies, D.; Kollman, P. A. *J. Am. Chem. Soc.* **1998**, *120*, 3448.
(37) Dafforn, A.; Koshland, D. E. *Proc. Nat. Acad. Sci.* **1971**, *68*, 2463.
(38) Bruice, T. C. *Cold Spring Harb. Symp. Quantum Biol.* **1972**, *36*, 21.
(39) Warshel, A.; Papazyan, A. *Proc. Nat. Acad. Sci.* **1996**, *93*, 13 665.
(40) Page, M. I.; Jencks, W. P. *Proc. Nat. Acad. Sci.* **1971**, *68*, 1678.
(41) Lan, E. Y.; Bruice, T. C. *J. Am. Chem. Soc.* **1998**, *120*, 12 387.
(42) Strajbl, M.; Florian, J.; Warshel, A. *J. Am. Chem. Soc.* **2000**, *122*, 5354.
(43) Strajbl, M.; Sham, Y. Y.; Villa, J.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **2000**, *104*, 4578.
(44) Holtzer, A. *Biopolymers* **1995**, *35*, 595.
(45) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047.
(46) Alagona, G.; Ghio, C.; Kollman, P. A. *J. Mol. Biol.* **1986**, *191*, 23.
(47) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707.
(48) Peräkylä, M.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 3436.
(49) Peräkylä, M.; Kollman, P. A. *J. Phys. Chem. A* **1999**, *103*, 8067.
(50) Bruice, T. C.; Lightstone, F. C. *Acc. Chem. Res.* **1999**, *32*, 127.
(51) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; Dover: New York, 1986.
(52) Menger, F. M. *Acc. Chem. Res.* **1993**, *26*, 206.
(53) Lee, T. S.; Chong, L.; Chodera, J.; Kollman, P. A. *J. Am. Chem. Soc.*, submitted.