Simulations of enzymic reactions

A. J. Mulholland* and M. Karplus††

*Department of Chemistry, Harvard University, Cambridge, MA 02138, U.S.A., and †Laboratoire de Chimie Biophysique, Institut le Bel, Université Louis Pasteur, 67000 Strasbourg, France

A fundamental problem in biology is the explanation of enzyme activity at the molecular level. The detailed origin of the rate acceleration of chemical reactions by enzymes is not fully understood, in spite of the immense amount of experimental work that has been done in this area [1,2]. To describe an enzyme-catalysed reaction, its mechanism must be firmly established and intermediates must be identified. It is necessary to know the roles of specific residues in transition-state stabilization and reactant destabilization. Further, the contribution to the rate enhancement of conformational changes and dynamical effects, including possible tunnelling, must be determined. Only when the description of an enzymic reaction can be given at the level of detail available for simple reactions, such as the hydrogen atom–hydrogen molecule (H+H₂) exchange reaction, will the problem of enzyme catalysis be solved in any given case. Such a description of a reaction in a mesoscopic system, such as an enzyme in solution, is a formidable challenge. It requires contributions from both computational approaches and experimental measurements, as in the case of H+H₂ [3–5]. Results at the required level of detail can be obtained from theory; e.g. the energy profile for the reaction in the enzyme can be calculated and the roles of various interactions such as hydrogen bonding and electrostatic stabilization in the rate acceleration can be quantified. However, because of the approximations inherent in such calculations, structural, mutational and kinetics measurements are necessary as input and for verification of the validity of the theoretical results.

A major difficulty in simulating enzymic reactions is the large size of the systems involved. This is particularly clear when one contrasts the 10 000 or so atoms of an enzyme-plus-solvent system with the three atoms in the H+H₂ reaction. Some simplification can be achieved because the chemical changes at the active site usually involve directly only the substrate atoms, certain catalytic residues and cofactors. However, a large portion of the remainder of the enzyme and the solvent environment can be significant in catalysis and must be included in a full treatment, albeit in a more approximate way. To represent the bond breaking and bond making and the consequent electronic redistribution that occur in many enzymic reactions, the quantum mechanical nature of the process must be taken into account. This contrasts with the large number of successful calculations on macro-molecular systems [6,7], for which an empirical 'molecular mechanics' (MM) potential energy function is adequate. Quantum mechanical (QM) methods generally require large amounts of computer time and memory even for small molecules. *Ab initio* methods can calculate accurately the properties of isolated small molecules and transition states for reactions, if a sufficiently high level of theory is used [8]. Semiempirical molecular orbital techniques (such as the widely used AM1 [9] and PM3 [10] formulations) can be used to treat larger systems because simplify-
ing approximations make them much less computer intensive [11]. However, even semiempirical molecular orbital calculations are still too slow for calculating the reaction paths and dynamics of enzymic reactions. It is also important to note that, because of their approximate nature, semiempirical methods suffer from certain weaknesses that limit their scope and accuracy [12].

As already mentioned, most simulations of proteins are carried out with empirical MM-type force fields, which allow the potential energy of the system and the forces acting on the atoms to be calculated simply and quickly [7]. Typically, atoms are represented by point partial charges with van der Waals radii that give rise to nonbonded interactions, and simple terms (e.g. harmonic bonds and bond angles) are used to represent the bonding terms. Through extensive parameterization, MM potential functions provide accurate results for the structures and dynamics of macromolecules. However, the forms of the energy terms (e.g. the harmonic bonds) and the parameterization to local ground-state geometries mean that standard force fields cannot be applied to processes involving chemical reactions. One approach that has been used to model transition states and unstable species is to introduce special functions and parameters (for example, by fitting to ab initio results). This has been used to study transition states in complex organic molecules [13] and reactions in solution [14].

An attractive alternative for simulating enzymic reactions is the use of potentials that combine a QM description of a small region, including the active site and the substrate, with a molecular mechanics description of the bulk of the enzyme and the solvent [15]. Such QM/MM methods can treat electronic and chemical changes and take account of the effects of the environment. The combination of the versatility of a QM method with the simplicity and speed of a molecular mechanics force field allows reactions in large systems to be studied. It makes possible the determination of the energy and the free energy along the reaction path of enzymes and of reactions in solution [16]. Dynamic effects [11] can be studied with special techniques, such as activated dynamics [7] based on the QM/MM potentials. QM/MM methods using ab initio techniques have also been developed [17], but the computing cost is such that, even for restricted QM regions, they cannot be used for full determinations of a reaction path. The ab initio approaches are most useful for checking a few essential points along the reaction path from a semiempirical QM/MM treatment.

A QM/MM potential incorporating the AM1 semiempirical formulation in the MM/dynamics program CHARMM has been developed, and has been successfully used to model enzymic reactions [15,18]. The system to be studied is divided into three regions: a QM region, an MM region and a boundary region that takes account of the finite system size; for the last, either periodic boundary conditions or the stochastic boundary method can be used [15,19]. The atoms in the QM region, treated by AM1, are influenced by the partial charges of the MM atoms and the other interactions between the QM and MM regions are included consistently; e.g. the MM atoms interact with the QM charge distribution and MM-type terms are retained for any covalent interaction, including at least one MM atom. If the QM and MM regions involve different molecules, the QM/MM formulation is particularly simple; if the dividing line between the QM and MM regions separates covalently bonded atoms, the bonding requirements of the QM atoms are satisfied by introducing 'link' atoms, equivalent to AM1 hydrogens, to 'cap' broken bonds. Similar QM/MM approaches have been used by other workers for the investigation of processes in solution [20] and in enzymes [21]. A related method for the study of enzyme-catalysed reactions is the empirical valence bond technique [22], which combines a simple valence bond picture of the reaction with a MM model of the protein and solvent.

Care must be taken in partitioning a system into QM and MM regions to represent the reaction realistically, and to avoid disruption of the electronic structure of the active site. The accuracy of an approximate QM method should be tested by ab initio QM calculations of model systems or ab initio QM/MM calculations before each application. Other factors which are of importance for MM simulations generally, such as how to include solvent dielectric effects, also need to be considered.

Computer simulations of enzymic reactions have been reviewed recently [23]. The present discussion focuses on studies of three enzymes that illustrate several techniques that can be used to study the catalytic mechanism and provide significant insight into the origin of the rate acceleration.
**Triose phosphate isomerase (TIM)**

TIM catalyses the interconversion of dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate in glycolysis. TIM has been intensively studied by a wide variety of experimental techniques [1]. Glu-165 has been identified as the likely base for substrate deprotonation, and His-95 is well positioned at the active site to act as an acid and protonate the carbonyl oxygen of DHAP. An enediol intermediate is believed to be formed. For His-95 to function as the general acid, it was expected that it would be positively charged (doubly protonated) [24]. An important observation is that TIM polarizes the carbonyl oxygen of bound DHAP [25]. Evidence for the role of a positively charged His-95 as the acid and polarizing electrophile was strengthened by the finding that replacement of His-95 with neutral residues by site-directed mutagenesis greatly reduces enzyme activity and polarization of the DHAP substrate carbonyl group [26,27].

The reaction pathway in TIM has been calculated with the AM1 QM/MM potential in CHARMM [18,28]. The results supported the proposal of Glu-165 as the general base, and suggested the novel possibility [18,24] of neutral histidine acting as an acid, although a concerted mechanism with His-95 acting as a proton relay could not be excluded. The conclusion that the catalytic His-95 is neutral has been confirmed by NMR spectroscopy [29]. The selection of neutral imidazole as the general acid catalyst has been discussed in terms of achieving a pKₐ balance with the weakly acidic intermediate. This avoids the 'thermodynamic trap' that would result from a too stable enediol intermediate, produced by reaction with the more acidic imidazolium [18,24].

A perturbation analysis of the contributions of individual residues to the energetics along the reaction pathway showed that electrostatic interactions with many residues, some as far away from the active site as 14 Å, contribute to catalysis (Figure 1). Lys-12, which lies close to

**Figure 1**

Residue contributions to the QM/MM energy difference between the enediolate and the DHAP substrate in the TIM reaction, plotted versus their centre of mass distances from the substrate.

Some residues making important contributions are indicated. 'Vacuum' refers to the energy difference when the charges on all amino acids, other than those treated as QM atoms, were set to zero. (See [18] for details.)
the active site, makes the most important contribution to stabilizing the enediolate; a Lys-12 to Met mutant is inactive, but this does not verify its role in catalysis because the substrate does not bind [30].

The presence of a neutral His-95 raises the problem of why its replacement by glutamine or asparagine (also neutral) abolishes polarization of the DHAP carbonyl [26]. This question was investigated by semiempirical and \textit{ab initio} calculations on models of the active sites of wild-type and H95Q mutant TIM [28] for which crystal structures were available [27,31]. It was shown that, in the native enzyme, Lys-12 and Glu-165 make the dominant contributions to polarizing the DHAP carbonyl group, whereas the neutral His-95 has only a small effect. In the H95Q mutant, Lys-12 still exerts a polarizing influence, but the contribution of Glu-165 is greatly reduced, so that there is a much smaller overall shift in the carbonyl stretching frequency. This is because the side chain of Glu-165 has a somewhat different orientation in the H95Q mutant [27,28]. It hydrogen bonds with Gln-95 and is no longer in a position to polarize the substrate. The reduction in the frequency shift is therefore due not to the mutation itself, but to a subtle structural change induced by the mutation. These results [28] show the importance of calculations for interpreting the effects of mutations, which may otherwise be misunderstood, particularly when they yield results in accord with expectations.

**Citrate synthase (CS)**

Like TIM, CS catalyses abstraction of a proton from a carbon atom adjacent to a carbonyl group, and also contains histidine residues implicated as general acids in the mechanism, although there is uncertainty concerning the histidine protonation states [32]. CS catalyses the Claisen-type condensation of acetyl CoA and oxaloacetate to form citryl CoA, which is then hydrolysed to products [33]. An enzymic base deprotonates acetyl CoA to produce a nucleophilic intermediate [34,35], which attacks the carbonyl group of oxaloacetate. Debate has centred on the nature of the intermediate; it is not clear whether it is the enol or the enolate of acetyl CoA, nor is its stabilization by the enzyme understood. It has been suggested that the enolate of acetyl CoA is too unstable to be compatible with the observed rate of reaction, unless significantly stabilized by the enzyme [36]. Formation of the neutral enol directly from the keto substrate by concerted acid–base catalysis has been suggested to be a

**Figure 2**

Enolization of acetyl CoA in CS

(a) Mechanism of acetyl CoA enolization by CS, indicating the keto (reactant), enolate and enol forms of the substrate. (b) Energy profile for the reaction shown in (a), calculated by the AM1/CHARMM QM/MM method, in a 17 Å radius sphere around the active site. The total energy is plotted against interatomic distances as a measure of progress along the reaction coordinate: for the keto-to-enolate reaction (left-hand side) the distance between OD1 of Asp-375 and the reacting proton is shown, and for the enolate-to-enol reaction (right-hand side) the distance between the carbonyl (enolate) oxygen of acetyl CoA and the reacting proton is used [40].
more likely, lower energy alternative. Based on crystallographic and mutagenesis results, it was proposed that Asp-375 (in pig CS numbering) acts as the general base to deprotonate acetyl CoA, and positively charged His-274 as the general acid in a concerted reaction to form the enol of the substrate \([37,38]\). The enol would then attack the oxaloacetate carbonyl, which is made reactive by polarization at the active site \([39]\), with His-320 acting as a general acid in another concerted reaction to form citrul CoA \([37]\). However, following the discovery that the histidine that serves as the general acid in TIM is neutral \([18,29]\), it was noted that the hydrogen-bonding patterns of the two proposed general acids in CS, His-274 and His-320, indicate that they are neutral, as well \([32]\).

The enolization of acetyl CoA in CS has been studied by semiempirical, \textit{ab initio} and QM/MM methods \([40]\, A. J. Mulholland and W. G. Richards, unpublished work); the CHARMM program \([41]\) was used for the calculations. A suitable high-resolution structure of CS with an inhibitor and substrate bound was taken as a model of the reactant complex \([42]\). The results indicated that His-274 and His-320 are indeed neutral, and that Asp-375 is the catalytic base. The enolate form of acetyl CoA was calculated to be more stable, within the enzyme, than the enol (Figure 2). No evidence was found for a concerted reaction. Studies on models of the condensation reaction showed that the enolate is a much more reactive nucleophile than the enol, and that for this next stage of the reaction to proceed rapidly the intermediate must have a high degree of enolate character \([40]\). Calculations also indicated that, if His-274 were positively charged, the enol of acetyl CoA would be formed. This does not occur in CS, which instead utilizes a neutral imidazole to stabilize the enolate, which reacts quickly with oxaloacetate.

Two hydrogen bonds were found to provide much of the stabilization of the enolate necessary to prolong its lifetime and allow it to participate as an intermediate in the reaction. In the substrate complex, His-274 and a conserved water molecule donate hydrogen bonds to the carbonyl oxygen of acetyl CoA. On enolate formation, negative charge shifts to this atom, and the hydrogen bonds are strengthened, specifically stabilizing the intermediate. Another interesting feature highlighted by analysis of residue contributions, using the perturbation method developed for TIM \([18]\), is the hydrogen bond from the hydroxyl side chain of Ser-244 to His-274. Ser-244 is an absolutely conserved residue, and it is noteworthy that the arrangement of acetyl CoA--His-274--Ser-244 is similar to the Asp--His--Ser catalytic triad of the serine proteases. As with TIM, charged groups were calculated to exert a considerable influence on the reacting species, and the size and long range of these effects again emphasizes the need to include the enzyme environment in realistic calculations.

The results for CS are relevant to the recent proposal that 'short, strong' or 'low-barrier' hydrogen bonds are an important means of stabilizing charged intermediates in a wide variety of enzymic reactions \([43,44]\). Such bonds can be exceptionally strong in the gas phase, and have been proposed to have energies of up to 20 kcal/mol in enzymes, although their existence and exact strengths in proteins remain uncertain \([45,46]\). Bonds of the low-barrier type are proposed to form between partners of approximately equal \( pK_a \) when one is charged and bulk solvent is excluded. In CS it has been proposed that the hydrogen bond between His-274 and an 'enolic' intermediate, thought to be of similar basicity, is of this type, and is responsible for stabilizing the intermediate \([44,47]\). The term enolic is used to indicate that the proton is effectively shared between the hydrogen-bonded partners.

The calculations showed that the hydrogen bond between His-274 and the enolate is considerably stronger than that with the neutral acetyl CoA, and specifically stabilizes the intermediate. However, it is important to consider the net effect of interactions in determining their contributions. The effective stabilization is affected by the proximity of Asp-375, which is also involved in the reaction. The increased strength of the hydrogen bond with His-274 is calculated to stabilize the enolate by approximately 5 kcal/mol relative to the substrate. A hydrogen bond with a water molecule (Wat-385) provides a similar amount of stabilization. Although these energies may be underestimated somewhat by the methods used, the effective stabilization falls well short of the values sometimes suggested for similar hydrogen bonds in proteins. The crucial factor in the formation of a strong hydrogen bond is that one of the partners is charged, whereas matching of their \( pK_a \) values appears to be less important \([45,48]\). Thus, the energies of the enolate and enol forms do not
have to be equal within the enzyme for the hydrogen bond with His-274 to stabilize the intermediate. The nucleophilic intermediate in CS is best described as the enolate of acetyl CoA, stabilized by normal hydrogen bonds from His-274 and a water molecule. This conclusion is in accord with recent discussions on the role of 'low-barrier' hydrogen bonds in proteins [45,49].

**Rotamase catalysis by the FK506-binding protein (FKBP)**

Proteins that bind the immunosuppressant FK506 catalyse the isomerization of proline peptide bonds [50]. Although inhibition of this peptidylprolyl isomerase (also known as a rotamase) activity is not thought to be related to the immunosuppressive action of FK506, the mechanism of peptidylprolyl isomerases is of interest in itself, particularly because of their apparent role in protein folding [51]. Rotamases accelerate protein folding by catalysing prolyl trans to cis isomerization, which can be the rate-limiting step. Experimental studies have shown that a covalent intermediate is unlikely, but left the exact mechanism unclear [52]. Details of the mechanism have come from a calculation of the trans–cis prolyl isomerization pathway in FKBP [53]. Structures of the bound cis and trans forms of a peptide substrate and of the transition state for isomerization were constructed. This was done by model building based on the structures of the FK506/FKBP complex [54]; it has been confirmed by a Monte Carlo docking study of the system (A. Caflisch, S. Fischer and M. Karplus, unpublished work). It was shown that the substrate binds as a type VIa proline turn, with the C- and N-termini exposed, so as to permit binding of a protein with such a turn.

The standard MM energy function for the imide bond of a proline residue is designed for the essentially planar cis and trans forms found in a stable protein. Since semiempirical methods, such as AM1, are known to be inaccurate for peptide bond rotation, an *ab initio* energy map for isomerization of the proline dipeptide was calculated and used to correct the MM expression. The modified potential function for the imide bond was combined with the standard MM force field for the rest of the system to calculate the reaction pathway for the interconversion of the cis and the trans forms (Figure 3). A newly developed technique [55] to locate the true saddle points from an approximate reaction pathway was used; it does not require a predetermined reaction co-ordinate.

The energy barrier found for the cis to trans isomerization was approximately 6 kcal/mol, compared with 19 kcal/mol for a dipeptide in solution, in good agreement with experiment. The reaction pathway was analysed to determine the origin of the barrier reduction. The analysis demonstrated that the binding energy of the substrate/FKBP complex was greatest at the transition state, in accord with the Pauling hypothesis of the complementarity of an enzyme to the transition state [1,56]. Interestingly, the only interactions observed to bind the transition state specifically are relatively weak C-H⋯O hydrogen bonds. The dominant contribution was found to arise from relief of repulsive interactions which destabilize the reactant and product ground states. Repulsions between the imide carbonyl oxygen and residues Asp-37 and Tyr-92 destabilize the cis and trans forms, respectively, in agreement with mutational studies. These repulsive interactions lead to reactant and product minimum energy structures with a twisted imide bond, so that the energy of the reactant state is raised and the effective barrier to isomerization is lowered.

Another important contribution to catalysis arises from desolvation of the imide carbonyl oxygen. Interaction with water molecules causes the enthalpy of isomerization of prolyl peptide bonds to be about 3 kcal/mol higher in solution than the barrier for the isolated prolyl dipeptide.
Because the imide carbonyl is not hydrogen bonded to FKBP, the increase in barrier height associated with carbonyl polarization in solution is avoided, and no hydrogen bonds need to be broken during twisting of the imide bond. Finally, there is a contribution from substrate autocatalysis. The transition state is stabilized by a hydrogen bond between the pyramidalized imide nitrogen and the amide NH group of the neighbouring residue in the type VIa turn bound by the enzyme.

**Conclusion**

Theoretical studies of enzymic reactions are an important tool for understanding the origin of the rate enhancement. They complement data available from structural, kinetic, spectroscopic and mutagenesis investigations and can answer questions that are difficult to resolve experimentally. One approach uses potential energy surfaces that are calculated with a Hamiltonian that couples an approximate QM description of the reaction with a simpler MM representation of the bulk of the protein and solvent. Properly applied and tested, such QM/MM potentials can be used to determine accurate reaction paths and dynamics for large systems. They can quantify the interactions involved and determine the contributions of individual residues and other aspects of the environment to catalysis. Moreover, QM/MM methods are applicable, with little or no modification, to a wide range of systems. It is likely, therefore, that the fundamental goal of enzyme kinetics – a determination of how a specific enzyme ‘works’ at the atomic level – will be achieved in the near future.

The authors would like to thank their collaborators in the work described here; they are acknowledged in the listed references. This work was supported in part by a grant from the National Science Foundation. A.J.M. is a Wellcome Prize Travelling Research Fellow.
Brownian dynamics simulations of enzyme–substrate encounter

R. C. Wade
European Molecular Biology Laboratory, Meyerhofstrasse 1, 69012 Heidelberg, Germany

Introduction
The first stage in the catalysis of a substrate by an enzyme is the diffusion of the substrate towards its binding site. For some enzymes, this is the rate-limiting step. Such diffusion-controlled enzymes typically have high, viscosity-dependent, bimolecular rate constants (approx. $10^9$–$10^{10}$ M$^{-1}$ s$^{-1}$) and have been described as ‘perfect enzymes’ [1]. With continuing improvements and applications of protein engineering, it can be expected that the rates of other enzymes that are not normally diffusion controlled will be increased by mutagenesis towards the diffusion-controlled limit, and that new diffusion-controlled enzymes will be developed. To predict the rate constants of diffusion-controlled enzymes and to design modified or new diffusion-controlled enzymes, it is necessary to calculate the rate of diffusional encounter between the enzyme and its substrate and the impact of changes in the properties of the enzyme and the substrate on the rate.

The diffusional encounter rate between highly symmetrical objects can often be calculated analytically. However, enzymes and their substrates are generally far from symmetric: they are flexible and have irregular shapes and charge distributions. The effects of these features on the diffusional encounter rate can be evaluated by using a Brownian dynamics simulation methodology. The next section provides an overview of the method, and this is followed by examples of its application.

Methodology
The steady-state bimolecular diffusion-controlled rate constant $k_D$ for the diffusional encounter of