Combined *ab initio* and Free Energy Calculations To Study Reactions in Enzymes and Solution: Amide Hydrolysis in Trypsin and Aqueous Solution

Robert V. Stanton, Mikael Peräkylä, Dirk Bakowies, and Peter A. Kollman*

Contribution from the Department of Pharmaceutical Chemistry, Box 0446, University of California at San Francisco, San Francisco, California 94143

Received August 6, 1997

Abstract: We present a new more general way to combine *ab initio* quantum mechanical calculations with classical mechanical free energy perturbation approach to calculate the energetics of enzyme-catalyzed reactions and the same reaction in solution. This approach, which enables enzyme and solution reactions to be compared without the use of empirical parameters, is applied to the formation of the tetrahedral intermediate in trypsin, but it should be generally applicable to any enzymatic reaction. Critical to the accurate calculation of the reaction energetics in solution is the estimate of the free energy to assemble the reacting groups, where the approach recently published by Hermans and Wang *(J. Am. Chem. Soc.* **1997**, *119*, 2707) was used. A central aspect of this new approach is the use of the RESP protocol to calculate the charge distribution of structures along the reaction pathway, which enables us to circumvent problems in partitioning the charge across a residue that is being divided into QM and MM parts. The classical mechanical free energy calculations are implemented with two different approaches, "Cartesian mapping" and "flexible FEP". The similarity of the results found by using these two approaches supports the robustness of the calculated free energies. The calculated free energies are in quite good agreement with available experimental data for the activation free energies in the enzyme and aqueous phase reactions.

Introduction

Enzyme catalysts are the machinery that drive biology and, therefore, have received much attention from the scientific community. The study of these catalysts has proven to be an area of prolific interdisciplinary collaboration. Biochemists and organic chemists have long studied enzyme reaction mechanisms (e.g., what is the actual course of molecular events and how can these same reactions be mimicked by smaller host−guest systems). Additionally structural biologists have contributed detailed snapshots of enzyme structures in many of the different states along the reaction pathway,[1] but usually no definitive evidence can be derived for mechanisms from experiment alone.[2] The role of physical chemists/theoreticians[3,4] has been to provide a link between structures, provided by X-ray crystallography or NMR, and function as determined by experiments on enzyme-catalyzed reactions.[5] This structure function relationship can, in principle at least, be determined in a quantitative fashion by using the system's three-dimensional geometry to model the free energy along putative reaction pathways.

Theoretical structure/function studies have proven very challenging for a variety of reasons. Foremost among the difficulties presented by enzyme systems is their size; enzymes are very large molecules and molecular dynamics studies on them are time consuming and difficult when the simplest potential functions are used. Even with a "perfect" classical force field that accurately describes the energy of a system, it would remain enormously difficult to adequately sample the conformational space of an enzyme or enzyme−substrate noncovalent complexes. An additional difficulty is that classical force fields are unable to simulate bond breaking and bond formation processes. The study of chemical reactions requires quantum mechanical (QM) methods. Unfortunately, the computational time required for QM calculations rises steeply with the number of atoms and electrons in the system, and therefore it is very difficult to apply QM calculations to large enzyme systems.

In 1976, Warshel and Levitt[6] introduced the concept of combining quantum mechanical (QM) and molecular mechanical (MM) methods. This approach limits the quantum mechanical description to the reaction center and uses a computationally efficient classical treatment for the remainder of the molecule. Various other QM/MM models have been reported subsequently,[4,7−15] which differed in the particular QM and MM methods used, and in the treatment of the QM/MM interactions. The level of quantum mechanical theory employed ranged from

(1) Hajdu, A.; Andersson, J. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 467−498.
(2) Fersht, A. R. *Enzyme Structure and Mechanism*; W. H. Freeman and Co.: New York, 1985.
(3) Kollman, P. A.; Merz, K. M., Jr. *Acc. Chem. Res.* **1990**, *23*, 246−252.
(4) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons, Inc.: New York, 1991.
(5) Knowles, J. R. *Science* **1987**, *236*, 1252−1258.
(6) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**,*103*, 227−249.
(7) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718−730.
(8) Bash, P.; Fields, M.; Karplus, M. *J. Am. Chem. Soc.* **1987**, *109*, 8092−8094.
(9) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comp. Chem.* **1990**, *11*, 700−733.
(10) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580−10594.
(11) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298−305.
(12) Matsubara, T.; Sieber, S.; Morokuma, K. *Int. J. Quantum Chem.* **1996**, *60*, 1101−1109.
(13) Monard, G.; Loos, M.; Thery, V.; Baka, K.; Rivail, J. L. *Int. J. Quantum Chem.* **1996**, *58*, 153−159.

S0002-7863(97)02723-6 CCC: $15.00
Published on Web 03/28/1998

time consuming but fairly accurate *ab initio* Hartree−Fock[7,12] and density functional[14] methods to semiempirical approaches[8−13] and empirical valence bond (EVB)[4] descriptions. Due to their computational expense, *ab initio* methods have found little use in simulations which require a large number of single point calculations. Notable exceptions do exist in which molecular dynamics simulations using pure density functional[16] or Hartree−Fock[17] potentials were used, but these applications have been limited to relatively short simulations of gas phase, cluster systems or of small periodic boxes of solvent molecules. Recently, semiempirical QM/MM potentials have been used in molecular dynamics and free energy simulations of enzyme systems.[18] The computational expense is much reduced compared to *ab initio* potentials, allowing for runs of several hundred picoseconds in length. Unfortunately, the low accuracy of currently available semiempirical methods limits their general application for enzymatic reactions.

The most practical QM/MM approach to study enzyme catalysis at present involves the EVB method.[4,19,20] The empirical valence bond Hamiltonian is first calibrated for a model reaction in solution and is then used to describe the reaction center of the enzyme substrate complex. This approach is computationally efficient due to the simplicity of the EVB Hamiltonian. The QM energies and gradients are easy to evaluate and typically require less computer time than the classical treatment of the remainder of the system. It is not clear, however, whether the parametrization for the solution reaction guarantees an equally appropriate description of the enzyme. A further disadvantage is the difficulty of finding the necessary valence bond structures for the particular reaction under study.

The limited predictive power of semiempirical and EVB methods warrants the development of a more "first principles" approach. Some time ago, Jorgensen[21,22] proposed a computationally feasible strategy to incorporate high level *ab initio* quantum chemical treatment of solutes with classical simulations of solute−solvent interactions. For the chemical reaction of a given solute, one first obtains the gas phase reaction path from *ab initio* calculations and then calibrates classical potential functions to reproduce solute−solvent geometries and interaction energies. These potential functions are then used along with standard solvent force fields to perform free energy simulations (FE) and to obtain the free energy of solvation as a function of the reaction coordinate. Jorgensen employed Monte Carlo simulations to generate statistical ensembles, but his protocol was later also used in the context of Molecular Dynamics (MD) simulations.[23] We call this approach QM-FE to distinguish it from fully coupled QM/MM.

The major characteristic that distinguishes this QM-FE approach from the QM/MM models discussed so far is the static (gas phase) description of the reacting solute. The static geometry allows for a completely classical simulation of a chemical reaction in solution to calculate the relative free energies. The approach assumes, however, that solvation has little or no influence on the course of the reaction and that it only affects the energetic profile. Many successful applications support the validity of this assumption for reactions of small organic solutes.[24−29] However, three major challenges must be overcome before this approach can be used for complex, enzymatic reactions.

The first challenge is that the model system fragments can usually not simply be optimized in the gas phase to obtain a relevant reaction pathway. This is due to the preorganization of the enzyme, which fixes the relative geometries of the fragments and keeps them from moving freely relative to one another as would happen on the gas phase reaction pathway. A possible exception to this is when a metal prosthetic group provides a very strong orienting influence on the pathway. Zheng and Merz[30] used this fact when they studied the mechanism of carbonic anhydrase using a QM-FE approach. Below we present a general approach that can meet this challenge for enzymatic reactions. It relies on the fact that noncovalent interactions involve a much less stringent directionality than covalent interactions.

The second major challenge is the "link atom" problem. This problem is simply that there is no obvious way to correctly describe the energies at the junction between covalently bonded molecular mechanical and quantum mechanical atoms, which almost always occur in enzymatic reactions. The simple organic reaction studies noted above are free from this concern because there are no covalent bonds between the reacting solutes (whose energies are evaluated quantum mechanically) and the solvent molecules (whose noncovalent interactions with each other and the solute are described with molecular mechanics). Below we present a way to mitigate this problem through treatment of the charges used on the QM atoms and a judicious choice of restraints.

The final problem we face is how to generate charges for the quantum mechanical atoms, in order to calculate their interaction with the molecular mechanical atoms. Although there are various choices one can make in this regard, we show that the RESP approach[31−34] used here has some excellent features: (1) it is identical to the approach used to derive molecular mechanical electrostatic charges for the protein, thus automatically leading to balanced protein−protein and protein−substrate interactions, and (2) because the Lagrangian constraints in the RESP software can be employed in a general way, it is

(14) Stanton, R. V.; Hartsough, D. S.; Merz, J. K. M. *J. Comput. Chem.* **1995**, *16*, 113−128.
(15) Thompson, M. A. *J. Phys. Chem.* **1996**, *100*, 14492−14507.
(16) Lassonen, K.; Klein, M. L. *J. Phys. Chem.* **1994**, *98*, 10079−10083.
(17) Gibson, D. A.; Carter, E. A. *J. Phys. Chem.* **1993**, *97*, 13429−13434.
(18) Hartsough, D. S.; Merz, K. M. *J. Phys. Chem.* **1995**, *99*, 11266−11275.
(19) Fothergill, M.; Goodman, M. F.; Petruska, J.; Warshel, A. *J. Am. Chem. Soc.* **1995**, *117*, 11619−11627.
(20) Aqvist, J.; Warshel, A. *J. Am. Chem. Soc.* **1990**, *112*, 2860−2868.
(21) Chandrasekhar, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 2974−2975.
(22) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 154−162.
(23) Sun, Y.; Kollman, P. *J. Chem. Phys.* **1992**, *97*, 5108−5112.
(24) Duffy, E. M.; Severance, D. L.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1992**, *114*, 7535−7542.
(25) Jorgensen, W. L.; Blake, J. F.; Lim, D. C.; Severance, D. L. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 1727−1732.
(26) Madura, J. D.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1986**, *108*, 2517−2527.
(27) Schreiner, P. R.; Severance, D. L.; Jorgensen, W. L.; Schleyer, P. v. R.; Schaefer, H. F. *J. Am. Chem. Soc.* **1995**, *117*, 2663−2664.
(28) Severance, D. L.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1992**, *114*, 10966−10968.
(29) Blake, J. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 7430−7432.
(30) Zheng, Y. J.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1992**, *114*, 10498−10507.
(31) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(32) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **l995**, *117*, 5179−5197.
(33) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comp. Chem.* **1995**, *16*, 1357−1377.
(34) Cornell, W. D.; Cieplak, P. C.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.

an essential component in our strategy to circumvent the link atom problem noted above.

We apply our simulation protocol to the first step of the amide hydrolysis reaction catalyzed by the serine protease trypsin. The entire reaction pathway has been studied previously by applying the semiempirical PM3[35,36] method to a gas phase model of the active site.[37] Although these calculations were not very accurate, they clearly indicated that the first step, the formation of a tetrahedral intermediate in acylation, is rate limiting and that it has the highest barrier on the entire pathway. This is consistent with experimental results and assumptions made by Warshel and Russell.[38] Below we report our calculations on the free energy difference between models of the Michaelis complex and of the tetrahedral intermediate for the acylation step both in trypsin and in water. Our first principles calculations are in good agreement with experimental results for trypsin and are consistent with indirect experimental information for the solution reaction. This connection with experiment helps to validate the methodology.

Methods and Computational Details

In this section we describe an extension of the QM-FE approach to study enzymatic reactions. In enzyme reaction mechanism studies typically only a small fragment of the protein substrate complex requires quantum mechanical treatment. It typically includes the active site residues of the enzyme and a model of the substrate. After initial MM geometry optimization and MD equilibration for the entire system, the QM region is cut out of the protein, and dangling bonds at the QM/MM interface are satisfied with hydrogen atoms (referred to as "link atoms"). After fixing certain anchor points of the QM region to their original position within the protein's frame of reference, restricted QM geometry optimizations are performed to release bond and bond angle strain and to define appropriate QM states for *ab initio* single point calculations. This procedure is applied to several points along the putative reaction path to yield the reaction energy of the QM region in the absence of the rest of the protein. Subsequently, appropriate atomic charges are derived for the QM atoms applying the RESP[31−34] with a Lagrangian constraint that ensures an integral total charge for the enzyme substrate complex. These charges are then used along with standard parameters to define the force field representation of the complex for FEP calculations.

Two sets of free energy simulations were performed to obtain the difference in free energy between points on the reaction pathway. In the first set, the QM region is static during the entire length of the simulation. This protocol constitutes the closest analogy to Jorgensen's approach. In a second set of simulations, internal degrees of freedom of the QM region are allowed to relax in response to the dynamics of the enzyme, while maintaining key geometry restraints to keep the system in an appropriate geometry. In this way, we can assess the impact of active site flexibility on the calculated free energies.

Our protocol allows us to examine the nature of enzyme catalysis by performing analogous simulations for a reaction in a solvent and in a protein environment. Comparison of these two simulations requires the determination of the work necessary to constrain the reacting molecules in a productive geometry. This free energy term, often referred to as cratic free energy, can have a significant effect for the solution reaction, and in contrast it is absorbed in the free energy of binding for the enzymatic reaction. We present two simple, independent approaches to estimate the magnitude of the cratic term.

Initial Structure Definition of Michaelis Complex (MICO). In our simulations the initial coordinates for trypsin were obtained from the X-ray crystal structure of bovine trypsin complexed with bovine pancreatic trypsin inhibitor (Brookhaven Protein Data Bank code 2PTC) determined to 1.9 Å resolution.[39] The trypsin inhibitor was removed from the complex and the catalytically active substrate Acetyl-Ala-Phe-Arg-Ala-NH_2 (k_{cat} = 52.7 s^{-1}, $k_{cat}/K_M = 2.4 \times 10^5$ L/M s)[40] was placed in the active site X-ray coordinates of the inhibitor. Energy minimizations of the substrate in the trypsin environment when the Cornell *et al.* force field[32] was used to position the substrate properly in the active site. The active site of the trypsin−substrate complex was then solvated by adding a sphere of TIP3P[41] water molecules with a 10 Å radius from the O_γ of the catalytic Ser195 with use of the AMBER[42] "cap" option. The protein complex, which included one crystallographically defined Ca^{2+} ion, was neutralized by placing 9 Cl^- ions in the positions of largest positive electrostatic potential as determined by the program CION[43] of the AMBER package. The counterions were located outside of the 20 Å water cap and their positions were fixed during all simulations to avoid artefactual long range electrostatic effects on the calculated free energies. A final energy minimization of 1000 steps was performed with use of the conjugate gradient algorithm. The result was an energy minimized trypsin substrate system, which had been neutralized by the addition of counterions and solvated with a cap of water around the active site.

The minimized system was used as the starting point for molecular dynamics (MD) simulations performed at a constant temperature of 300 K with use of the Berendsen algorithm.[44] Standard parameters of the Cornell *et al.*[32] force field were used in all calculations unless otherwise noted. A time step of 1.5 fs and the SHAKE algorithm[45] to constrain bond distances were used along with a 10 Å nonbonded cutoff. In the protein simulations, we allowed movement only in residues within 14 Å of the O_γ of Ser195. Additionally, after an initial 15 ps of molecular dynamics 30 TIP3P water molecules were added to refill the 20 Å water cap as some of the initial water molecules were able to diffuse into gaps in the protein. The protein system now included 374 water molecules of which 320 were within the moving portion of the system. A further 30 ps of MD equilibration was performed. In these two MD simulations the substrate was kept in a productive binding geometry by constraining the distance between the O_γ of Ser195 and the carbonyl carbon of the scissile amide bond (carbonyl carbon of Arg) to 3.2 Å by a harmonic constraint force of 20 kcal/(mol Å).

Choice of Quantum Mechanical Model. A model system must be chosen that balances the need for including all of the catalytically important residues with the computational limitations inherent to *ab initio* calculations. In the trypsin system several active site residues need to be considered for inclusion in the QM model; these are the Ser195, His57, and Asp102 of the catalytic triad, the oxyanion hole residues, and the substrate (see Figure 1). The decision of which residues or atoms to include effectively determines the level of *ab initio* theory that can be employed. If all of the substrate and active site residues mentioned above were included in the model, then only semiempirical theory could be practically applied for the QM calculations.

Of the residues we considered above, Ser195, His57, and the substrate are the most important, since they are directly involved in the bond breaking and bond forming process (see Figure 1 for the QM model system used). To allow for the highest possible level of QM calculations, we decided to further reduce our model system by only including the key fragments of these residues: methanol and imidazole for serine and histidine, respectively, and *N*-methylacetamide for the

(35) Stewart, J. J. P. *J. Comp. Chem.* **1989**, *10*, 209−220.
(36) Stewart, J. J. P. *J. Comp. Chem.* **1991**, *12*, 320−341.
(37) Daggett, V.; Schröder, S.; Kollman, P. *J. Am. Chem. Soc.* **1991**, *113*, 8926−8935.
(38) Warshel, A.; Russell, S. *J. Am. Chem. Soc.* **1986**, *108*, 6569−6578.
(39) Marquart, M. J. W.; Deisenhofer, J.; Bode, W.; Huber, R. *Acta Crystallogr. B* **1983**, *39*, 480−485.
(40) Dutler, H.; Bizzozero, A. *Arch. Biochem. Biophys.* **1987**, *256*, 662−676.
(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−933.
(42) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 4.0*, 1991, University of California, San Francisco,
(43) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. A. *Comp. Phys. Com.* **1995**, *91*, 1−41.
(44) Berendsen, H. J. C.; Potsma, J. P. M.; van Gunsteren, W. F.; DiNola, A. D.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(45) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

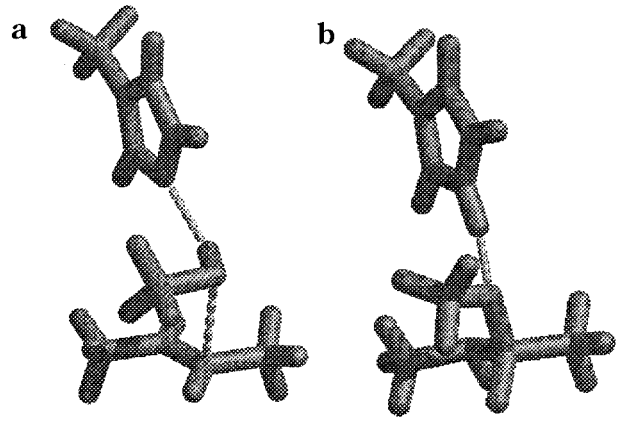

Figure 1. (a) The quantum mechanical model of MICO. (b) The quantum mechanical model of TET.

substrate. This model system maintains the essential aspects of the acylation reaction.

Quantum Mechanical Optimization. The MM geometry obtained from MD and minimizations is most likely not appropriate for the QM potential energy surface of the reaction center. For this reason, *ab initio* geometry optimizations of the QM model are necessary. Initial calculations showed, however, that unrestricted QM optimizations lead to very unreasonable orientations of the active site residues. This problem could certainly be prevented by using a traditional QM/MM scheme that considers the influence of the environment through a mechanical (and possibly electrostatic) embedding of the QM portion. Unfortunately, full QM/MM geometry optimizations are currently unfeasible at an appropriately high level of *ab initio* theory. We may, however, simulate such an embedding by fixing certain anchor atoms of the QM model to the protein coordinate frame. In practice, we performed geometry optimizations for the monomers with the C_β and C_γ atoms of imidazole (His), the C_β and O_γ atoms of methanol (Ser), and the carbonyl O and attached methyl C atoms of the substrate (Sub) fixed.

Approximate reaction coordinates were calculated for the dimers of methanol−imidazole (Ser-His) and methanol−substrate (Ser-Sub) to ensure that the MM optimized intermolecular distances were reasonable for the QM description of the reaction center. For this purpose we reinserted the QM optimized geometries of the monomers into the original protein coordinate frame. A set of single point *ab initio* calculations was carried out for various N_ϵ (His)−O_γ(Ser) and carbonyl C (Sub)−O_γ(Ser) distances, with one of the monomers moved relative to the other along the common bond vector. The optimized MM structure showed a N_ϵ (His)−O_γ(Ser) distance of 2.9 Å, which was confirmed to be an energy minimum on the QM reaction coordinate. For the methanol−substrate dimer, the QM calculations suggested a slightly longer C (Sub)−O_γ(Ser) distance than the MM geometry optimizations (3.2 Å vs 2.9 Å). Hence we decided to perform two sets of free energy simulations, using either the MM or the QM dimer distance. The final results were so similar, however, that we will only discuss in detail the calculations for the dimer distance of 3.2 Å (which we will refer to as MICO hereafter), although we briefly discuss the results of the model with the distance of 2.9 Å, referred to below as MICO1.

All QM geometry optimizations were performed at the HF/6-31+G* [46] level with the Gaussian94 suite of programs.[47] Optimizations were also performed by using the semiemperical molecular orbital models PM3[35,36] and AM1.[48] Additional single point calculations were performed at higher levels (MP2/6-31+G*,[46] MP2/AUG-cc-pVDZ[49]) for the trimer composed of the His, Ser, and Sub mimics, using the *ab initio* optimized geometries of the monomers reinserted into the protein coordinate frame. We used the 6-31+G* basis set because it is, in our view, the smallest one with sufficient accuracy for anions with which we could do geometry optimizations and expect reasonable agreement with experiment. This modest size lets us do the optimization on the largest possible system efficently. Figure 1a shows a view of the quantum mechanically optimized structure of MICO.

The Tetrahedral Intermediate: TET. An initial model of the first tetrahedral intermediate was obtained from the refined structure of the Michaelis complex by forming the appropriate C−O bond between Ser195 and the substrate and by transferring the Ser195 H_γ proton to N_ϵ of His57. The resulting geometry was then optimized for 1000 steps by using our standard force field and the conjugate gradient algorithm. Further refinement with *ab initio* geometry optimizations was performed in close analogy to the procedure discussed above. In the case of TET, however, only two geometry optimizations needed to be done: one for protonated imidazole and one for the covalent complex between *N*-methylacetamide and methanolate. In the latter case, only the carbonyl oxygen and neighboring methyl carbon of the adduct complex were held fixed, in analogy with the optimization of NMA in MICO. Figure 1b shows a quantum mechanically optimized structure of TET.

A Possible Intermediate State: PTMICO. In addition to MICO and TET we examined one further state that represents a possible intermediate point on the free energy surface. This state (PTMICO) is derived from the structure of MICO by only transferring the Ser195 H_γ proton to N_ϵ of His57. The *ab initio* geometry of PTMICO was refined by using the procedures discussed above.

Atomic Point Charges. Atomic point charges were calculated with the RESP[31−34] method at the HF/6-31+G* level for the various model systems. We used this model since it is very similar to the 6-31G* model used to derive the protein charges[32] yet has extra flexibility to describe anionic species which play a role in this enzyme mechanism.

The RESP approach to generate charges goes beyond simply fitting electrostatic potentials by (a) damping out the charges of statistically ill-determined atoms, (b) allowing simultaneous multimolecule and multiconformational fitting,[32,33,50] and (c) allowing a general use of Lagrangian restraints to fix the net charge of the system. For example, if one begins with the charges of the protein residues Ser195 and His57 and removes the atoms from C_β to the end of the side chain, the net charge remaining, albeit small, is not exactly zero. Lagrangian restraints on the fitting of the charges of the quantum mechanical atoms allow us to ensure that the sum of the charges of the quantum mechanical atoms plus the charges of the molecular mechanical atoms of Ser195 and His57 including up to C_α is exactly zero.

This approach is a good one for our trypsin model as the charge deviation from zero is small ($|0.07e^-|$). If it were not, an alternative strategy could be adopted by rederiving the charges for the Ser and His amino acids by using the Lagrangian constraints to ensure that the sum of the backbone charges and the side chain charges each separately were zero. The side chain charges for Ser and His could then be replaced without forcing the net charge of the QM atoms to be slightly different than zero.

In order to minimize partial charge artifacts at the link atoms between the QM and MM regions, we also constrained the hydrogens on the boundary atoms (e.g. C_β in Ser and His) to have zero charge. This is equivalent to using a "united atom" charge model at this junction. The charges for the various systems are given in Figure 2.

It is desirable to use the same set of RESP charges for the solution reaction as were used in the protein. However, as the solution model lacks the protein backbone attachments, when these charges are used

(46) Hehre, W. J.; Radom, L.; van Schleyer, P. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley: New York, 1986.

(47) Frish, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94, D.3*, 1994, Gaussian, Inc.: Pittsburgh.

(48) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(49) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806.

(50) Howard, A.; Cieplak, P.; Kollman, P. *J. Comp. Chem.* **1995**, *16*, 243−261.

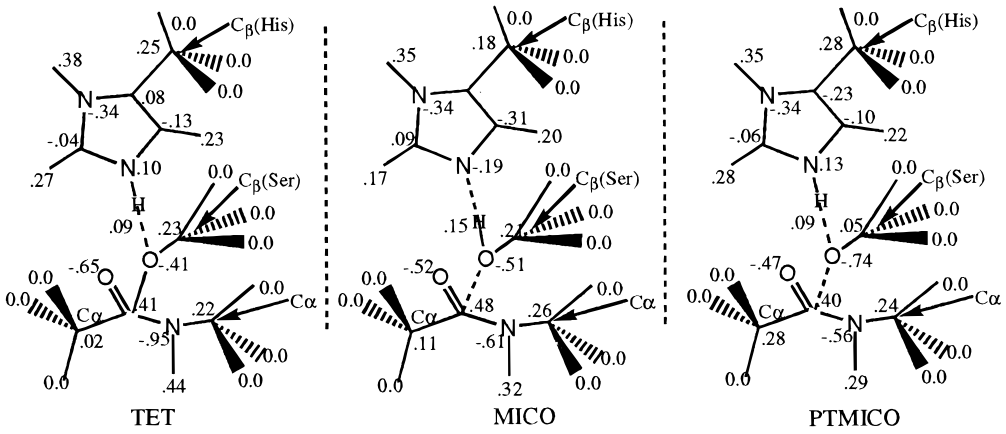

Figure 2. Schematic of enzyme active site with charges.

in solution they add up to a non-zero value (−0.07e for both TET and MICO). Non-unit charges are certainly unphysical, but their effect may be small in practice. To estimate this effect we performed short (38 ps) free energy simulations for the solution reaction, perturbing between the two possible charge distributions (model 1: RESP charges with constraints as above; model 2: unconstrained RESP charges with total charges = 0). The calculated free energy differences (model 1 → model 2) are 2.l ± 0.3 kcal/mol for MICO and 3.2 ± 0.2 kcal/mol for TET, respectively, where the error estimates indicate the difference between forward and backward runs. The difference in relative free energies is thus fairly small (ca. 1 kcal/mol), justifying the use of the same set of RESP charges (Figure 2) for both the protein and solution simulations.

Free Energy Calculations. The above derived geometries and charges were then used in free energy perturbation (FEP) calculations. Two different protocols were used for the protein simulations. In the first, the active site geometry was kept rigid (Cartesian mapping) while in the second the active site geometry was allowed to relax and move dynamically. For each of the protocols, we performed 30 ps equilibration runs prior to data sampling.

We used a 12 Å cutoff for all simulations after showing in shorter (∼40 ps) simulations that the calculated free energies were similar (<1 kcal/mol different) with 10, 12, and 14 Å cutoffs. We also carried out a free energy calculation on MICO to TET in the enzyme with a dual cutoff[51] (12 Å primary and 22 Å secondary), and the free energy was within 1 kcal/mol of the 12 Å cutoff value.

Cartesian Mapping. The relative free energies were first calculated by using the Cartesian mapping algorithm.[23] In this algorithm the geometry of the active site/model system (Figure 2) is rigid and a function of the perturbation parameter (λ) while the remainder of the system is allowed to move freely. When $\lambda = 1$ the fixed active site geometry and charges correspond to the initial state (MICO, MICO1, or PTMICO), and when $\lambda = 0$ the charge and fixed geometry are those of TET. We did allow motion in QM atoms at the QM/MM boundary (any atom within the model system that was covalently bonded to an atom outside of the model). Allowing these atoms to move greatly improved the results of our calculations, incorporating a "buffer" zone between the rigid model and remainder of the system. The flexibility in these atoms did not affect the positions of the catalytically active atoms, but reduced inappropriate strain at the QM/MM junction. The free energy obtained in this way for the interaction of the fixed active site residues within the remainder of the protein can then be used to estimate the environmental effects on the *ab initio* models. When combined with the difference in energies between the *ab initio* states an overall relative free energy between the states is calculated.

Flexible FEP Simulations. For the flexible FEP simulations several "pseudobond" and "pseudoangle" harmonic restraints were introduced to ensure proper active site geometries throughout the simulation. (See Table l). The perturbation in the flexible calculation included changes in the charge distribution and bond topology. The relative free energy for each perturbation was calculated as the average from "forward" and "reverse" sampling in a single run (see Table 4). Each perturbation (in both the flexible and Cartesian mapping calculations) was run from both $\lambda = 1$ to 0 and $\lambda = 0$ to 1, to obtain an estimate of the error in the calculations from the hysteresis.

Table 1. Harmonic Constraints Used in Flexible FEP Calculations

model		
MICO	TET	PTMICO
	bonds[a]	
Ser Oγ−Sub C	Ser Oγ−His Hϵ	Ser Oγ−Sub C
Ser Hγ−His Nϵ		Ser Oγ−His Hϵ
	angles[b]	
Ser Oγ−Ser Hγ−His Nϵ	Ser Oγ−His Hϵ−His Nϵ	Ser Oγ−Ser Hγ−His Nϵ
Ser Hγ−His Nϵ−His C_δ	His Hϵ−His Nϵ−His C	Ser Hγ−His Nϵ−His C_δ
Ser Oγ−Sub C−Sub O*	Sub C−Ser Oγ−His Hϵ	Ser Oγ−Sub C−Sub O*
Ser Oγ−Sub C−Sub N*		Ser Oγ−Sub C−Sub N*
Ser Oγ−Sub C−Sub CH_3*		Ser Oγ−Sub C−Sub CH_3*

[a] All bond equilibrium values are from *ab initio* calculations. All force constants = 100 kcal/(mol Å^2). [b] All angle equilibrium values from *ab initio* calculations; force constants = 30 kcal/(mol radian^2) except for those (*) where 40 kcal/(mol radian^2) was used. The higher value was required to keep the geometry close to the *ab initio* calculated value.

Aqueous Phase Calculations. The active site models were placed in a box of TIP3P water with dimensions ∼36 × 31 × 33 Å. Simulations employed periodic boundary conditions and were run at constant volume and temperature, after an initial equilibration of the box dimensions to those proper for a pressure of 1 atm. Rigid and flexible simulations identical in length and procedure to those carried out for the protein were done in solution (Table 3).

Cratic Free Energy Contributions. For the solution reaction there are important contributions to the free energy which arise from bringing the imidazole, methanol, and *N*-methylacetamide together in a reactive geometry. Below, we estimate this free energy using two different models.

In the first model, we assume that, relative to methanol, both imidazole and NMA must occupy very specific positions in Cartesian space in order to react. From a series of molecular dynamics runs we derive the restraint force constants required to keep the two molecules in a proper reactant geometry with standard deviations of 0.2 Å and 20° for bonds and angles, respectively. The specific requirements are as follows: (1) The methanol H−imidazole N distance should be l.9 Å and the corresponding hydrogen bond angle should be linear. (2) The substrate carbonyl C−methanol O distance should be 3.5 Å, and the corresponding OCO angle should be 90°, with the methanol O lone pair facing the carbonyl C. A C−O distance of 3.5 Å (instead of 3.2 Å, *vide supra)* was used because the potential of mean force calculations showed a too repulsive interaction for 3.2 Å, with a minimum at 3.5 Å. After establishing the force constants needed to fulfill the above geometric restraints, the cratic free energies were calculated from the analytical formulas given in the paper of Hermans and Wang.[52]

(51) van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992−1023.

Table 2. Cratic Free Energy Contributions (kcal/mol)

free energy components	methanol−NMA	methanol−imidazole	methanol−NMA−imidazole[a]
		Method 1[b]	
translational	4.8	4.0	8.8
angular	0.6	0.6	1.2
PMF	0.6 ± 0.2[c]	−0.3 ± 0.1[c]	0.3
dihedral		1.1	1.1
total	6.0	5.4	11.4
		Method 2[d]	
translational	2.4	2.4	4.8
angular	1.7	2.1	3.8
PMF	0.6 ± 0.2[c]	−0.3 ± 0.1[c]	0.3
total	4.7	4.2	8.9

[a] The energies for the complete system were taken as the sum from methanol−NMA and methol−imidazole interactions. [b] The first method used formulas of Hermans and Wang.[52] This should provide an upper bound to the total possible cratic free energy contribution. [c] These PMF contributions were determined by using AMBER[42,43] GIBBS to determined the free energy of the complex in a box of water as a function of distance ($O_\gamma\cdots C$ in the case of methanol−NMA and $H_\gamma\cdots N_\epsilon$ in the case of methanol−imidazole). In both cases, there was a free energy minimum near ∼3 Å; in methanol−NMA it was 0.6 kcal/mol above the free energy at ∼6 Å; in methanol−imidazole, it was 0.3 kcal/mol below the free energy at ∼6 Å. [d] The second, more approximate method used simple statistical mechanics volume and angular considerations. See text in the Methods section.

Table 3. ΔE for TET → MICO[a] Model Systems from *ab Initio* Calculations (kcal/mol)[b]

	HF/6-31+G*	MP2/6-31+G*	HF/AUG-cc-pVDZ	MP2/AUG-cc-pVDZ
TET → MICO[c]	−70.8	−58.9	−68.6	−53.8
TET → MICOl[d]	−64.2	−54.5	−65.0	−51.3

[a] MICO = Michaelis complex; TET = tetrahedral intermediate. [b] The model systems used consisted of NMA, methanol, and imidazole. The geometries used are fully described in the text. [c] The absolute QM energies in hartrees for TET are as follows: HF/G-31+G*, −625.809610; MP2/6-31+G*, −627.730983; HF/AUG-cc-pVDZ, −625.891836; MP2/AUG-cc-pVDZ, −627.986133. [d] Slightly modified Michaelis complex structure—see Computational Details.

Table 4. Calculations of ΔG_{int} (kcal/mol) for TET → MICO[a]

		TET → MICO[b]	
length (ps)	model[c]	protein	solvent
40	fixed	36.2/36.1	33.2/33.7
80	fixed	36.2/36.3	32.5/32.3
150	fixed	37.0/35.1	33.0/32.4
75	flexible	39.7/39.4	34.6/34.0
150	flexible	37.6/38.6	34.4/29.7

[a] TET = tetrahedral intermediate; MICO = Michaelis complex. [b] The values from forward and reverse runs are given. [c] Two models were used: (1) fixed, using the Cartesian mapping[23] on the active site while the remainder of the protein was flexible, and (2) flexible, using standard free energy calculation techniques (see text) with a few harmonic restraints to maintain the proper geometries.

An alternative, simpler method to estimate these contributions is to assume that imidazole and NMA are replacing a water molecule each in the solvation shell of methanol. From simple concentration arguments, the free energy change of both molecules can then be evaluated as,

$$\Delta G_{trans} = -RT \ln \frac{1}{[H_2O]}$$

where $[H_2O] = 55$ M. To estimate the angular contribution we can use a ratio of the total angular space to that portion which might represent a reactive geometry. A reasonable assumption for this estimate is that any angle within 20° of the reference angle might result in a reactive geometry. If we consider in spherical polar coordinates the angles of interest (both the imidazole−methanol H bond and the methanol−NMA approach of the methanol oxygen perpendicular to the NMA plane) it is simple to calculate the area of the core of angles within 20° and compare this to the volume of the entire sphere. The free energy can then be estimated as *RT* ln (allowed angular space/total angular space).

Finally, in both of the approaches above, the free energy for the molecules to be in the first solvation shell of methanol as compared to being surrounded by water must be calculated. These free energies are estimated by carrying out potential of mean force (PMF) calculations for methanol−NMA and methanol−imidazole as a function of distance. These results are reported in Table 2, where the free energies for restraining positions of the catalytic groups in a reactive geometry in water range from 8.9 to 11.4 kcal/mol.

Results

In analogy with the approach introduced by Jorgensen in 1984−85,[21,22,53] we estimate the free energy difference ΔG^* between two structures in solution as,

$$\Delta G^* = \Delta E + \Delta G_{int} \quad (1)$$

Here ΔE is the *ab initio* energy difference between the two QM models and ΔG_{int} is the difference in free energy of interaction.

We focus on the formation of the first tetrahedral intermediate (TET) which has been established[37] as the rate limiting step in the acylation pathway of trypsin. The energy difference (ΔE) between the two model states was determined at a number of theoretical levels, including both MP2/6-31+G* [46] and MP2/AUG-cc-pVDZ.[49] These results are presented in Table 3.

We calculated ΔG_{int} in two different ways. In the first approach, the free energy difference between MICO and TET was determined with a flexible model. We neglect any free energy changes within the perturbed group (Figure 1), since these are included in ΔE. These results are presented in Table 4. Using the average of the longer 150 ps flexible model calculations, we find that ΔG_{int}(protein) ∼ 38 kcal/mol and ΔG_{int}(solution) ∼ 32 kcal/mol. The active sites of MICO after equilibration and TET after the perturbation are shown in Figure 3. In a second set of simulations we also evaluated these free energies using Cartesian coordinate mapping. In this approach we rigidly perturb the model systems (whether within protein or solvent) from one structure to the other. With this method we calculate ΔG_{int}(protein) ∼ 36 kcal/mol and ΔG_{int}(solvent) ∼ 33 kcal/mol.

The estimated values of ΔG^* (from eq 1) are summarized in Table 5. In the flexible and Cartesian mapping calculations the ΔG^*(protein) value of ∼16−18 kcal/mol is in good agreement with the experimental value of ∼15.1 kcal/mol for the activation free energy ($\Delta G^‡$) of acylation by trypsin (using the single Eyring equation relating k_{cat} to $\Delta G^‡$). Comparing ΔG^*(protein) and ΔG^*(solution) we find a $\Delta\Delta G^*$ difference of ∼3−6 kcal/mol depending on the model. Although it is not obvious how to compare ΔG^*(solution) with experiment, a reasonable estimate is provided by the $\Delta G^‡$ for base-catalyzed hydrolysis of amides extrapolated to pH 7, which is ∼33 kcal/mol.[53] As one can see, our estimates of ΔG^*(solution) of 21−22 kcal/mol are far from this value. However, if we add the cratic free energy correction of 11 kcal/mol (see above), the theoretical value (32−33 kcal/mol) is in reasonable agreement with the experimental estimate. We also carried out some free energy calculations on MICO1 (not reported in detail). Using the ΔE from Table 3 for MICO1 and adding the ΔG_{int}, we find

(52) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2707−2714.
(53) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184−189.
(54) Guthrie, J. P. *J. Am. Chem. Soc.* **1974**, *96*, 3608−3615.

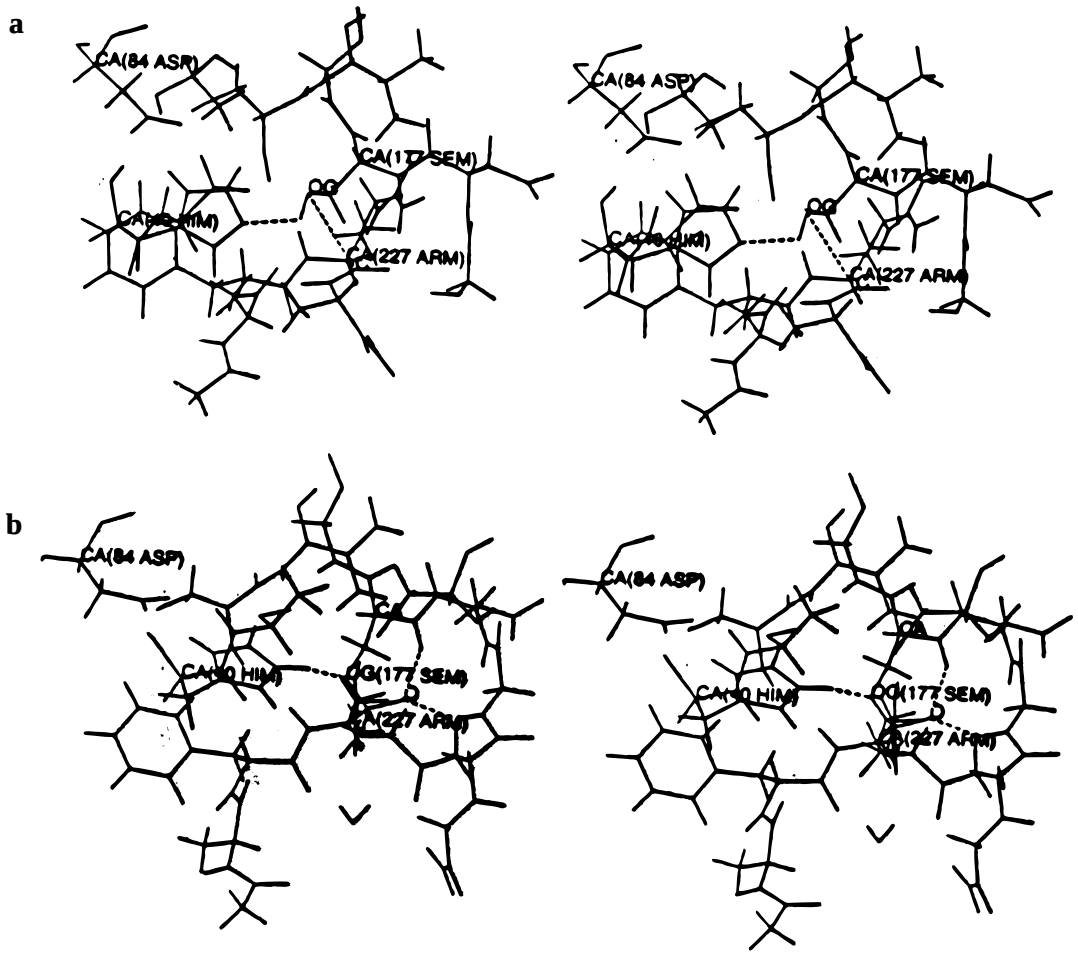

Figure 3. (a) Stereoview of a snapshot of the equilibrated MICO structure. Displayed are residues with an atom within 5 Å of Ser177Oγ as well as residues 175−177, Asp84, and the entire substrate. (Note that in trypsin, the catalytic triad is Asp84-His40-Ser177). The distance between Ser177 Hγ and His40Nϵ and Ser177 Oγ and the substrate scissile carbonyl C are dotted. (b) Stereoview of a snapshot of TET after perturbation from MICO. Same residues displayed as in part a. Dotted lines are distances between His He and Ser O_γ and between NH of Ser177 and NH of Gly175 and the scissile carbonyl oxygen (the latter two showing the oxyanion hole interaction).

Table 5. Total Calculated ΔG^* (kcal/mol)[a]

	MICO → TET	
model	ΔG^* (prot)	ΔG^*(solv)
flexible	16	22
fixed	18	21

[a] Using ΔE from MP2/AUG-cc-pVDZ and ΔG_{int} from 150 ps flexible runs.

that the ΔG^* for MICO1 is within 1−2 kcal/mol of that of MICO. Thus, our calculated ΔG^* seem robust to small structural changes.

We should note that we are comparing the calculated free energy ΔG^* for forming the tetrahedral intermediate with the experimental $\Delta G^{\ddagger}$ for forming the transition state. However, earlier work by Daggett *et al.*[37] and Warshel and Russell[38] and a more recent study on formamide hydrolysis[54] suggest that the two are likely to be close in energy. The calculated value of ∼16 kcal/mol is consistent with this, but one should not overemphasize the quantitative agreement.

The difference between the ΔG^* calculated with the flexible and Cartesian mapping models is of the sign one would expect. While the flexible model allows for internal relaxation, it only includes interactions between the active site and the environment in the free energy estimate. Interactions within the active site are neglected. Hence we may expect that the model only captures the attractive part of the orientational polarization, but neglects the work required to polarize the active site. From classical theory of polarization we conclude that the difference between the flexible and nonflexible model due to this effect is (−1/2) times the difference between the ΔG_{int}(flex) and the ΔG_{int}(rigid) values, or ∼1 kcal/mol.

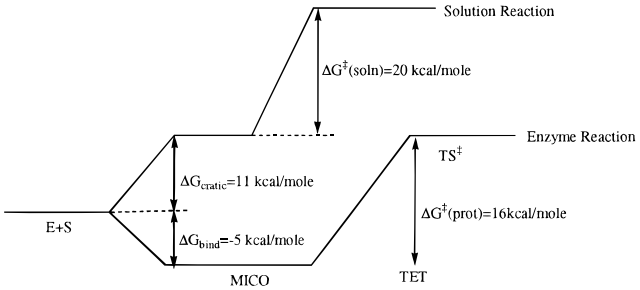

Figure 4. Schematic free energy diagram (see text) assuming the free energy for the transition state is approximately that of TET. ΔG_{bind} has not been determined by simulation, but inferred from the experimental K_M.

We summarize our relative free energies for TET and MICO in a schematic free energy diagram (Figure 4). It is apparent from this diagram that the predominant difference between enzyme and aqueous phase reactions lies in the formation of the Michaelis complex (ES). By forming such a complex, the

(55) Bakowies, D. 1997. Unpublished results.

Table 6. ΔE, ΔG_{int}, and ΔG^* for TET → PTMICO[a] (kcal/mol)[b]

	HF/ 6-31+G*	MP2/ 6-31+G*	HF/AUG-cc-pVDZ	MP2/AUG-cc-pVDZ
ΔE	5.5	−0.9	−3.3	3.0
	flexible		fixed	
	enzyme	solution	enzyme	solution
ΔG_{int}[a]	−5.8 ± 0.9	9.3 ± 0.5	25.6 ± 0.2	0.1 ± 0.2
ΔG^* [b]	−2.8	12.3	28.6	3.1

[a] Error bars given as range of forward and backward results for 150 ps simulation. [b] Using ΔE from MP2/AUG-cc-pVDZ.

enzyme keeps the free energy of ES below that of E + S, whereas in solution, it costs ∼10 kcal/mol of free energy to preorganize the reactants. It is also likely that there is some "cratic" free energy involved in forming the enzyme−substrate complex, but this is included in the experimental ΔG_{bind}, which includes both favorable interaction free energies between enzyme and substrate and unfavorable cratic terms.

Our next series of simulations estimated the free energy difference between PTMICO and TET. This is of interest since this is another easily defined point in possible reaction pathways for this system and addresses the issue of whether proton transfer precedes or accompanies carbonyl attack by Ser Oγ. We separately consider PTMICO because there is no experimental evidence, in contrast to MICO and TET, that this state is along the *productive* catalytic pathway or whether proton transfer and C−O bond formation are concerted. In Table 6 we present the ΔE, ΔG_{int}, and ΔG^* for PTMICO. The relative free energies show that PTMICO is much closer in energy to TET than MICO in both protein and aqueous reactions.

It is important to note that the close correlation between the results obtained by using the fixed and flexible calculations in most of our simulations was not observed for PTMICO. First, considering the reaction in solution, it is interesting that ΔG^* is 3 kcal/mol for the "fixed" model and 12 kcal/mol for the "flexible" model. This is not as surprising when one looks at the final structures for the free energy perturbations of TET → PTMICO. In the case of the "fixed" model, the structure is essentially identical to the quantum mechanical model because Cartesian coordinate mapping forces it to be so. In this structure, the Oγ, which contains most of the negative charge, forms three strong hydrogen bonds with water molecules. On the other hand, in the "flexible model", the Oγ···N and Oγ- -C distance restraints are satisfied, but the imidazolium$^+$ has moved considerably from the structure optimized quantum mechanically. Water is not able to H bond to Oγ as effectively as in the "fixed" model. Thus, in the case of the solution reaction we feel that PTMICO is reasonably represented in the fixed model, and its similar free energy relative to TET suggests it could be along the productive reaction pathway.

In the enzyme, the flexible model finds ΔG^* to be 3 kcal/mol *lower* for PTMICO than TET, whereas in the fixed model it is ∼29 kcal/mol *higher*. To understand this, one has to realize that in TET, most of the anionic charge is on the oxygen stabilized in the oxyanion hole, whereas in PTMICO it is on the Oγ, which has no apparent H-bond donor except for the Histidine H$^+$ to which it delivered the proton. In the fixed model, the peptide plane (*N*-methylacetamide fragment) of the substrate is forced to adopt the position from the quantum mechanical model, which leaves the rest of the substrate a considerable distance from the serine Oγ. On the other hand, in the flexible model, only the Oγ- -C substrate distance is constrained and the substrate is able to make subtle adjustments which enable the N−H bonds of residues 227 and 228 (Arg and Ala on the substrate) to form an oxyanion like hole for the Oγ (Figure 5). There are likely other subtle adjustments in the flexible model that enable it to much more effectively stabilize the anionic Oγ than in the rigid model, but these oxyanion hole H bonds are likely to be the major contributors. Thus, our calculations with the flexible model, which we consider to be a more realistic value for the protein system for PTMICO, suggest that it could be along the reaction pathway for formation of TET and that concerted proton transfer and Oγ- -C attack is not essential for effective enzymatic activity. This is supportive of the assumption made by Warshel and Russell,[38] based on the pK_a of His and Ser that proton transfer precedes Oγ- -C attack.

Discussion

We believe that the protocol presented above addresses the major problems associated with the application of the QM-FE approach to complex biomolecules and opens the door for a general application of the QM-FE approach for any enzymatic reaction, and any analogous solution reaction or biomimetic reaction model. First, our limited requirement for geometry optimization, using the constraints imposed by the enzyme, allows us to use the highest possible level of *ab initio* theory. In addition to the values reported in Table 3, we have also calculated ΔE for PM3 and found TET → MICO to be 61.9 kcal/mol and TET → MICO1 to be 58.3 kcal/mol and with AM1 TET → MICO was 69.1 kcal/mol and TET → MICO1 was 69.2 kcal/mol. It is clear from comparing these energies with those in Table 3 that the error in using lower level (semiempirical or *ab initio)* values to calculate ΔE is large and likely to be significantly larger than errors introduced from the fact that we cannot establish exact transition states and thus employ accurate vibrational corrections in our approach. We have optimized the intermolecular parts of the quantum mechanical atoms at the RHF 6-31+G* level, assuming that this level of quantum mechanical model and optimization is sufficient. We have also used a molecular mechanical approach to optimize most of the much "softer" intermolecular degrees of freedom, assuming that this will not introduce too large an error, particularly if we constrain key intermolecular distances, e.g., Ser O$_\gamma$−carbonyl C, to that found in more limited quantum mechanical optimizations.

Given that the choice of constraints used is not uniquely defined and may affect the results of the single point calculations for the QM model, it seemed appropriate to assess the sensitivity of the QM calculations by comparing the results of another optimization protocol. Thus, we studied an alternative protocol that constrains the Z-matrix of the QM atoms rather than Cartesian coordinates relative to the backbone positions we had found during molecular mechanics minimizations. Such a protocol allows the removal of rigid body translations and rotations of the monomers without constraining any of their internal degrees of freedom. Further internal coordinate constraints may then be added as needed. The details of this protocol are described in ref 56 and show that the calculated ΔE are not overly sensitive to the precise approach for limited geometry optimizations.

Secondly, our use of both Cartesian mapping and flexible FEP calculations offers a useful internal control on our calculations of ΔG_{int}. The fact that both lead to similar values for MICO → TET is encouraging. When they do not give similar ΔG_{int}, as found in TET → PTMICO, this is significant and requires further analysis to see which model is more realistic.

Thirdly, we note that previous implementations of the QM-FE approach do not need to consider covalent interactions

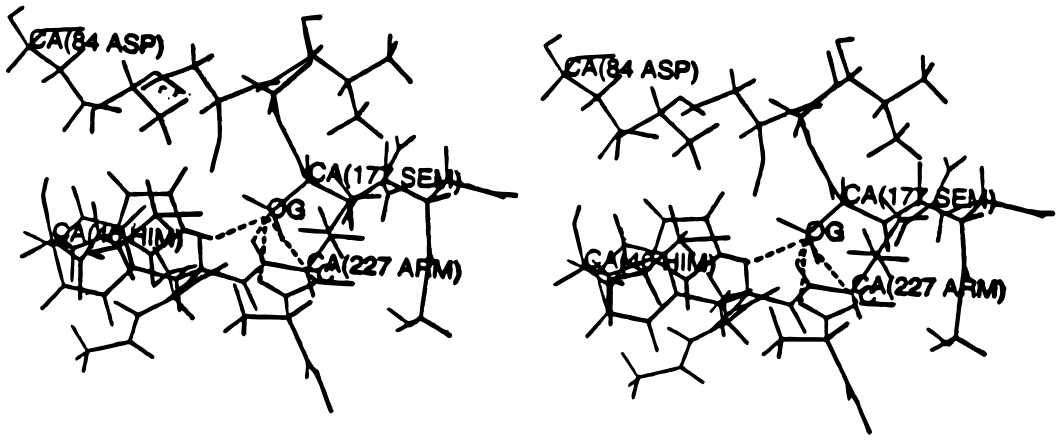

Figure 5. Stereoview of a snapshot of PTMICO. Residues are the same as in Figure 2. Dotted lines are distances between Ser Oγ and the following four atoms: His Hϵ, substrate scissile carbonyl carbon, and the NH of Arg227 and Ala228 of the substrate.

between the QM solute and the MM solvent (the link atom problem[57]). In the case of enzymatic reactions, however, we address the problem of cutting covalent bonds between the QM active site and the MM part of the protein. This involves the definition of link atoms to satisfy the free valences for the QM calculation and can be a source of considerable error. Nonetheless, the use of the RESP methodology[31−34] allows us to create a set of charges for the reacting atoms that is internally consistent with those used in the force field, which should help minimize this source of errors.

Since the 6-31G* and 6-31+G* basis sets tend to lead to RESP charges with enhanced polarity, mimicking the polarization inherent in "effective two body" models, it would be inappropriate to additionally calculate the polarization of the QM part by the MM atoms and vice versa. If one wished to do that, RESP charges from higher QM levels should be used to derive the RESP charges for both MM (to use in polarizing the QM model via incorporation in the one-electron Hamiltonian) and QM (to use in polarizing the MM model via classical polarization calculations[58]) atoms.

As noted above, Zheng and Merz[30] used QM-FE to study carbonic anhydrase (CA). The zinc in CA imposes a geometry in the enzyme very similar to that found in the gas phase (especially when NH_3 is used to model histidine). This allowed them to fully optimize and characterize the gas phase reaction pathway and then keep the structure constrained to this pathway in the FEP calculations. However, in many enzymes, the structural constraints of the enzyme preclude such an approach until a much larger QM model of the active site than is currently possible with suitably high level *ab initio* calculations is used. Thus, our approach employs limited QM optimization within the constraints of the enzyme rather than letting a gas phase pathway drive the enzyme atoms. An additional difference between the approaches used is that our RESP methodology provides a clean way to incorporate the charges for the QM atoms into the FEP calculations, whereas Zheng and Merz fit the charges for the Zn bound NH_3 to a united atom model (just N) and then "evenly dispersed" that charge around the other atoms of the imidazole ring. This is reasonable for their model, since the imidazoles are just anchors and not closely involved in the catalytic reaction, but the RESP methodology appears to be a more general and useful way to derive charges for the QM atoms.

Conclusion

Until this time, Warshel's EVB method has been practically the only method that could provide useful mechanistic insight into enzyme-catalyzed reactions.[4] By calibrating the model on the solution reaction, the EVB method allowed one to compare enzyme-catalyzed reactions with corresponding solution reactions.

We have presented a new *ab initio*/free energy (QM-FE) approach to studying enzyme-catalyzed reactions and the corresponding solution reaction. We have shown a method to estimate free energies for both solution and enzyme-catalyzed reactions with a QM-FE model. Our rigid model finds the TET-MICO free energy differences between solution and enzyme to be 3 kcal/mol while the flexible model finds a difference of 6 kcal/mol. The latter is reasonably consistent with Warshel's estimate of 8 kcal/mol for this free energy difference. However, our calculations suggest that the cratic free energy terms are the larger contribution to the free energy difference between

(56) In the case of MICO, we defined the orientation of His by fixing C_β to $C_\alpha-C{=}O$, C_γ to $C_\beta C_\alpha-C$, and N_δ to $C_\gamma-C_\beta-C_\alpha$. Any of the Z-matrix coordinates that involve at least one atom outside the QM region (C_α, C=O) are constrained; all others are optimized. In this way, all internal degrees of freedom of the monomer are optimized, and only the orientation of the monomer with respect to the protein backbone is fixed. The orientation of Ser was defined in an analogous manner, relating C_β, O_γ, and the O_γ hydrogen to the backbone atoms $C_\alpha-C{=}O$. For the substrate we explored two choices, relating either C_α (Arg) and the carbonyl C=O of the scissile bond to N−C−O(Phe) (variant A) or C_α (Ala), N, and C of the scissile bond to the carbonyl C=O (Ala) and terminal N (variant B). For TET we used the same protocol to define the orientation of His. The covalent complex of Ser and the substrate was attached to the backbone of Ser, very much in the same way as for MICO, except that the carbonyl C of the scissile bond now replaces the hydroxy hydrogen of Ser. It appeared necessary to constrain additional dihedrals which connect Ser and the substrate to avoid rotation of the substrate. In variant A, we only constrained $N-C(Sub)-O_\gamma(Ser)-C_\beta(Ser)$; in variant B, we also fixed $H-N-C(Sub)-O_\gamma(Ser)$. The geometry optimizations were performed in mixed internal and Cartesian coordinates. Defining the MM anchor atoms (dummy atoms in QM optimizations) in Cartesian coordinates and relating the QM atoms by Z-matrix coordinates to the MM coordinate frame avoids the necessity of subsequent rigid body rotations to reinsert the QM monomers into the protein coordinate frame. Performing MP2//6-31+G* calculations for the QM models of MICO and TET, we obtained results similar to those reported in Table 3. For MICO, variant A leads to a slightly higher energy (+1.2 kcal/mol), and variant B leads to a slightly lower energy (−0.3 kcal/mol). For TET, we now calculate energies that are 0.6 and 3.0 kcal/mol lower with variants A and B, respectively. All of these discrepancies are well within the expected accuracy of our combined QM-FE approach. Hence we conclude that the relative quantum mechanical energy (ΔE) is not highly sensitive to the exact choice of constraints in the QM geometry optimizations.

(57) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. *Int. J. Quantum Chem.* **1996**, *60*, 1189−1200.

(58) Dang, L. X.; Rice, J. E.; Caldwell, J.; Kollman, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 2481−2486.

solution and enzyme, a suggestion congruent with analysis of anhydride formation in model reactions.[59] Thus, we conclude that the enzyme catalysis in trypsin has the largest contribution from preorganization of the reacting groups and smaller, but significant contributions from the enzyme groups that stabilize TET such as Asp102 and the oxyanion hole. Further calculations on other enzyme-catalyzed reactions are required to see how general these results are.

The methodology presented here should be applicable to a wide variety of enzyme systems and biomimetic models.[60] One can also imagine variants, where one performs some fully coupled QM/MM with a semiempirical model and then locally reoptimizes and does single points at a higher QM level. Semiempirical models specifically reparametrized[61] to reproduce *ab initio* results may be useful in this regard. In any case, our approach to deriving charge models for the QM portion that can be used in the calculation of the interaction free energy of the QM atoms with their environment should be of general use.

Acknowledgment. We would like to thank the NIH for support (NIH-GM29072). D.B. gratefully acknowledges support from Deutsche Forschungsgemeinschaft through a research scholarship.

JA972723X

(59) Lightstone, F. C.; Bruice, T. C. *J. Am. Chem. Soc.* **1996**, *119*, 2595−2605.

(60) Zheng, B. L.; Breslow, R. *J. Am. Chem. Soc.* **1997**, *119*, 1676−1681.

(61) Bash, P. A.; Mackerell, A. D.; Lavine, D.; Hallstrom, P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3698−3703.