Simulation of Enzyme Reactions Using Valence Bond Force Fields and Other Hybrid Quantum/Classical Approaches

Johan Åqvist*,† and Arieh Warshel*,‡

Department of Molecular Biology, Uppsala University, Biomedical Centre, Box 590, S-75124 Uppsala, Sweden, and Department of Chemistry, University of Southern California, Los Angeles, California 90089-1062

Received February 3, 1993 (Revised Manuscript Received July 21, 1993)

Contents

1. Introduction

Ever since the first three-dimensional structures of enzymes were solved[1] it has been one of the fundamental challenges of molecular biology to establish and explain the relationships between enzyme structure and catalytic activity. Although it is now widely accepted that enzymes operate by lowering the activation energies of chemical reactions, it is not clear how such a stabilization of the transition state relative to the ground state is really accomplished. In principle, the transition state can be stabilized in many ways[2,3] and the fundamental issue is thus not whether the transition state is stabilized or not, but what the most important contributions to this stabilization are, and how the structure of the enzyme is related to its catalytic power. A partial answer to these questions can be provided by site-directed mutagenesis of enzyme active sites and other experimental approaches.[4] However, no experimental technique allows the entire energy profile for an enzymic reaction to be studied. It is also difficult to experimentally quantify the contributions from different types of catalytic effects, such as electrostatic stabilization, steric strain, entropy, etc., as well as the actual energy contributions associated with each amino acid residue of the protein.

This situation certainly challenges theoretical chemists to try to explain and model the catalytic effect of enzymes. The development of a completely reliable modeling approach would provide, at least in principle, the ability to explore the details of enzyme catalysis that are inaccessible to experiment. One could then also predict the outcome of any protein engineering experiment aimed at modifying the activity of a given enzyme. Such prospects, as well as the shear complexity of enzyme molecules and their enormous catalytic power, has attracted a lot of attention and effort to field of enzyme modeling and simulation. Of course, the progress in this challenging field has not quite yet reached the level where computer simulations can provide a real substitute for experiments. Nevertheless, useful insight is starting to emerge and it is already clear that computer modeling is becoming an essential part in the process of rationalizing experimental findings about enzyme catalysis.

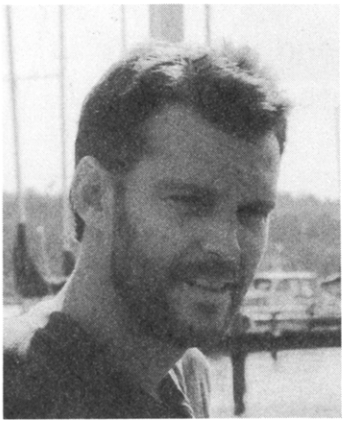
In this review we will try to summarize the progress that the field of enzyme modeling and simulation has witnessed since the first attempts were made in the mid 1970s. The focus will be on methods that attempt to treat the entire system in which the catalytic process takes place. The main problems with such approaches are associated with the incorporation of the environment surrounding the reacting groups. We will discuss different methods that have been used in the literature for accomplishing this and try to highlight their characteristic features. The latter part of the review then deals with applications of the methodology to various problems in enzyme catalysis and, in particular, we will illustrate some recent examples employing the empirical valence bond (EVB) method[2,44] in more detail.

2. Some Basic Problems in Describing Reactions in Complex Systems

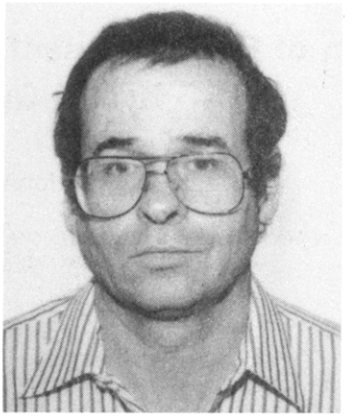
The difficulties in describing chemical reactions in enzymes as well as in solutions are associated with the complexity of these systems. That is, such systems do neither involve a small number of degrees of freedom as is the case for gas-phase systems, nor do they have any appreciable degree of symmetry as one finds in solids. Furthermore, enzyme reactions *always* involve at least two "phases", namely the protein itself and the surrounding solvent which is usually water. So, apart

† Uppsala University.
‡ University of Southern California.

0009-2665/93/0793-2523$12.00/0



Johan Åqvist was born in 1959 in Uppsala, Sweden. He received a M.Sc. degree in Engineering Physics at Uppsala University in 1983. His graduate research was carried out at the Department of Chemistry and Molecular Biology of the Swedish University of Agricultural Sciences under the supervision of O. Tapia. His thesis dealt with theoretical studies of protein dynamics and he received the degree of Doctor of Technology in 1987. From 1987 to 1989 he held a postdoctoral fellowship from the NFR and worked with A. Warshel at the University of Southern California. He then returned to Sweden to join the faculty at Uppsala University. Dr. Åqvist's research interests are in theoretical approaches to the understanding of protein structure–function relationships, in particular, enzyme reactions, protein–ligand interactions, ion channels and dynamic processes. He is the father of two children and among his other interests are skiing and yacht racing.



Arieh Warshel is a professor of Chemistry and Biochemistry at the University of Southern California. He was born in Kibutz Sdhe Nahum, Israel, in 1940. After receiving his B.Sc. from the Technion Institute in 1966, he went on to earn his M.Sc. in 1967 and Ph.D. in 1969 from the Weizmann institute Israel. His Ph.D. work with Professor Shneior Lifson involved the development of the Cartesian consistent force field approach which is the basis of many current programs for macromolecular simulations. From 1970 to 1972, he was a postdoctoral associate at Harvard University. From 1972 to 1977, he was senior scientist and associate professor at the Weizmann Institute and, during which time, he also was an EMBO fellow at the MRC in Cambridge, England. In 1976, he joined the faculty of the Department of Chemistry at the University of Southern California. Professor Warshel has pioneered several computer modeling approaches for the study of protein functions. These include the development of consistent treatments of electrostatic energies in proteins, quantum/classical methods for studies of the energetics of enzymatic reactions and the introduction of molecular dynamics simulations to studies of biological processes (in a study of the primary event in the visual process). Professor Warshel also introduced computer simulation approaches to studies of electron-transfer reactions and pioneered microscopic evaluation of thermodynamic cycles in biological systems, including the use of the free-energy perturbation method in studies of proteins. Professor Warshel is the author of the book *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, published in 1991 by Wiley and Sons. He has been involved in recent years in extensive simulation studies of enzymatic reactions, and other biological processes including photosynthesis and vision. He received the Alfred P. Sloan Fellowship and EMBO Senior Fellowship, the USC award for creativity in research, and the Phi Kappa Phi award for his book.

from the complication that enzyme molecules can contain thousands of atoms the effects of the surrounding medium must also be taken into account in any realistic description of the catalytic reaction process. This dilemma of complexity poses two major problems, namely, (1) that it is impossible to obtain reaction potential energy surfaces for the whole system by rigorous quantum mechanical (QM) methods and (2) that the dimensionality of such a potential energy surface is truly enormous. The latter point affects the feasibility of evaluating reliable minimum energy paths for the reaction by standard optimization methods. Since one also eventually wants to obtain *free energies* it is necessary that relevant parts of the configurational space can be sampled and this demands considerable computing power. To overcome these difficulties one has to resort to various approximations and we will discuss the most important of these below.

As it is not possible even with semiempirical methods to solve the Schrödinger equation for the entire protein/substrate/water system, one must adopt some kind of partitioning of the system into regions that are treated quantum mechanically and those that are treated classically. This division into a "reacting" part (r) and a surrounding (or "solvent") part (s) is common to most quantum treatments of solvated systems.

The general form of the Hamiltonian for this mixed system can be written

$$\mathcal{H}_{tot} = \mathcal{H}_r^0 + \mathcal{H}_{rs} + \mathcal{H}_s \qquad (1)$$

where $\mathcal{H}_r^0$ represents the vacuum Hamiltonian for the quantum system, $\mathcal{H}_{rs}$ the interaction between r and s and $\mathcal{H}_s$ the interactions within the outer system. The latter term may also reflect particular boundary conditions of the system. One might be led to believe that once the partitioning of eq 1 has been made, it is mainly on the accuracy of the QM method applied to the subsystem r that the overall effectiveness of the model hinges. This is, however, far from being true, as we shall emphasize below, and a reliable representation of the surrounding medium turns out to be a prerequisite for approaching a quantitative description of the reacting system.

With the above perspective in mind we will next consider different theoretical approaches that have been used in studies of enzymic reactions.

3. Calculations on Reacting Systems in Vacuum

The simplest way to approach enzyme reactions from a theoretical/computational perspective is to start by focusing on the chemistry of the isolated reacting fragments in vacuum. With this simple type of model system one can often afford to carry out high-level ab initio calculations of the corresponding gas-phase reaction. An illustrative example of a reaction for which numerous studies of this type has been reported is that catalyzed by carbonic anhydrase (CA), viz. the reversible hydration of carbon dioxide. The mechanism of CA,

deduced from experimental work,[5] can minimally be described as a two-step process involving water ionization and subsequent nucleophilic attack by hydroxide ion on the carbon dioxide substrate (a more detailed discussion of this mechanism is given in section 6.1).

Early calculations on the $OH^- + CO_2 \rightleftharpoons HCO_3^-$ reaction in vacuum by Jönsson and co-workers[6] demonstrated that the gas-phase reaction is barrierless and very exothermic. Subsequent studies that employed more sophisticated basis sets[7a-c] as well as semiempirical AM1 calculations[7d] have confirmed these conclusions. Encouragingly, the most accurate ab initio calculations reported for this system[7c] could reproduce the experimental gas-phase enthalpy of the reaction, but less accurate treatments resulted in substantial errors (see section 6.2). Several studies have been reported that include some basic elements of the enzyme's active site, in particular the catalytic zinc ion and its ligands.[7b,d,8] The latter which are histidyl groups are usually represented by simplified models such as ammonia. Interestingly, although not really unexpectedly, even such minimal descriptions of the active site produce rather drastic changes in the reaction profile compared to the bare gas-phase process. The interaction between the reacting species and the zinc complex affects the energetics considerably and various types of transition states, corresponding to different mechanistic pathways, could be characterized.[7b,d,8]

Many other calculations of this type have been reported that attempt to describe various aspects of enzyme mechanisms. Some important examples are the serine and thiol proteases,[9] carboxypeptidase,[10] alcohol and lactate dehydrogenases,[11] aspartic proteases,[12] triose phosphate isomerase,[13] superoxide dismutase,[14] and rubisco.[15] In order not to broaden this review too much we have felt it necessary to limit the discussion of the above type of calculations, since we want to focus on actual enzyme simulations. The interested reader is instead referred to the recent excellent review by Richards and co-workers.[16] However, we should perhaps emphasize that vacuum studies of the above type are of considerable interest in the sense that they allow the intrinsic gas-phase chemistry to be examined in detail. This may include the electronic characterization of different mechanistic pathways and their transition states which can be useful for interpreting various aspects of actual enzymic reactions. As will be discussed below, a common approach is also to use potential surfaces for the reacting fragments calculated in the gas-phase to describe these fragments also in solvated systems. The interaction with the surrounding medium is then simply added onto this vacuum surface.[7c,8g,17,18] Accurate vacuum calculations may also be of great use for calibration purposes in semiempirical treatments of solvated systems.

It is, however, becoming increasingly clear that gas-phase QM calculations themselves do not tell us so much about the energetics of reactions that take place inside enzymes. The catalytic effects provided by enzyme active sites can often be interpreted as a type of "solvation phenomena" and it is therefore not surprising that methods which do not include the surrounding protein or the solvent often fail to describe the energetics correctly. This will usually be the case even if a few catalytic groups (or, e.g., a metal ion) are included in the model, as long as no attempt is made to take into account the overall dielectric properties of the real system. In spite of this rather trivial point there has perhaps been tendency to overemphasize the capability of gas-phase quantum chemical methods in explaining enzymic reactions, and to overinterpret the corresponding gas-phase results. But, if used in a proper context, reliable QM methods for treating the reacting groups can be most valuable in various different respects when one is concerned with the modeling of enzymes.

4. Incorporating the Environment in the Calculations

It should thus be clear that an appropriate representation of the environment surrounding the reactants is crucial for describing chemical reactions in enzymes and in solution. The main problem, apart from the treatment of the reacting groups themselves, then becomes how to design such a model for the surroundings. If we consider the reacting fragments immersed in a *homogeneous* solvent the simplest type of approach would be to represent the latter as a dielectric continuum. The leading free-energy term in such a treatment, when the reacting groups are overall electroneutral, is the interaction between the solvent and the solute dipole moment (μ_r) which can be expressed in the reaction field formulation as[19]

$$\Delta G_{rs} = \frac{\mu_r^2}{a^3} \frac{(\epsilon - 1)}{(2\epsilon + 1)} \qquad (2)$$

where ϵ is the dielectric constant of the solvent and a is the effective radius of the cavity occupied by the reacting fragments. The reaction field model appears rather attractive considering the great qualitative progress that has been made with this approximation in the last 50 years. It should, however, be realized that the cavity radius is not reliably given by any theoretical considerations and thus appears rather as a parameter of the reaction field theory. It can, however, be successfully parametrized as in the applications of the generalized Born formula recently used by Cramer and Truhlar[20] or as described in the appendix of ref 21. While continuum treatments do neglect the detailed structure of the solvent they can still yield very good results for reactions in homogeneous media as has been demonstrated in ref 20b.

The continuum treatment becomes more questionable when one is dealing with enzyme reactions since it is not evident that the combined environmental effect of protein and solvent can be adequately modeled by a continuum dielectric model. It is also inevitable that one ends up with more or less undefined parameters such as dielectric constants and cavity radii (see ref 22 and the discussion in ref 68). While the latter could probably be satisfactorily parametrized (see above), it would clearly be undesirable to have to assume rather than calculate the dielectric properties of the system. It is also unlikely that any general parametrization of dielectric constants for solvated enzyme active sites is to be found at all, although this quantity can be calculated by microscopic simulations.[23]

At the expense of increased complexity the main alternative to continuum models are the microscopic ones in which both the protein atoms and the sur-

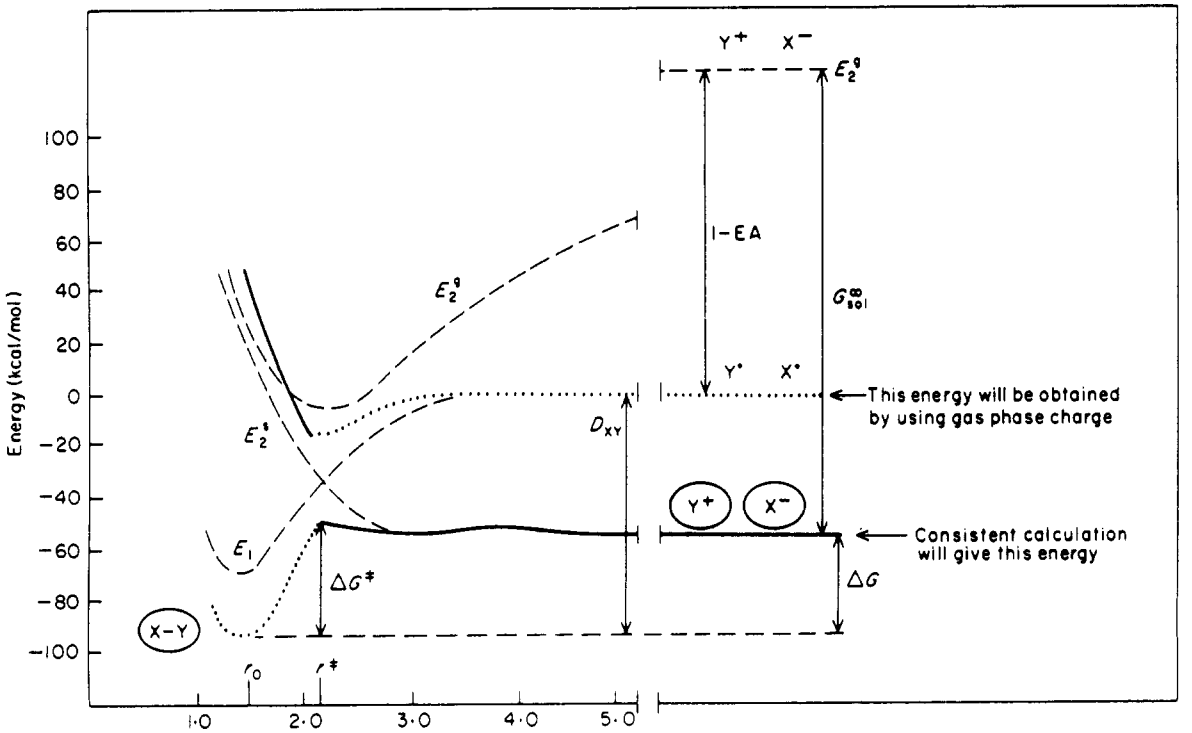


Figure 1. Energy diagram representing the heterolytic cleavage of a Y–X bond. The bond is described by a mixture of a covalent state (Y–X) and an ionic state ($Y^+ X^-$), with energies E_1 and E_2, respectively. While the energy curve E_1 is approximately unaffected by transfer from gas phase to solution, the gas-phase curve for the ionic state, E_2^g, is shifted downward in energy due to stabilization of the ion pair by the polar solvent, yielding E_2^s. The ground state obtained by mixing E_1^g and E_2^g becomes a pure covalent state ($Y^{\bullet}$ $X^{\bullet}$) with neutral atoms, when the bond is broken in the gas phase. Using this gas-phase charge distribution in the calculations will give a solvation energy close to zero and a large dissociation energy (typically ~90 kcal/mol) for the bond-breaking reaction in solution. On the other hand, if the solvent effect is included in the solute Hamiltonian, one obtains a solvated ionic ground state ($Y^+ X^-$) for the broken bond in solution. The corresponding error is around 60 kcal/mol for the case described here.

rounding solvent molecules are explicitly treated. Simplified solvent models have also been invented to reduce the computational cost. In the first microscopic theoretical study of enzyme catalysis, Warshel and Levitt[24] introduced a solvent model that represents each molecule by a polarizable point dipole located on a three-dimensional grid with cubic unit cell. The dipoles were assumed to obey the Langevin polarization law and accordingly the model was named the Langevin dipole (LD) model. With increased computer power the focus nowadays is turning toward explicit all-atom representations of the entire protein/water system.[16,25] This is, if affordable, of course the most desirable type of model since it can provide more accurate structural information.

With the system partitioning prescribed by eq 1 the degrees of freedom belonging to the outer system, s, are usually treated classically (see exception in ref 26). Thus, in fully microscopic schemes $\mathcal{H}_s$ is represented by an empirical force field of the standard molecular mechanics (MM) type.[24] The quantum treatment is then restricted to the evaluation of the potential energy surface associated with the vacuum Hamiltonian $\mathcal{H}_r^0$, perturbed by the interaction with s via $\mathcal{H}_{rs}$. Charge transfer between r and s is usually not considered and the potential energy associated with $\mathcal{H}_{rs}$ can then be represented by a function of the type

$$V_{rs} = \sum_{i,k} \frac{Q_i q_k}{r_{ik}} - \frac{1}{2}\sum_k \alpha_k \vec{\xi}_k^{\,2} + \sum_{i,k} \frac{A_{ik}}{r^{12}} - \frac{B_{ik}}{r^6} \qquad (3)$$

where the indices i and k run over the quantum and classical atoms, respectively. The Q_i's are the effective charges on the atoms in r and the q_k's denote the permanent partial charges in the outer system. The second term in eq 3 is the induced electronic polarization of the system s and the third term denotes the van der Waals interaction (represented by a Lennard-Jones potential here) between the two systems.

Having a reasonable representation of the solute–solvent interaction term is, however, not by itself sufficient for obtaining accurate results, even if $\mathcal{H}_r$ for the isolated reacting system is given by a very accurate method. For instance, if one takes the best *gas-phase* charge distribution of the reacting region and simply evaluates its interaction with the solvent, such a procedure will not correctly reflect the polarization of the "solute" by the field from the solvent and may correspond to a completely erroneous solute charge distribution. Perhaps the best illustration of this problem is to consider the heterolytic bond cleavage depicted in Figure 1.[27] As seen from the figure, the most reliable gas-phase calculations will give neutral fragments as the product when $r \rightarrow \infty$, since this is the correct result in the gas phase for the given electron affinity and ionization energy. In this case V_{rs} will be approximately zero. Only when the solute Hamiltonian, $\mathcal{H}_r$, feels the solvent polarization will we get a dissociation to X^+ Y^- for $r \rightarrow \infty$. In this case the free energy associated with V_{rs} will correspond to the large solvation energy of the dissociated ion pair. We will consider ways of coupling the solvent polarization to the solute Hamiltonian, in practice, below.

With a reasonable model for the solvent and a clear realization of the importance of the solute–solvent coupling, the next issue becomes the choice of the QM treatment for obtaining the effective wave function and charge distribution of the subsystem r in the field imposed by s.

The main alternatives for representing the wave function of the quantum system are the molecular orbital (MO) and valence bond (VB) approaches. We will briefly discuss different implementations of these methods for modeling enzyme systems below and try to highlight some of their characteristic features.

4.1. MO-Based Methods

The first example of a coupled hybrid quantum/classical scheme for describing enzymatic reactions[24] made use of semiempirical (QCFF/ALL,[24] MINDO/2[28]) MO methods within the standard Hartree–Fock SCF scheme. In these calculations the classical part of the enzyme/substrate system was treated by a molecular mechanics force field that includes electronic polarizability and the solvent was represented by the LD model mentioned above. The coupling of the solvent field to the solute Hamiltonian was introduced by formally considering the solute and the solvent as a supermolecule and then freezing the solvent electron distribution, while neglecting the overlap between the solute and the solvent orbitals. The resulting Fock matrix for the SCF calculations of the solute electronic states was expressed as[24] (see ref 21 for a recent discussion of this treatment)

$$F^{r}_{\mu\mu} = (F^{r}_{\mu\mu})_0 - \sum_k \boldsymbol{m}_k \boldsymbol{r}_{kA}/r^3_{kA}$$

$$F^{r}_{\mu\nu} = (F^{r}_{\mu\nu})_0 \qquad (4)$$

where μ designates an atomic orbital on the Ath solute atom, and k runs over the solvent molecules which are described here in terms of their dipole moment $\boldsymbol{m}$. The solvent dipole involved permanent and induced components which respond iteratively in a *nonlinear*, self-consistent way to the field from the solute charges (which in turn depends on the solvent polarization). The dipole term in eq 4 corresponds to the potential of the solvent at the Ath solute atom. Obviously, when the "solvent" involves protein residues the potential of the corresponding charges and induced dipoles is also included in $F_{\mu\mu}$. The quantum/classical model obtained by using eq 4 together with the protein force field thus provided the potential surface for the substrate–enzyme–solvent system. The energy profile for the reacting system was then evaluated by using a conventional energy-minimization technique to relax the system along the reaction path by a procedure which has been named adiabatic mapping.[29] Later versions of this hybrid quantum/classical model use an all-atom solvent model and a molecular dynamics (MD) simulation approach.[21,30]

An analogous approach that combines ab initio or selfempirical quantum models with a MM treatment of the surroundings, including explicit water molecules, has been developed by Kollman and co-workers[31] and used in studies of the catalytic reaction of trypsin.[32,74b] This implementation has also been used by Waszkowycz et al.[33] in a recent study of phospholipase A_2. A similar type of combined QM/MM potential utilizing the semiempirical AM1 parametrization[34] has been developed by Karplus and co-workers[35a] and applied, e.g., to the catalytic reaction of triose phosphate isomerase.[35b]

Tapia and co-workers[36] presented a QM scheme for solving the nonlinear Schrödinger equation that arises from considering the reaction field with which the system s responds to quantum charges in r. This approach is conceptually rather similar to that of Warshel and Levitt,[24] but it differs both with respect to the QM method used (CNDO-INDO in ref 36b) and in its representation of the surrounding medium s. The essence of the approach of Tapia et al. is that the dielectric properties of the surrounding protein are modeled by a dielectric response tensor, $\boldsymbol{g}$, that may be determined either by theoretical procedures or from experimentally available data. It is, however, not clear how the solvent around the protein would best be incorporated within this framework. That is, it would either have to be incorporated in a (statistically significant) average way or one would have to sample a large number of configurations for each of which the nonlinear Schrödinger equation would be solved. Several other authors have also employed reaction field (RF) methods for studying enzymic reactions.[37,38] Papain[37a,b] and actinidin[37c] are examples for which these types of calculations have been carried out. The approach of van Duijnen and co-workers that was used in ref 37a is of particular interest since it combines the RF approach with a microscopic representation of water. This treatment is much more reasonable than earlier ones that did not include the solvent at all and, e.g., seemed to overestimate the so-called helix dipole effect[9e] (see discussion in ref 38).

The earlier mentioned approach of Cramer and Truhlar[20] is a more recent example of the combination of a semiempirical MO approach with a continuum dielectric model, although not designed for describing enzyme reactions. This method, which relies on an empirical calibration of atomic parameters related to solvation (viz., "Born radii", surface tension parameters, etc.), seems quite promising for describing solvation effects in homogeneous media upon electronic properties. One of the nice features with the method is that although it treats the solvent macroscopically it attempts to take nonpolar interactions between solute and solvent into account and can thus reproduce some hydrophobic effects.[20a]

Various other interesting implementations of continuum formulations of the reaction field type have also been reported in the literature.[39,40] It seems, however, that continuum models will always encounter difficulties when the chemical reaction takes place in an inhomogeneous environment such as an enzyme active site surrounded by solvent. The progress that has been made in solving electrostatic problems involving solvated proteins by discretized continuum methods might turn out to be useful also in this respect.[41] The main drawbacks with continuum approaches, however, is probably the fact that they disregard structural details of the environment, which are clearly important for some phenomena, and also that they are not so easily integrated with dynamic simulation methods in a consistent way.

We have used the term "coupled hybrid" schemes here for the type of implementations discussed above to emphasize that the potential from the classical system actually enters into the Hamiltonian of the quantum region as a reaction field. There are, however, also several examples in the literature of what might be called decoupled hybrid schemes.[7c,8g,17,18,80] In this type of approach the unperturbed vacuum wave function and charge distribution of the subsystem *r* is calculated for the *isolated* fragments. The surrounding medium, *s*, then interacts with *r* without influencing its (vacuum) charge distribution. Often the calculated gas-phase geometry of the reacting fragments is also kept fixed so that the interaction with *s* does not allow any conformational effects on *r*. This type of approach has been used both in studies of enzymes[8g,80] and solution reactions.[7c,17,18] One may raise some criticism to the above procedure on the grounds that the electronic structure of the quantum system is not allowed to respond to the "solvent" reaction field. As was indicated in Figure 1, the problems can be expected to be particularly severe for charge-separation reactions. Studies adopting the above approximation have also exclusively dealt with charge-transfer processes (in which no new charges are created) rather than charge-separation reactions.

In the realm of MO schemes for calculations of enzyme reactions it would certainly be desirable to work at the ab initio level with inclusion of correlation effects etc., but this is usually too expensive. That is, accurate ab initio calculations can at present only be carried out for small fragments and, in particular, if we want to perform statistical averaging over many configurations (see below) which is a requirement for obtaining *free energies*, they become prohibitive. Nevertheless, ab initio solvation free energies were calculated recently using a classical force field as a reference state;[26] this trick allows one to evaluate the ab initio energy much less frequently than in a direct calculation. The local density functional (LFD) method[42] may provide a cheaper alternative to traditional ab initio schemes and more applications of this kind will probably emerge in the future. The only application reported[43] so far, however, evaluated the change in charge distribution of the reacting fragments in an enzyme active site (compared to vacuum) and this type of problem can easily be addressed by semiempirical models. The true challenges for density functional methods would rather be the evaluation of reliable reaction potential surfaces.

One of the main questions to be asked is of course what degree of accuracy can be attained with a given method and it is certainly not trivial to devise reliable models for reactions in such complex systems as solvated enzymes. Both ab initio and semiempirical MO methods are today quite successful in describing vacuum reaction energetics, at least for small systems. In solutions and enzymes, however, the situation becomes more difficult as we have very large solvation energy terms entering (on the order of 100 kcal/mol for charged groups). Therefore, unless these energy terms can be accurately accounted for, the efforts spent on obtaining accurate vacuum wave functions will be more or less wasted. In this respect, it can be most useful to calibrate the system Hamiltonian not only against gas-phase data (as is usually done with semiempirical MO Hamiltonians) but also against solution experiments.[2] This is one of the big advantages with the empirical valence bond (EVB) method[2,44,48] that will be briefly outlined below.

4.2. The EVB Model

One of the appealing features of VB methods is that their basic concepts, viz. bond functions and ionic terms, have a simple and clear physical meaning. Consequently, it is at least conceptually easy to define different states along a chemical reaction path in terms of VB configurations. It is, however, a more tricky business to try to solve the Schrödinger equation ab initio by expanding the wave function in terms of usual VB functions. This is basically due to mathematical difficulties associated with constructing a sufficiently complete basis in terms of VB resonance structures and with their mutual orthogonality properties. (Many of the drawbacks with the traditional VB formulation can, however, be overcome by more modern approaches.[45]) It turns out, however, that the VB framework easily lends itself to parametrization, whereby the number of resonance structures can be reduced considerably while still retaining an accurate description of the potential surface. This is to say, that with a fairly small number of VB functions it is possible to fit a VB model so that it, e.g., reproduces a given ab initio surface obtained by MO methods or available gas-phase experimental data on the relevant reaction.[46] Moreover, it is possible to use *solution experiments* to calibrate the VB Hamiltonian thereby avoiding some of the problems associated with incorporating surrounding medium effects on a vacuum potential.[46a] This is again due to the fact that the most important VB configurations for a given reaction do have a clear physical meaning and that it is possible to interpret the different structures (e.g., reactants and products) along a reaction path in terms of these VB states. The analytical form of these VB functions can be made rather simple by making use of appropriate MM potentials. One should also emphasize that the errors introduced by reducing the number of resonance structures do not at all become as serious as in ab initio VB formulations, since the calibration procedure is used to assure that the reaction surface asymptotically behaves in accordance with experimental facts (or whatever information is used for the parametrization). These are the basic ideas behind the EVB model which has been described in full detail elsewhere.[2,27,44,48,75] The interested reader may also consult the early VB treatment of hydrogen bonding in ref 47 which partly inspired the treatment of the solute in the present EVB method. In order to illustrate the method more clearly we will examine a simple test case.

4.2.1. VB Potential Surface for Proton-Transfer Reactions in Solutions

Let us consider a proton-transfer reaction in solution, which can be written as

$$RXH + YR' \rightarrow RX^- + HY^+R' \qquad (5)$$

In the language of VB theory, such a reaction can be described by the three resonance structures

$$\phi_1 = \mathrm{RX{-}H\ YR'}$$

$$\phi_2 = \mathrm{RX^-\ H{-}Y^+R'} \qquad (6)$$

$$\phi_3 = \mathrm{RX^-\ H^+\ YR'}$$

The electrons involved in the actual reaction (referred to here as the active electrons) can be treated according to the general prescription of the four-electron three-orbital problem with VB wave functions.[48]

$$\Phi_1 = N_1\{|\mathrm{XH\bar{Y}\bar{Y}}| - |\mathrm{\bar{X}HY\bar{Y}}|\}\chi_1 = \phi_1\chi_1$$

$$\Phi_2 = N_2\{|\mathrm{X\bar{X}H\bar{Y}}| - |\mathrm{X\bar{X}\bar{H}Y}|\}\chi_2 = \phi_2\chi_2 \qquad (7)$$

$$\Phi_3 = N_3|\mathrm{X\bar{X}Y\bar{Y}}|\chi_3 = \phi_3\chi_3$$

where X, Y, and H designate atomic orbitals on the corresponding atoms, the N's are normalization constants and the χ's are the wave functions of the inactive electrons moving in the field of the active electrons. We have thus partitioned the molecular electronic space into active and inactive parts and assumed no interaction between these parts. The three resonance structures Φ_1, Φ_2, and Φ_3 can be treated by the approach detailed in ref 48 and can be reduced to an effective two-state problem, where one state is mostly Φ_1 and the other is mostly Φ_2. The corresponding matrix elements can be evaluated by standard quantum chemical methods but this evaluation is very tedious. Instead we can exploit the simple physical picture of Φ_1 and Φ_2 and describe H_{11} and H_{22} by analytical potential functions that can be calibrated by both experimental information and accurate quantum mechanical calculations. That is, the function H_{11} will be given, at the range where the X···Y distance is large as compared to the X–H bond length, by a Morse potential function that depends on the distance $R_{X\text{-}H}$. When the H atom approaches Y we have electrostatic and van der Waals interactions between the groups. We can describe both of these forces using analytical potential energy terms (see below). The same argument applies to H_{22}. As far as H_{12} is concerned we can approximate it by an exponential term and fit the parameters in this term to experimental information on the gas-phase potential energy surface of the reaction or, if needed, to accurate gas-phase calculations.[2,46,48] One may also chose different functional forms for the off-diagonal matrix elements.[46b] The main issue here is that the energetics and geometry of the transition-state region, where the "coupling" between H^0_{11} and H^0_{22} is stronger, can be reasonably well represented. The efficiency of the EVB model in this context has recently been examined by Chang and Miller[46b] who found that gas-phase ab initio surfaces can be reproduced with high accuracy.

We will thus describe the gas-phase potential by

$$\epsilon^0_1 = H^0_{11} = \Delta M(b_1) + U^{(1)}_{\mathrm{nb}} + (K/2)(\theta_1 - \theta^0_1)^2 + U^{(1)}_{\mathrm{inact}}$$

$$\epsilon^0_2 = H^0_{22} = \Delta M(b_2) + U^{(2)}_{\mathrm{nb}} + (K/2)(\theta_2 - \theta^0_2)^2 + \alpha^0_2 + U^{(2)}_{\mathrm{inact}} \qquad (8)$$

$$H^0_{12} = A\exp\{-\mu(r_3 - r^0_3)\}$$

where b_1, b_2 and r_3 are, respectively, the X–H, H–Y, and X···Y distances, and θ_1 and θ_2 are, respectively, the R–X–H and H–Y–R′ bond angles, ΔM is a Morse potential function taken relative to its minimum value ($\Delta M_b = M(b) - D$) and U^i_{nb} is the nonbonded interaction in the given configuration. The parameter α^0_2 expresses the difference between the energy of ψ_1 and ψ_2, where the fragments of each resonance structure are at infinite separation. The potentials $U^{(i)}_{\mathrm{inact}}$ represent the interaction with the inactive part of the reacting system and are described by:

$$U^{(i)}_{\mathrm{inact}} = \langle\chi_i|\mathbf{H}_{\mathrm{inact}}|\chi_i\rangle$$

$$= {}^1/_2\sum_{\mathrm{bonds}}K^{(i)}_b(b - b^{(i)}_0)^2 + {}^1/_2\sum_{\mathrm{angles}}K^{(i)}_\theta(\theta - \theta^{(i)}_0)^2 + \sum_{\mathrm{torsions}}K^{(i)}_\phi[1 + \cos(n^{(i)}\phi^{(i)} - \delta^{(i)})] + U^{(i)}_{\mathrm{nb,inact}} \qquad (9)$$

where the b's, θ's, and ϕ's are, respectively, the bond lengths, bond angles, and dihedral angles in the fragments R and R′.

The effect of the solvent (or, more generally, the surrounding medium) on our reaction Hamiltonian is obtained by adding the corresponding energies to the diagonal matrix elements

$$\epsilon^S_1 = H_{11} = H^0_{11} + V^{(1)}_{\mathrm{rs}} + V_{\mathrm{ss}}$$

$$\epsilon^S_2 = H_{22} = H^0_{22} + V^{(2)}_{\mathrm{rs}} + V_{\mathrm{ss}}$$

$$H_{12} = H^0_{12} \qquad (10)$$

where $V^{(i)}_{\mathrm{rs}}$ is the interaction potential between the solute atoms in the ith VB configuration and the surrounding solvent and V_{ss} the solvent–solvent interaction. The ground-state potential surface for the reaction is then obtained by solving the secular equation

$$\boldsymbol{H}\mathbf{C} = E_g\mathbf{C} \qquad (11)$$

A theory that resembles the above EVB approach has recently been presented by Kim and Hynes.[40] Their model treats the solvent as a continuum, obtaining a nonlinear Schrödinger equation, but in contrast to most other continuum formulations it also considers the effect of electronic polarizability or nonequilibrium coupling between the solute and solvent (these effects are usually explicitly treated in the microscopic EVB model[2]). The qualities of this description have been illustrated for S_N1 ionic dissociation using valence bond type of Hamiltonians.[40]

4.2.2. The EVB Calibration Procedure

A key feature of the EVB method is its unique *calibration* possibilities that allows for the incorporation of reliable experimental information into the Hamiltonian. That is, returning to our proton-transfer example, after evaluating the free-energy surface (see below) with the initial parameter α^0_2 we can use the fact that the free energy of the proton-transfer reaction is given by

$$\Delta G_{PT} = \Delta G[(A - H + B) \rightarrow (A^- + HB^+)] = 2.3RT[pK_a(A - H) - pK_a(B^+ - H)] \quad (12)$$

Now we can adjust α_2^0 until the calculated and observed ΔG_{PT} coincide. This calibrated surface can then be used with confidence for studying the reaction in different solvents (or environments such as an enzyme's active site) since α_2^0 remains unchanged and only $V_{sol}^{(i)}$ is recalculated. In this way the error associated with the evaluation of H_{ii}^0 does not affect the calculations of the relative effects of different surrounding media.

In general, when one deals with a more complicated reaction, for which it is hard to obtain gas-phase estimates of α_i^0, it is convenient to use solution experiments to obtain the first estimate of α_i^0. This is done by using

$$\epsilon_i(\infty) - \epsilon_1(\infty) \simeq \alpha_i^0 + \Delta g_{sol}^i(\infty) - \Delta g_{sol}^1(\infty) = \alpha_i^0 + \Delta\Delta g_{sol}^i \simeq \Delta G_S^i(\infty)_{obs} \quad (13)$$

where $\epsilon_i(\infty)$ is the energy of the ith resonance structure when the corresponding fragments are held at infinite separation. Δg_{sol}^i is the solvation energy of the ith resonance structure and can be estimated by the free-energy perturbation approach that will be described below or by simpler models.[2,46a] Similarly, $\Delta G_S^i(\infty)_{obs}$ is the free energy involved in forming the ith configuration from the first configuration where the fragments in each configuration are held at infinite separation, this leads to the useful estimate:

$$\alpha_i^0 \simeq \Delta G_S^i(\infty)_{obs} - \Delta\Delta g_{sol}^i(\infty) \quad (14)$$

The off-diagonal matrix elements H_{ij} can also be determined by ab initio calculations (see discussion in ref 48) or by semiempirical procedures (e.g. ref 44). In most applications, however, we have followed the procedure outlined in ref 48 and represented H_{ij} by a simple function

$$H_{ij}^0 = \sum_{(k,l)} A_{ij}^{(k,l)} \exp\{-\mu_{ij}^{(k,l)} r_{kl}\} \quad (15)$$

where the atom pair (k,l) is chosen according to the specific H_{ij}. This function is fitted to experimental information about the activation free energy for the different steps of the relevant *reference* reaction in solution, using

$$(\Delta g_{i\rightarrow j}^{\ddagger}(H_{ij}))_{calc,w} = (\Delta g_{i\rightarrow j}^{\ddagger})_{obs,w} \quad (16)$$

This procedure is less reliable than that used for the diagonal energies and can benefit from ab initio calculations on the gas-phase reaction (see ref 48), which can be used as extra constraints on the parameters of eq 15. However, the calculated *difference* between the free-energy surface in solution and in the enzyme is not very sensitive to the exact value of the H_{ij}'s. It has previously been demonstrated that the dependence of $\Delta g^{\ddagger}$ on the reaction free energy is often almost linear.[48] Moreover, the relation between $\Delta g^{\ddagger}$ and ΔG_0 is virtually independent of the magnitude of the particular H_{ij}. (This is one of the reasons for why linear free-energy relationships have proven to be so powerful in physical organic chemistry.[49,50])

5. Evaluation of Reaction Free-Energy Profiles by the Thermodynamic Perturbation Technique

What we really want with a microscopic modeling approach for enzyme reactions is, of course, the ability to calculate the relevant rate constants and also how these are affected by various perturbations done to the system, e.g., by site-directed mutagenesis experiments. The rate constant for a given reaction can usually be written in terms of the free-energy difference between the reactant and transition states as

$$k = Fk_{TST} = F\frac{k_B T}{h} e^{-\Delta g^{\ddagger}/k_B T} \quad (17)$$

where k_B and h are the Boltzmann and Planck constants, respectively. k_{TST} is the transition state theory (TST) rate constant and F is the transmission factor that contains dynamical corrections to TST. This factor represents the probability that a trajectory, in which the system has enough energy to overcome the barrier, actually *will* cross and not be reflected back to the ground state. The most important factor in determining the rate is, however, the Boltzmann factor $e^{-\Delta g^{\ddagger}/k_B T}$ or more precisely the free-energy difference $\Delta g^{\ddagger}$. So far, we have only discussed how to obtain the potential energy surface $V_{pot}(\boldsymbol{X})$ for a given reaction. Although this entity is clearly of interest in itself, since it represents the zero temperature enthalpy surface, the main obstacle now is to calculate the corresponding free-energy function $\Delta g(\boldsymbol{X})$. To be formal, one should make a distinction here between the Gibbs' and Helmholtz' free energies, depending on which ensemble we chose to deal with (constant volume or pressure), but this difference is of little concern for actual calculations.

It is well-known that the evaluation of the free energy through direct calculation of the partition function is not tractable due to sampling and convergence difficulties.[51] The most successful approaches instead make use of perturbation and so called umbrella sampling procedures.[51,52] The perturbation formula[53] for the free-energy difference between two states turns out to be very useful for computer implementation and reads

$$\delta G(1 \rightarrow 2) = -RT \ln\langle \exp\{-(\epsilon_2 - \epsilon_1)/RT\}\rangle_1 \quad (18)$$

Here, $\epsilon_1(\boldsymbol{X})$ and $\epsilon_2(\boldsymbol{X})$ are the potential energy functions for the two states and $\langle\ \rangle_1$ denotes an ensemble average obtained on the potential ϵ_1. Although the expression of eq 18 is exact, it is not of much use if the minima of ϵ_1 and ϵ_2 are far away from each other in the configurational space, $\boldsymbol{X}$. This is again due to sampling problems which makes eq 18 converge very slowly. However, by introducing a coupling parameter approach the two states can be "connected" via a set of intermediate potentials that are sufficiently alike that adequate sampling can be obtained for adjacent potentials. This is usually done by defining these "mapping" potentials as[54]

$$\epsilon_m = \theta_1^m \epsilon_1 + \theta_2^m \epsilon_2$$

$$\theta_1^m + \theta_2^m = 1 \quad (19)$$

We can now envisage the two states above corresponding, e.g., to the two states ϕ_1 and ϕ_2 in our proton-

transfer example of section 4.2.1. By changing the mapping vector $\vec{\theta} = (\theta_1, \theta_2)$ in small increments in an MD simulation from (1,0) to (0,1), we can thus drive the system from the reactants via the transition state to the products.

The free energy associated with changing ϵ_1 to ϵ_2 in n discrete steps can now be obtained from

$$\delta G(\vec{\theta}_m \rightarrow \vec{\theta}_{m'}) = -RT \ln\langle \exp\{-(\epsilon_{m'}-\epsilon_m)/RT\}\rangle_m \quad (20)$$

$$\Delta G(\vec{\theta}_n) = \Delta G(\vec{\theta}_0 \rightarrow \vec{\theta}_n) = \sum_{m=0}^{m=n-1} \delta G(\vec{\theta}_m \rightarrow \vec{\theta}_{m'}) \quad (21)$$

where the average $\langle\ \rangle_m$ is evaluated on the potential surface ϵ_m. One can also modify the mapping potential ϵ_m by instead writing it as[27]

$$\epsilon_m = \theta_1^m\epsilon_1 + \theta_2^m\epsilon_2 - 2|H_{12}|\sqrt{\theta_1^m\theta_2^m} \quad (22)$$

For $\vec{\theta} = (1/2, 1/2)$ the mapping potential will then approach the true ground-state potential at the transition-state region, $E_g^\ddagger = 1/2(\epsilon_1 + \epsilon_2) - |H_{12}|$.

It must now be emphasized that $\Delta G(\vec{\theta})$ represents the free energy associated with moving on the constraint potential ϵ_m. We therefore still need to obtain the free energy, $\Delta g(\boldsymbol{X})$, corresponding to the trajectories moving on the actual ground-state potential $E_g(\boldsymbol{X})$. This is done using a trick that is sometimes referred to as umbrella sampling[52a]

$$\exp\{-\Delta g(X^n)/RT\} \approx \exp\{-\Delta G(\vec{\theta}_m)/RT\} \times \langle \exp\{-[E_g(X^n) - \epsilon_m(X^n)]/RT\}\rangle_m \quad (23)$$

where the reaction coordinate X^n can be defined in terms of the energy gap, $\Delta\epsilon = \epsilon_1 - \epsilon_2$, between the potential surfaces.[2] What this means is that we calculate the energy difference between the mapping potential and the ground-state potential (given by eq 11) at each point of the MD trajectory and use the Boltzmann average of this difference to correct the free energy obtained on the mapping potential. In connection with the EVB scheme outlined above, we can now use $\Delta g(X^n)$ to determine the values of $(\Delta g^\ddagger_{1\rightarrow 2})_{calc,w}$ and $(\Delta G_{1\rightarrow 2})_{calc,w}$, and adjust $\alpha^{(2)}$ and H_{12}^0 of eq 8 until the calculated and observed values of these free energies coincide and satisfy eqs 13 and 16.

The advantage of the EVB approach in implementing this free-energy perturbation (FEP) technique should now become obvious. That is, since the VB basis functions closely correspond to the actual reactant and product states of the reaction they are ideally suited to be used in eq 19 when constructing the mapping potential. The reaction coordinate can then be defined directly in terms of the VB states (or rather the energy gap between these states) and there is no need to use any particular constrained geometric variable as the reaction coordinate. This is more of a problem if we want to use an MO description of the reaction in connection with the FEP approach. In principle, one would then like to define the reactant and product states in terms of the corresponding Hamiltonian eigenvectors and orbital coefficients. These are, however, not really known beforehand and will also fluctuate rather unpredictably during an MD simulation. In applications that combine MO descriptions with FEP simulations it has therefore been necessary to assume some reaction coordinate in terms of geometric variables (and sometimes also charge distribution) and then constrain this "coordinate" to different values along the reaction path.[7c,17,55,80] This solution is not always satisfactory since it involves constraining of part of the quantum system and since it may often be difficult to make a good guess at a proper reaction coordinate. The possibility of using mixed MO/EVB schemes has also been suggested[25,56a] and might be rather useful. In such a case one could evaluate the relevant potential energy surface by MO methods (e.g., using the adiabatic mapping procedure[30]) and then fit this surface to an EVB model which would then be used for the FEP calculations. Another promising strategy is provided by using the EVB surface as a reference state for MO calculations.[21]

6. An Illustrative Example of EVB Simulations: Carbonic Anhydrase

In order to clarify the actual procedure that is employed in modeling an enzyme reaction with EVB approach we will discuss some recent work on carbonic anhydrase.[56] This enzyme provides a suitable test case for several reasons. First, high-resolution crystal structures of various native and inhibited forms of the enzyme are available[57] and, second, extensive kinetic data on the reaction (both the catalyzed and uncatalyzed) has been gathered during the years.[5,58] As discussed in section 3, a variety of different theoretical works have been addressed at the catalytic mechanism of CAs[6-8] and it is therefore interesting to compare what we can learn from different theoretical approaches.

Carbonic anhydrases catalyze the reversible hydration of carbon dioxide:

$$E\text{–}H_2O \rightleftharpoons H^+\text{–}E\text{–}OH^- \overset{B}{\underset{BH^+}{\rightleftharpoons}} E\text{–}OH^- \quad (24)$$

$$E\text{–}OH^- + CO_2 \rightleftharpoons E\text{–}HCO_3^- \overset{H_2O}{\rightleftharpoons} E\text{–}H_2O + HCO_3^- \quad (25)$$

Here, eq 24 represents the proteolysis of the reacting active-site water molecule and subsequent transfer of a proton from the active site to the surrounding solution, while eq 25 denotes the "half-reaction" corresponding to the interconversion between CO_2 and HCO_3^-. The former half-reaction has been found to be rate limiting both in isozymes I and II,[5,58b] although it is not known exactly where on the proton transfer pathway (eq 24) the rate-limiting barrier is located. In isozyme II it is rather well established that the imidazole moiety of His64 acts as an intermediate proton acceptor/donor during the translocation process, while the exact pathway in isozyme I is still uncertain.

The active site Zn^{2+} ion is thought to play a very important role in all the CAs. One of its key functions is probably to lower the pK_a of the reacting water molecule from its normal value of 15.7 (in bulk water) to about 7–8 in the enzyme's active site.[58c]

6.1. The Initial PT Step

Let us start by considering the first reaction step of eq 24, namely the initial proteolysis of water at the

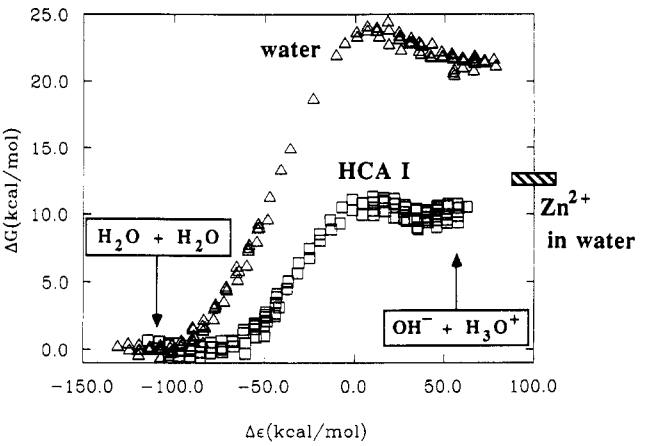


Figure 2. Calculated free-energy profiles for the reference reaction ($2H_2O \rightleftharpoons OH^- + H_3O^+$) in water after calibration of the $\Delta\alpha$ and H_{12} parameters (upper curve, open triangles) and for the proton-transfer step in HCAI (lower curve, open squares). The calculated energetics of proton transfer from a zinc-bound water molecule in aqueous solution[60b] is also shown for comparison. The reaction coordinate $\Delta\epsilon$ denotes the energy gap between the two diabatic surfaces.[2]

catalytic zinc site. The proton transfer (PT) reaction between two water molecules can be described as an effective two-state problem corresponding to the valence bond (VB) structures:

$$\psi_1 = (HOH)\ (OH_2)$$

$$\psi_2 = (HO^-)\ (HOH_2^+) \qquad (26)$$

The reaction can be described by more states but these higher energy configurations can be incorporated into a two-state model.[46a] The calculations of ref 56 therefore used a model for this PT step that is equivalent to the description in section 4.2.1 above. The EVB parameters $\Delta\alpha^0 = \alpha_1^0 - \alpha_2^0$ and A_{12} and μ_{12} which determine the function $H_{12}(r)$ were fitted to experimental information[59] about the uncatalyzed reference reaction in water. This is described in ref 60 where the relevant parameters can be found and the procedure for their calibration has been explained in section 4.2.2.

The water simulations[60] were carried out with the reactants immersed in a sphere of SPC water,[61] the boundary of which was subjected to radial and polarization restraints according to the SCAAS model.[62] In the subsequent simulations of the enzyme (HCAI) reaction a spherical protein/water system of radius 14 Å (centered on the Zn^{2+} ion) was treated. Further details concerning these simulations can be found in ref 56.

The corresponding EVB reaction free-energy profile in water is shown as the upper curve in Figure 2. The free-energy curve corresponds to the ground-state free energy, $\Delta g(X)$, calculated according to eq 23. This reaction profile reproduces the experimentally observed values of the activation energy and the reaction free energy, $\Delta G^{\ddagger}_{obs} = 23.8$ kcal/mol and $\Delta G^{\circ}_{obs} = 21.4$ kcal/mol.[59] The range of the abscissa, $\Delta\epsilon$, corresponds at its endpoints to a hydrogen bonded water pair and a contact ion pair, respectively. In the enzyme, the proton will subsequently be translocated from the active site to the surrounding medium,[5] and this process will be considered below.

The result of the "thermodynamic cycle" associated with moving the reacting fragments into the active site of HCA I is shown by the lower free-energy curve of Figure 2 (open squares) and a snapshot of the active site at a high energy "transition structure" where the proton is about halfway transferred between the two water molecules is shown in Figure 3. The values obtained for $\Delta G^{\ddagger}$ and ΔG° in the enzyme are 10.8 kcal/mol and 10.0 kcal/mol, respectively (with FEP hysteresis errors of ±1.0 kcal/mol). The calculated value of the reaction free-energy corresponds to 7.4 ± 0.7 pK units. This number is very close to the experimentally measured pK_a (≃ 7.5) in HCA I[58b,e] that has been ascribed to the ionization of the zinc-bound water. However, it should be noted that the experimentally measured pK_a value does not exactly correspond to the process described by the simulations, since in the former case the proton (or H_3O^+ ion) is transferred from the enzyme out into solution. Thus, in order to be able to make a direct comparison with the observed pK_a one must also consider the energetics of transferring the proton from the enzyme's active site into solution. This issue was addressed by also calculating the free energy associated with transforming a hydronium ion into a water molecule (viz. $H_3O^+ \leftrightarrow H_2O$) in the active site and in aqueous solution. (For related thermodynamic cycles in pK_a calculations, see ref 63.) This type of calculation is associated with some difficulties due to the fact that it involves annihilation/creation of a net charge inside the protein and the corresponding errors are larger than for the reaction profiles in Figure 2.[56a] However, the calculated binding energy for H_3O^+ was found to be close to zero,[56a] wherefore the calculated value of the reaction free energy should be a reasonable estimate of the pK_a associated with zinc-bound water ionization.

As for the free-energy barrier associated with the initial water ionization step, eq 17 with a transmission factor of unity would imply a rate for this reaction step of $k \simeq 7 \times 10^4$ s^{-1}, using our calculated value of $\Delta G^{\ddagger}$. Here, it may be appropriate to point out that the above treatment implicitly assumes that the contribution from quantum motion (zero-point energy and/or tunneling) to the *difference* between the enzyme and solution reactions is negligible. In other words, the quantum corrections to the classical rate constant (eq 17) for the catalyzed and uncatalyzed reactions are considered to be equal and thus "hidden" in the calibration of the latter free-energy profile. Actual calculations of these effects will, however, be addressed below.

The above estimate of the rate constant for the PT step should probably not be directly equated to the catalytic rate, since the exact "location" of the limiting barrier along the proton translocation pathway is not known. That is, the proton is being transferred from the zinc center in the active site some 10 Å out into the solution and what is known is that this *entire process* is rate limiting.[5] Exactly where the proton is at the highest point on the free energy curve is thus not known. We can only note that the overall process corresponding to the ionization of the zinc-bound water and subsequent translocation of the proton to solution (i.e. including possible intramolecular transfer steps) is known to be rate limiting. If the transfer of the proton to solution proceeds via a histidine residue, as has been

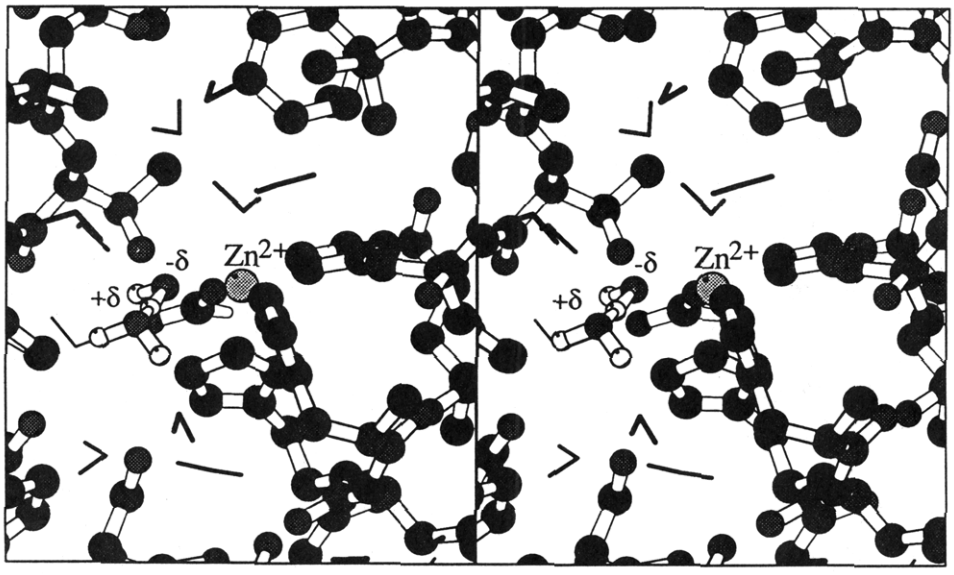


Figure 3. Stereo snapshot from the simulations of a high-energy transition structure at which the proton is about halfway transferred. The zinc ion is depicted in light gray color and the hydrogens on the two reacting water molecules in white.

suggested (see, for example, ref 58b), this second step would correspond to a negative $\Delta G°$ as long as the acceptor (histidine) pK_a is higher than that of the donor (H_3O^+). One would therefore expect such a second PT step to involve only a small barrier. It thus seems reasonable to expect that the rate of the initial PT step, obtained from the simulations, is fairly close to the observed rate of catalysis. Encouragingly, this is also what the calculated free energy profile suggests and it gives, in fact, a rate in rather close agreement with the experimental value of $k_{cat}^{obs} = 6.8 \times 10^4\ s^{-1}$ for HCAI at pH = 7.1.[58f]

While the above discussion of the PT step is instructive since it can be related to clear experimental findings, it is important to keep the perspective of the overall catalytic mechanism. A schematic free-energy diagram of the two major reaction steps, i.e. excluding the binding and release of substrate and product is shown in Figure 4. The figure depicts on its back panel the energetics of the simple sequential mechanism where the OH^- attack on carbon dioxide takes place while the H_3O^+ ion has not yet dissociated from the active site. If this were the actual mechanism in the enzyme one would presumably observe a $\Delta G_{cat}^{\ddagger}$ which is much larger than the current estimate. The value of $\Delta G^{\ddagger}$ for the $OH^- + CO_2 \rightleftharpoons HCO_3^-$ reaction in water is about 12 kcal/mol, as deduced from the corresponding rate constant,[64] and one would expect this $\Delta G^{\ddagger}$ in the protein to be at least, say, 5 kcal/mol. The free-energy barrier of the interconversion step would thus add on to the $\Delta G°$ of the first step and consequently determine the rate (unless the second barrier is extremely small). However, the reaction in the protein probably follows a different pathway in which the proton of the initially formed H_3O^+ ion is transferred to an accepting base (which is ultimately a solvent molecule) before the hydration step takes place (point 3 in Figure 4). This lowers the ΔG for the formation of the OH^- ion to about zero at pH = 7. Note, that this is not at variance with the previous result that the cost of moving the H_3O^+ ion from the active site to solution is close to zero, since in this case the calculations referred to standard-state free energies without any pH effect (see, for example, ref 63b). Hence, the nucleophilic attack would not be rate determining provided that its barrier is lower than that of the initial PT, as can be seen from Figure 4, and the overall rate would then be limited by either $\Delta G_{1\rightarrow2}^{\ddagger}$ or $\Delta G_{2\rightarrow3}^{\ddagger}$. Experiments also show that it is indeed this intramolecular proton transfer which limits the rate of the enzyme.[5]

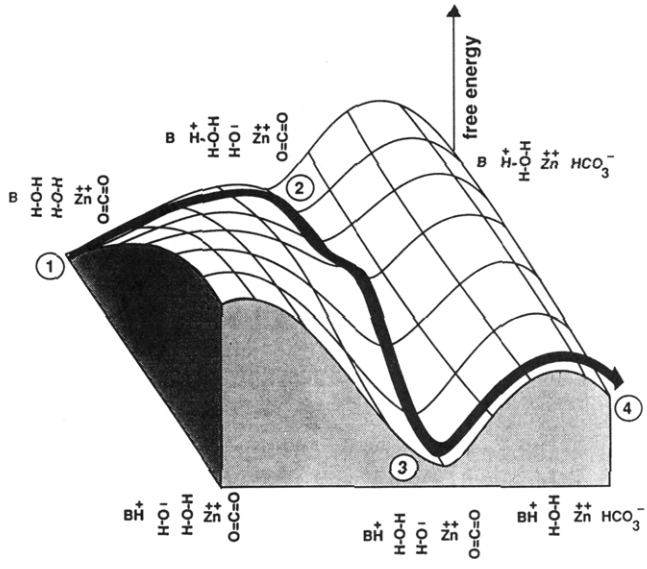


Figure 4. Schematic free-energy diagram for the water *proteolysis and* CO_2/HCO_3^- *interconversion steps* in HCAI at pH = 7. B denotes the final proton acceptor which is a solvent or buffer molecule.

6.2. The Nucleophilic Attack

Let us now turn to the actual chemical conversion step of this enzyme. As mentioned in section 3, several *earlier theoretical studies have examined the* CO_2/HCO_3^- interconversion process in vacuum for simple model systems with ab initio and semiempirical MO methods.[6-8] Solvation effects on an ab initio energy profile have also recently been reported.[7c,8g] Merz has *evaluated the energetics of* CO_2 *binding and the* pK_a *of* Glu106 in HCAII by FEP methods[65] and Liang and

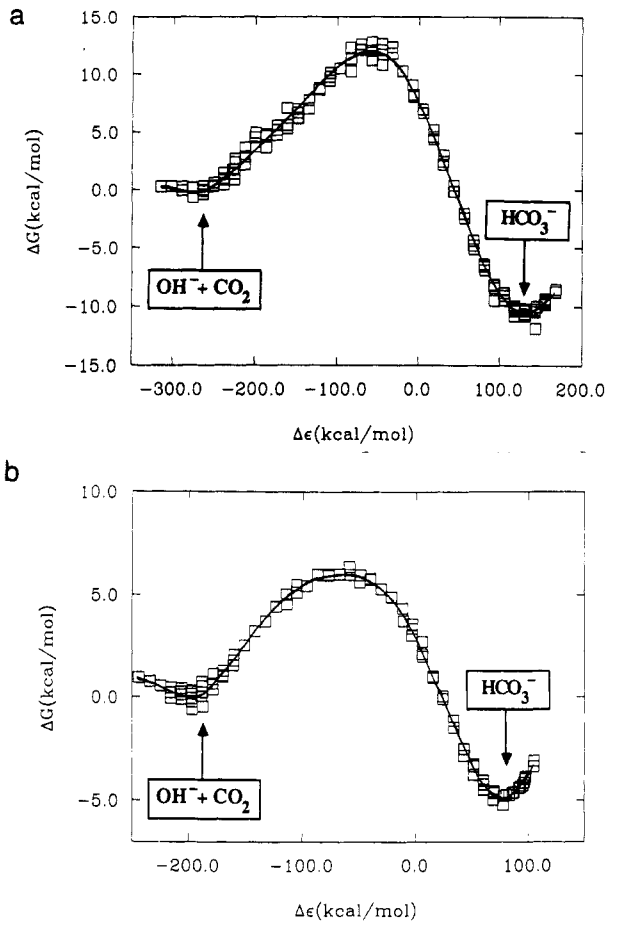


Figure 5. (a) Calculated free-energy profile for the reference reaction $OH^- + CO_2 \rightleftharpoons HCO_3^-$ in water after calibration of the $\Delta\alpha$ and H_{12} parameters and (b) calculated free-energy profile for the interconversion reaction (eq 25) in the active site of HCAI.

Lipscomb have also studied CO_2 binding and diffusion inside the enzyme with MD simulations.[66]

The interconversion reaction can be described by the two VB structures

$$\psi_1 = (HO^-)\,(CO_2)$$

$$\psi_2 = (HOCO_2^-) \qquad (27)$$

which obviously are the most important resonance forms for the reactant and product states, respectively. The experimental data on CO_2 hydration in aqueous solution from ref 64 was used as before to calibrate the $\Delta\alpha$ and H_{12} parameters of the EVB Hamiltonian by simulating this reference reaction in water.[56b] The off-diagonal matrix elements are again approximated by a simple exponential function, $H_{12} = A_{12}e^{-\mu_{12}r_{OC}}$, where r_{OC} denotes the separation between the atoms forming the new bond.

The forward and reverse rates for the uncatalyzed attack of OH^- on CO_2 in water are given in ref 64a (see also ref 64b) and can be translated into values of $\Delta G°$ = −10.5 kcal/mol and $\Delta G^‡$ = +11.9 kcal/mol for the reaction and activation free energies, respectively. Figure 5a shows the calculated free-energy profile for the *solution* reaction after calibrating the EVB Hamiltonian with respect to the $\Delta\alpha$ and H_{12} parameters ($\Delta\alpha$ = −15.8 kcal/mol, A_{12} = 107.5 kcal/mol, μ_{12} = 0.40 Å^{-1}). One can note in the case where $|H_{12}| \rightarrow 0$ as $|\Delta\epsilon| \rightarrow \infty$, $\Delta\alpha$ is determined by the thermodynamic cycle

$$\begin{array}{ccc} R^\infty(g) & \xrightarrow{\Delta\alpha} & P^\infty(g) \\ \Sigma\Delta G_{sol}(R)\downarrow & & \downarrow\Sigma\Delta G_{sol}(P) \\ R^\infty(aq) & \xrightarrow[\Delta G^{aq}_{rxn}]{} & P^\infty(aq) \end{array}$$

where R^∞ and P^∞ denote the reactant and product states with the relevant fragments noninteracting. The way we have defined the H_{12} function above, however, will add an extra (constant) term to the energy at the product. This term can be evaluated from a series expansion of the ground-state energy (see ref 56b) giving an extra contribution ΔE_{offd} = −23.3 kcal/mol that should be added to the gas-phase shift, $\Delta\alpha$, in order to be consistent with the thermodynamic cycle above ($\Delta\alpha_{eff} = \Delta\alpha + \Delta E_{offd}$ = −39.1 kcal/mol). It is thus possible to get a back-of-the-envelope check of the consistency of the computational reaction model from

$$\Delta G^{aq}_{rxn} - \Delta\alpha_{eff} = \Delta\Delta G^{R\rightarrow P}_{sol} \qquad (28)$$

Insertion of the relevant numbers yields $\Delta\Delta G^{R\rightarrow P}_{sol}$ = +30.4 kcal/mol which is in good agreement with experimental estimates for the difference in solvation free energies.[67] Thus, this type of check shows that the model for the solution reaction is physically consistent with experimental data.

It is also interesting to note that, in vacuum, the interconversion reaction is barrierless and exothermic by 47–49 kcal/mol.[7a,c] The difference in solvation energy between OH^- and HCO_3^- and the delocalization of charge along the reaction pathway, which is associated with a solvent reorganization barrier, thus produces a reaction profile in solution that is very different from the gas phase.[7c,56a] As is evident from eq 28, the quantity $\Delta\alpha$ is the gas-phase reaction free energy which can be obtained from heats of formation, e.g., if the ΔS contribution can be estimated. The value of $\Delta\alpha_{eff}$ = −39.1 kcal/mol obtained in ref 56b can be compared to the experimental estimate of −38.2 kcal/mol given in ref 7c. Hence, it is clear in this case that the calibration of the EVB Hamiltonian against *gas-phase* experimental data would also have resulted in a very accurate value for ΔG_{rxn} in water.

After calibration of the EVB surface for the solution reaction calculations of the same reaction catalyzed by HCAI were carried out, without any reparametrization of the EVB Hamiltonian. The resulting free-energy profile for the enzyme catalyzed reaction is shown in Figure 5b. Two important features of the enzyme profile can be immediately appreciated from the figure. First, the activation barrier for the nucleophilic attack is reduced by the enzyme (by 5–6 kcal/mol) and, second, the relative stability of the states ψ_1 and ψ_2 is shifted such that this reaction step becomes less exothermic than in water. The calculated values for $\Delta G^‡$ and $\Delta G°$ are 6.3 kcal/mol and −4.8 kcal/mol, respectively, with convergence errors of ±1.6 kcal/mol. Here, it should be noted that the lowering of the transition state for the interconversion process is much less than for the preceding water ionization step. That is, the rate enhancement (compared to water) of the PT step was calculated to be about a factor of 10^8 while the rate acceleration for the OH^- attack on CO_2 is only

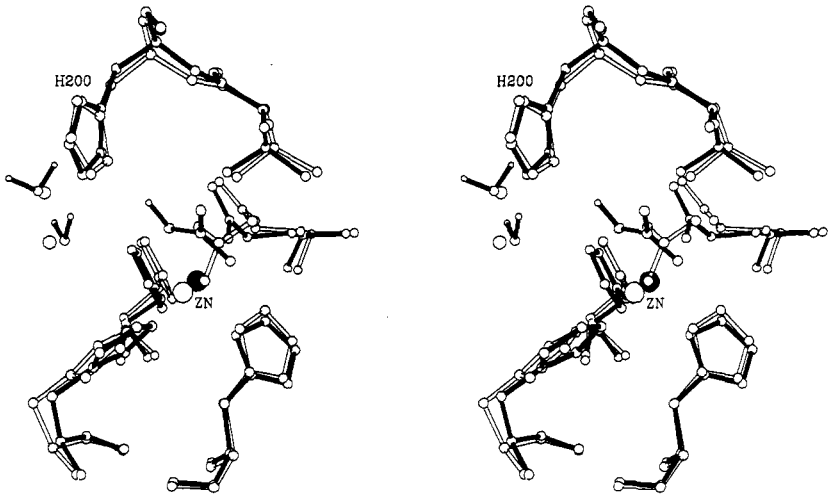


Figure 6. Comparison of the average MD structure of the HCAI–bicarbonate complex with that of the same complex in the T200H mutant of HCAII.[69] The simulated structure is depicted with black bonds and the Zn^{2+} ion is in this case also drawn in black.

about 10^3–10^4. The latter reaction is, however, already quite rapid in solution[64] and the enzyme does probably not need to put a larger effort into reducing the interconversion barrier since a reduction of a few kilocalories per mole is enough to not make it rate-limiting overall.

The simulations also show that the "reactants" (ψ_1) in this case are stabilized relative to the product (ψ_2). This reflects the strong interaction between the zinc and OH^- ions that is also essential for making the OH^- nucleophile available at a low energetic cost. Moreover, it is important from the functional viewpoint that the enzyme is able to "level out" the free-energy difference between ψ_1 and ψ_2 somewhat, since there are requirements for it to operate also in the reverse direction. The calculated free-energy profile for the CO_2/HCO_3^- interconversion step appears to be in reasonable agreement with the experimental data of Behravan et al.[58c] These authors have estimated the forward and reverse rate constants for the nucleophilic attack in HCAI as 3.4×10^7 s^{-1} and 3.8×10^4 s^{-1}, respectively, yielding values of $\Delta G° = -4.1$ kcal/mol and $\Delta G^‡ = 7.1$ kcal/mol.

Just as for the water ionization step, it seems clear that the zinc ion plays a major role also in controlling the energetics of the interconversion step. One reason for the stabilization of ψ_1 and the subsequent transition state relative to the product is presumably the stronger interaction between Zn^{2+} and OH^- than the corresponding one with HCO_3^-. Moreover, when the product is formed it will inevitably find some of its negative charge located in a relatively hydrophobic environment, although the region is not completely devoid of dipolar groups. (In particular, the backbone –NH group of Thr199 interacts with the carboxylate moiety of the product.) It thus appears that the active-site microenvironment is well suited to actually *destabilize* the product (compared to the solution reaction) somewhat, thereby enabling the reversibility of the interconversion process. Furthermore, it can be noted that the *reorganization energies* relevant to both the PT as well as the interconversion step are reduced by the enzyme compared to the solution reaction (cf. Figures 2 and 5). This indicates that the change in polarization of the active site, in response to the changing charge distribution during the reaction, is smaller than the corresponding one in water. (This type of effect has also been observed earlier in calculations of the lactate dehydrogenase catalyzed reaction.[68]) Here again, the zinc ion is likely to be important together with other polar groups of the active site. Xue et al.[69] have recently determined the crystal structure of the Thr200→His mutant of HCAII in complex with the substrate/product HCO_3^-. This mutant thus mimicks the structure of HCAI which has a histidine residue in position 200. It has also been found that the catalytic parameters of the T200H mutant have changed in the direction of HCAI.[58c] In particular, the affinity of T200H and of HCAI for bicarbonate is about 1.5–1.8 kcal/mol more favorable than for native HCAII.[58c] Figure 6 shows a comparison between the crystal structure of Xue et al.[69] and the average MD structure at the product state (ψ_2) with HCO_3^- bound. It can be seen that the position of the bicarbonate molecule is rather similar in the two cases, the main difference being a slight rotation about the C–OH bond. (This rotation determines whether the hydroxyl hydrogen of the product is pointing toward His200 ND1 or Thr199 OG1, both of which appear to provide good hydrogen bond acceptors.) Furthermore, in both cases is the hydroxyl group still ligated to the Zn^{2+} ion and no rotation of the bicarbonate molecule, which would bring the two carboxylate oxygens into the metal coordination sphere, has occurred. Interactions with the peptide–NH group of Thr199 and the side chain of His200 are evident and the latter of these can probably explain the increased affinity for HCO_3^- that is observed when the amino acid in position 200 is a histidine.[58c] Note also the excellent agreement between calculated and observed positions for the two water molecules near His200.

Another point, which is interesting in the context of methodology here, is the comparison of the transition state (TS) structure in the enzyme obtained by the EVB method to the corresponding ones from gas-phase ab initio calculations. Tapia and co-workers have located the saddle point for the interconversion step in vacuum in the presence of bare[7b] as well as ammonia-ligated Zn^{2+}.[8f] Optimized TS structures with NH_3-ligated zinc have also been obtained recently using different basis

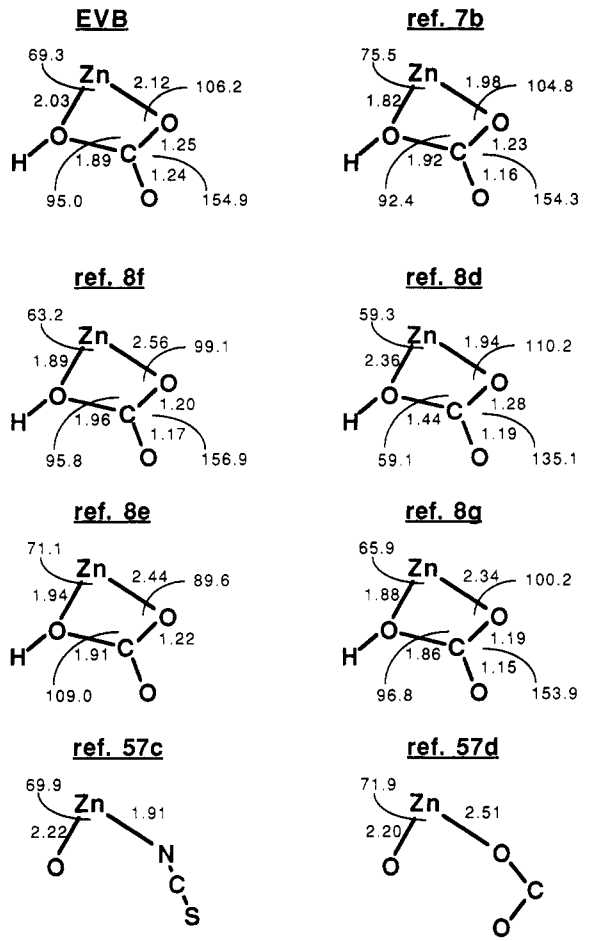


Figure 7. Comparison of the average transition-state geometry from the EVB simulations with those from the ab initio SCF calculations of Tapia et al.,[7b,8f] Solà et al.,[8d] Krauss and Garmer,[8e] and Zheng and Merz.[8g] Three ammonia molecules were used to model the Zn^{2+} ligands in ref 8. The figure also shows the observed coordination geometries of the inhibitors SCN^- and $HCOO^-$ bound to HCAII.[57c,d]

sets by Solà et al.,[8d] by Krauss and Garmer,[8c] and by Zheng and Merz.[8g] These five ab initio structures are shown in Figure 7 together with the time-averaged EVB TS structure. Here, a note of caution should perhaps be given concerning the ammonia-ligated ab initio structures. If the product (HCO_3^-) conformation is artificially stabilized by the interaction with an ammonia molecule[8c,g] this may make the TS look "earlier", i.e. shorter Zn–OH distance and longer Zn–CO_2 distance, than would be the case without such an interaction (the imidazole ligands in CA do not have the same capability of hydrogen bonding to the substrate). Although the TS geometries for these vacuum systems clearly depend on basis sets and the model used for the metal ligands, the comparison in Figure 7 demonstrates that the general features of the TS sampled in the EVB simulations are similar to those obtained from ab initio studies on model systems. Also shown in Figure 7 are the coordination geometries of two experimentally determined inhibitor structures of HCAII, namely the complexes with SCN^- and $HCOO^-$.[57c,d] It is, of course, difficult to say to what extent these inhibitors mimick the actual TS of the interconversion reaction in the enzyme, but their structures seem to be compatible with the overall picture emerging from the theoretical results of Figure 7.

7. Have Theoretical Calculations Contributed to Our Understanding of Enzyme Mechanisms?

In the previous sections we have tried to outline the most important theoretical approaches to the modeling of enzymic reactions. In doing so we have also emphasized the type of treatments that take into account the combined "solvent effect" from the mixed protein/water environment. That this route seems most promising for obtaining quantitative results is nowadays probably agreed upon by most workers in the field. Although it may be too early to pass a definite judgement on the success of hybrid QM/MM models, it seems appropriate to ask what the applications of the above methods have actually contributed to our understanding of enzyme reactions.

The early work of Warshel and Levitt on the catalytic reaction of lysozyme[24] demonstrated that ground-state destabilization via steric strain imposed on the substrate by the protein is unlikely to be a major source of catalysis, at least in this specific case. The rationalization of this observation is that the enzyme structure itself is not rigid enough to prevent relaxation of such a strained conformation, wherefore most of the strain energy would dissipate into the surrounding protein. Their calculations instead suggested that electrostatic stabilization of the carbonium ion intermediate is a much more important factor for the catalytic rate enhancement of lysozyme. Whether this is a general conclusion is not yet clear but considerable evidence seems to be accumulating in favor of electrostatic interactions as being the most important ones for enzymatic rate enhancement.[2]

Similar conclusions concerning the key role of electrostatic interactions, between groups of the enzyme and ionic forms of the reacting species along the catalytic pathway, have been reached in several studies of the serine proteases.[2,9,27,32,74] Computer simulations of site-directed mutagenesis experiments in which constituents of the so called "oxyanion hole" are deleted have also been carried out.[70,71] These studies reproduce the observed free-energy changes rather well and illustrate how the oxyanion hole, by its hydrogen bonds to the tetrahedral intermediate, is essential for the catalytic effect. A widely debated issue in the context of serine proteases has been whether the "charge-relay" or "double proton transfer" mechanism[72] is actually operational during catalysis. Early semiempirical calculations on model systems[9a,c,g] had indicated that this was the case, while subsequent studies of varying degrees of sophistication[73] as well as simple pK_a arguments suggested the contrary. More recent theoretical studies both by Warshel et al.[74a] and by Kollman and co-workers[74b] have disproved this mechanism rather convincingly, although the hypothesis might still find some supporters.

EVB simulations of the cleavage of nucleic acids by the enzyme staphylococcal nuclease (SNase) have been reported by us.[75] This work showed that the large catalytic effect associated with the active-site Ca^{2+} ion can also be understood as a rather straightforward electrostatic phenomenon. These simulations detected the movement of one active-site arginine residue during the course of the hydrolytic reaction which results in a stronger hydrogen-bonding interaction with the rate-limiting transition state. Such a structural change had

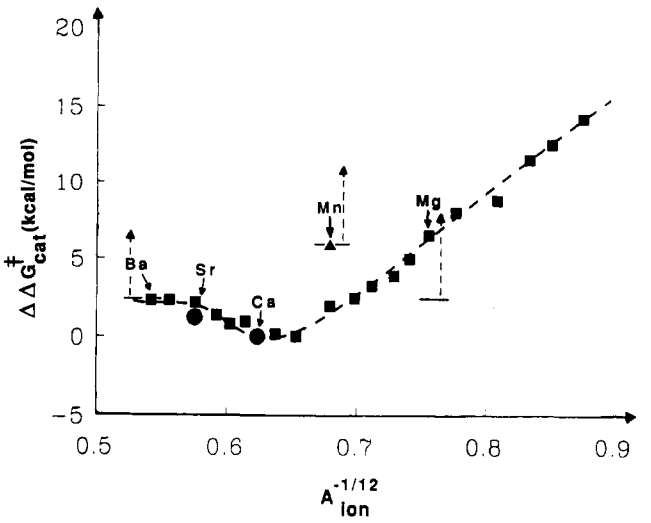


Figure 8. Calculated effect of metal ion substitutions on the overall activation barrier of SNase relative to the case with Ca^{2+} bound. The dashed curve represents the relationship obtained for the alkaline-earth metal ion series. Mn^{2+} was described by the octahedrally delocalized charge model of ref 77. The observed values for Sr^{2+} and Ca^{2+} are denoted by circles and experimentally estimated limits for Ba^{2+}, Mg^{2+}, and Mn^{2+} by upward dashed arrows. The scale of the abscissa is a measure of the effective ion radius. Note that the figure predicts the enzyme to be more sensitive to substitution by smaller than by larger divalent cations.

been conjectured earlier on the basis of kinetic and binding data[76] and was thus confirmed by the calculations. A subsequent FEP study[77] of metal ion replacement in the active site of SNase showed how the enzyme was optimized to work exactly with Ca^{2+} as its catalytic ion (Figure 8). The free-energy relationships underlying such optimization phenomena were also elicited. Moreover, a classification scheme for hydrolytic and related enzymes based on free-energy considerations was presented which rationalizes the choices of some catalytic groups that are commonly used (Figure 9).

Several theoretical works have also been published that address the catalytic reaction of triose phosphate isomerase (TIM).[78,35b] The early ab initio calculations by Alagona et al.[78] on relatively simple model systems as well as the recent work by Bash et al.[35b] that includes the protein/solvent environment both indicate that the enzyme excerts a major stabilizing effect on the ionic enediolate forms of the substrate. The latter study also provides some evidence in favor of neutral His95 as being the donor for enediolate protonation. However, the potential energy profile obtained in ref 35b depicts (doubly protonated) enediol and neutral His95 (after having protonated the enediolate) as the most stable configuration along the reaction pathway. Since this is not observed experimentally,[79] one would expect the cost of protonating His95 and/or binding the substrate in presence of the protonated histidine to prevent this "thermodynamic trap".

It should perhaps be mentioned that, so far, almost all calculations of *free-energy* surfaces for enzyme reactions have been done using the EVB model, the only exceptions being the work by Singh on PT in dihydrofolate reductase[80] and by Zheng and Merz on HCAII.[8g] Several studies of relative free energies of inhibitor binding to enzyme active sites using FEP methods have, however, been reported.[81-85] These enzymes include thermolysin,[81] trypsin,[82] DHFR,[83] α-lytic protease,[84] and HCAII.[85]

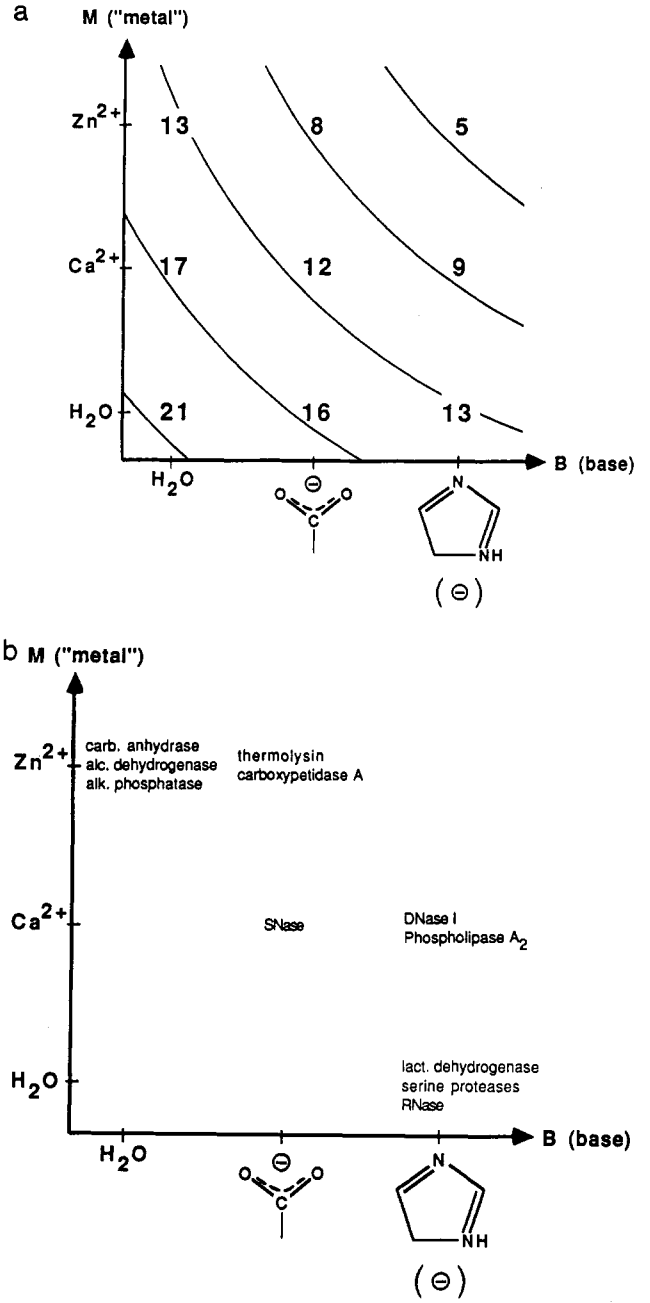


Figure 9. (a) Diagram showing the effects of metal ion and general base catalysts on the ionization of an R–OH group. The abscissa denotes increasing general base strength, represented by a water molecule, a carboxylate ion, and an imidazole ring. The ordinata represents increasing metal ion electrophilicity, where a water molecule denotes the case where no metal is present. The energy values, ΔG_{ij}, correspond to the proton transfer reaction $(M_i^{2+})H_2O + B_j \rightleftharpoons (M_i^{2+})OH + BH_j^+$, where each entry is obtained from $\Delta G_{ij} = -2.3RT(pK_j - pK_i)$. For example, in the case of a metalloenzyme using Ca^{2+} and glutamate, respectively, as M and B, i denotes Ca^{2+}-(H_2O) and j denotes (Glu)–COOH (the pK_a's for metal bound water are taken from ref 107). Part b shows a number of different enzymes, that catalyze reactions involving a proton-transfer step of the above type, plotted according to their use of metal ion and general-base catalysis (SNase, DNaseI, and RNase denote staphylococcal nuclease, deoxyribonuclease I, and ribonucleases [e.g., A and T_1], respectively). It is interesting to note that the "high-energy" region of part a seems to be avoided by most enzymes.

As far as CA anhydrase is concerned, the wealth of theoretical studies have certainly provided interesting information about different mechanistic options,[7,8,56,65,66] although the results have sometimes been conflicting. However, the combination of FEP simulations either with the EVB approach, as detailed in section 6, or with docking of ab initio structures into the active site as described by Zheng and Merz,[8g] allows for a direct comparison with experimental energetics. In this respect, we find the simulations summarized in section 6 very encouraging since they essentially yield a reasonable energetic picture of the catalytic reaction. The often-held view that the key role of the enzyme is to provide reactive hydroxide ions[86] for attacking CO_2 is quantitatively demonstrated by the simulations.[56] Furthermore, it is evident from the calculations that the enzyme *destabilizes* the bicarbonate form relative to OH^- and CO_2, thereby enabling reversibility of the reaction. This is achieved by a stronger "binding" of OH^- to the active site than HCO_3^-, where the catalytic zinc ion plays a major role and probably also the hydrophobic nature of the pocket in which the carboxylate moiety finds itself when formed.

Other interesting insights have been provided in a recent study of the catalytic reaction of lactate dehydrogenase.[68] There it was demonstrated that the reorganization energy, which determines the activation barrier when $\Delta G^\circ \simeq 0$ (see section 8.2), is significantly smaller in the enzyme than in the corresponding solution reaction. It would thus appear that one way for an enzyme to catalyze a given reaction is to minimize the reorganization of the polar groups in the active site.[91] This is best accomplished when the protein uses its folding energy to preorient dipolar groups toward the relevant transition-state configuration.[91,2]

Although most enzymes are not limited by the rate of collision with their substrates, but by chemical conversion steps, there are examples of the former type.[87a] The encounter between enzyme and substrate for the diffusion-limited reaction of superoxide dismutase has been studied by McCammon and coworkers[87b] and by Sharp et al.[87c] using a Brownian dynamics simulation method. This work demonstrated that electrostatic steering of the substrate can indeed increase the normal diffusion limit for the rate of encounter, in agreement with experimental observations.[87a]

The few examples that we have discussed in this section are some of the most relevant ones in our view and in our opinion they do motivate a rather positive answer to the question posed in the section's title.

8. New Challenges for Theoretical Methods

There are of course a multitude of questions associated with the actions of enzymes that are still being widely debated. Some of the most interesting discussions regard the relative importance of contributions from various sources to the catalytic effect.[2,3,9g,24,88] Among the effects that have been suggested to be of major relevance we find, e.g., entropy,[89] proximity,[90] electrostatic stabilization,[91] desolvation,[9g] orbital steering and stereoelectronic effects,[92] etc. It is obviously not within the scope of this review to discuss all these proposals in detail. However, we would like to suggest that theoretical simulation approaches can, in fact, be used to explore various hypotheses about catalytic processes that are often discussed in rather vague or abstract terms. Some recent examples of this have been reported by us,[95] and they deal with the concept of linear free-energy relationships in enzymes and the effect of "quantum motions", viz. zero-point energy and nuclear tunneling effects, in catalytic reactions. We will briefly describe how the EVB method has been employed in these studies in sections 8.2 and 8.3. The proposal that enzymes provide a gas-phase like environment around the reactants and thereby enhance the reaction rate[9g] has been critically examined in an earlier paper.[88] There it was demonstrated, in the case of peptide hydrolysis by trypsin, that this type of desolvation hypothesis did not seem to be valid. Another interesting proposal in this context was recently put forward by Menger and termed the "split-site model".[93] He suggests that binding energy at a site distant from where the reaction takes place can be used for ground-state destabilization in terms of steric strain and/or desolvation of charged groups. (See also discussion in the appendix of ref 90.) The relevance of this type of mechanism in catalysis certainly deserves further study and microscopic simulations could probably make significant contributions here. Another important issue regards the extent of entropic contributions to the lowering of the activation energy in enzymes. This question could now be addressed by computer simulations, and we will discuss it briefly in the next section.

8.1. Entropic Contributions to Catalysis

Although most theoretical and many experimental studies have indicated that electrostatic effects are the most important factor in enzyme catalysis (see discussions in ref 2), entropic effects may also play a significant role.[3,89,94] Unfortunately, it is extremely hard to experimentally determine the actual magnitude of entropic contributions in enzyme reactions. For example, if one considers certain intramolecular reactions, such as ring closure, where two reacting groups are covalently linked to each other with varying degrees of rigidity there are clearly large entropic effects manifested in the corresponding rate constants.[94] However, the direct relevance of entropic factors in such systems to enzyme catalysis is very difficult to assess (except for enzymes that actually catalyze ring closure). For many enzyme-catalyzed intermolecular reactions one might argue that the same steric effect that reduces the available conformational space in the ground state also will operate at the transition state (see the analysis in Chapter 9 of ref 2).

One way to examine the importance of the entropic factors is to try to analyze the configurational space available to the system in its ground and transition states, both in the enzyme and in solution. That is, provided that we have a force field that reliably can describe the system, the (generalized) volumes of the conformational space defined in terms of some appropriate reaction coordinates can be estimated for a given temperature. The entropic contribution to the catalytic effect, measured relative to the uncatalyzed solution reaction, can then be expressed as[2]

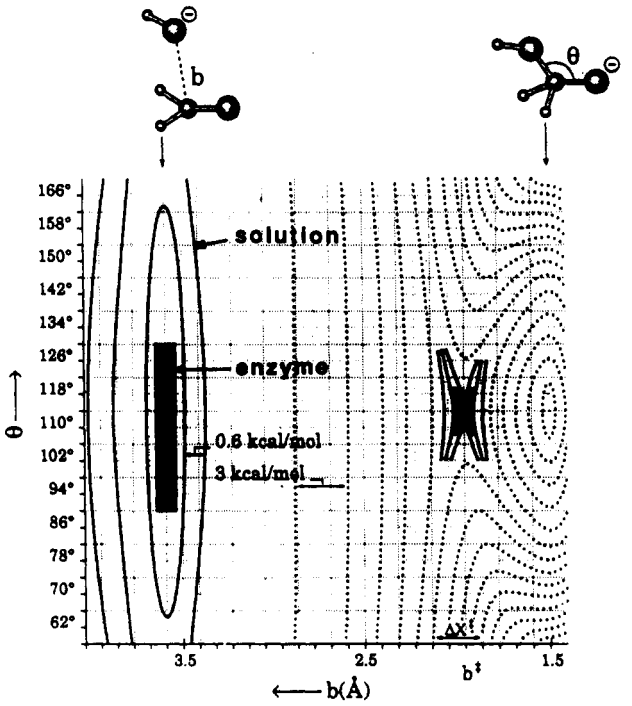


Figure 10. Energy diagram showing the potential surface for nucleophilic attack by R–OH⁻ on a carbonyl group in amide hydrolysis. The reaction is described by the bond distance, b, and the attack angle, θ. The heavy contour lines are spaced by kT at room temperature (~0.6 kcal/mol) and can be used to estimate entropic effects. The figure also shows the regions (in dark) where the potential is less than kT for the corresponding reaction in the active site of subtilisin.

$$-T\Delta\Delta S \simeq -\beta^{-1}\ln\left(\frac{v_p^*/v_p^0}{v_s^*/v_s^0}\right) \tag{29}$$

where v^* and v^0 denotes the configurational volumes in the transition and ground states, respectively, while p and s denote protein and solution, respectively. One extreme case is the orbital steering model of Storm and Koshland[92a] that considered v_s^* and v_p^* to be very small and implicitly assumes that $v_p^0 \simeq v_s^*$ (see ref 2). This model can now be conveniently examined using EVB potential surfaces (see Chapter 9 of ref 2) where experimentally known force constants of the reactant and product states provide an accurate upper limit for the transition-state force constants and the corresponding v_s^*. As is shown in exercise 9.3 of ref 2, v_s^* is not extremely small and even the unrealistic case where $v_p^0 \simeq v_s^*$ does not give very large entropic contributions. If one assumes that $v_p^* = v_s^*$, then the entropic effect can be approximated by

$$-T\Delta\Delta S \simeq -\beta^{-1}\ln(v_s^0/v_p^0) \tag{30}$$

In this case, the statement that enzyme catalysis involves large entropic effects reduces to the statement that the available configurational space in the ground state in the enzyme is much smaller than the corresponding space in solution. A qualitative examination of the entropic contribution in the catalytic reaction of serine proteases has been presented in ref 2 and is depicted in Figure 10. At least in this case, it would thus appear that the magnitude of entropic effects is not very large.

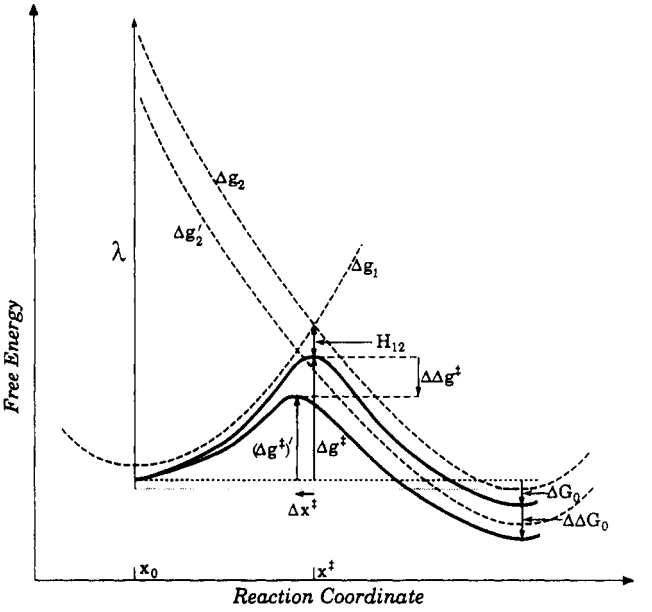


Figure 11. A schematic description of the relationship between the free-energy difference ΔG_0 and the activation free energy Δg^*. The figure illustrates how a shift of Δg_2 by $\Delta\Delta G_0$ (that changes Δg_2 to $\Delta g'_2$ and ΔG_0 to $\Delta G_0 + \Delta\Delta G_0$) changes Δg^* by a similar amount.

It is also important to comment here on the other type of entropic effect that is associated with bringing two reacting groups within bonding distance of each other.[90] This contribution has not been considered as part of entropic "puzzle" above. That is, the (proximity) effect of bringing the reactants to a contact distance in a solvent cage can be evaluated by rather trivial concentration considerations[2] and does not contribute further to the difference in activation energy between the enzyme and solution reactions once this cage has been defined. The really interesting entropic contributions are those associated with differences in the motion of the reacting fragments in the enzyme active site and in the reference solvent cage. These contributions seem to be smaller than previously thought since enzyme molecules are rather flexible. However, a definite answer to the entropic problem is clearly lacking. This is partially due to the difficulties with obtaining reliable estimates of entropy effects by FEP methods and it presents a major challenge for the future.

8.2. Examination of Linear Free-Energy Relationships

Linear free-energy relationships (LFERs) are among the most fundamental concepts in physical organic chemistry. The validity of such relationships for chemical processes in solutions has been the subject of many experimental and theoretical studies (e.g. refs 48–50 and 96–99) and appears to be reasonably established. The concept of LFERs can be best understood from Figure 11. For two harmonic free-energy functions of equal curvature one obtains in the range where $|\Delta G_0| < \lambda$ the expression (see ref 95)

$$\Delta g^* \approx (\Delta G^\circ + \lambda)^2/4\lambda - \bar{H}_{12}(X^*) + \bar{H}_{12}^2(X_0)/(\Delta G^\circ + \lambda) \tag{31}$$

$$\bar{H}_{12}(X_0) < |\Delta G^\circ + \lambda|/2$$

where X_0 and X^* are defined in Figure 11 and $\bar{H}_{12}$ is the average value of H_{12} at the given X. The first term of eq 31 is simply the well-known Marcus expression[96] for the diabatic case where $H_{12}(X) = 0$ and the two last terms reflect the effect of the adiabatic coupling of the two surfaces on the transition and reactant states, respectively. Obviously, the validity of eq 31 requires that the free-energy functions would be quadratic or in other words that the system will follow the linear response approximation. This indeed seems to be the case for charge-transfer reactions in solutions.[48,99,100] Although eq 31 formally represents a quadratic free-energy relationship (and not an LFER) there are certain ranges of $\Delta G°$ and λ where one will observe a linear dependence of Δg^* on $\Delta G°$. That is, the quadratic term in $\Delta G°$ is equal to $\Delta G°^2/4\lambda$ and the reorganization energy (λ) is often quite large (on the order of 100 kcal/mol, or so). Hence, for a small shift in $\Delta G°$, the quadratic dependence of $\Delta\Delta g^*$ on $\Delta\Delta G°$ can be rather small and an LFER may become observable.

The validity of LFERs in enzymatic reactions has not been fully established. Recent experimental studies[101,102] as well as early theoretical studies[2,52a,77,103] seem to indicate that such relationships are also valid in proteins. On the other hand it has been recently argued[104] that such relationships are not valid in general since the relevant interactions (e.g. the forces associated with hydrogen bonding) are not linear. It seems, however, that such an argument overlooks the fact that LFERs are relationships between Δg^* and $\Delta G°$ and that they are not related to the issue of whether the forces that determine $\Delta G°$ are linear or not (see refs 48 and 96 and also ref 105). It is, however, of great interest to try to examine the validity of LFERs in proteins in a systematic way.

If we again consider the initial proton-transfer step in HCAI (section 6.1), it would, in principle, be possible to examine LFERs by correlating the effect of mutations on $\Delta G^0_{1\to 2}$ with their effect on Δg^*. However, a given mutation may change both λ and $\Delta G°$ as well as the energetics of other reaction steps, thus complicating the analysis. Furthermore, as discussed in section 6, it is not known whether Δg^* actually reflects the initial ionization of the zinc-bound water (that corresponds to $\Delta G^0_{1\to 2}$) or a subsequent barrier for proton transfer out from the active site.

However, as was demonstrated in earlier studies of electron transfer (ET) and S_N2 reactions in solutions[48,99] it is possible to explore the validity of LFERs by computer simulation approaches. This can be accomplished by calculating the free-energy functions and the corresponding Δg^* for different assumed values of $\Delta G°$. One may thus take the protein contribution to the free-energy functions of the initial proton transfer step in HCAI as a generic test case and examine the relationship between Δg^* and $\Delta G°$ by repeating the simulations of Figure 2 for different values of the α^0_2 of eq 8. This type of calculation that examines the parametric dependence of Δg^* on $\Delta G°$ (viz. α) has been reported in ref 95, as well as calculations of Δg^* and $\Delta G°$ using different metal ions as catalysts in water.[60,106] Figure 12 shows the Δg_1 and Δg_2 curves for the reaction in the enzyme active site and demonstrates that these curves are almost quadratic (see ref 68 for a related study). Figure 13 summarizes the calculations of the

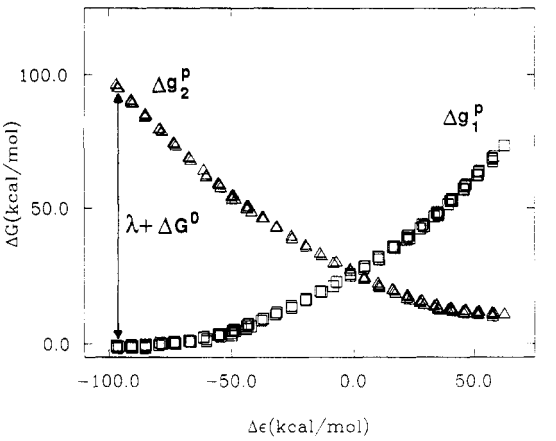


Figure 12. The calculated diabatic free-energy curves Δg_1 and Δg_2 for the initial proton transfer step in HCAI as a function of the corresponding energy gap.

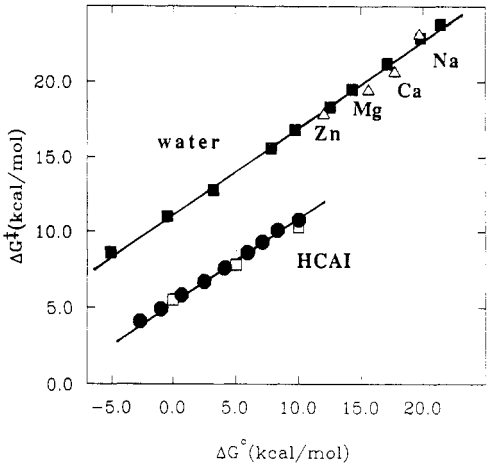


Figure 13. Examination of LFERs for the PT transfer reaction in water (upper curve) and the active site of HCAI (lower curve). Solid squares denote calculations in water where the parametric dependence of Δg^* on ΔG^0 is evaluated by changing the parameter $\Delta\alpha$ of the potential. Open triangles are data from separate calculations with different metal ions in the system.[60] Solid circles denote the calculations in HCAI where $\Delta\alpha$ is varied. Open squares represent the results predicted by eq 31 using the values $\lambda = 80$ kcal/mol, $\bar{H}_{12}(X^*) \simeq \bar{H}_{12}(X^0) \simeq 19$ kcal/mol.

dependence of Δg^* on $\Delta G°$ both in water and in the active site of HCAI. Also shown in Figure 13 is the relationship between Δg^* and $\Delta G°$ in the protein *predicted* by eq 31 using the values of λ, $\bar{H}_{12}(X^*)$, and $\bar{H}_{12}(X_0)$ obtained from the simulations.

As can be seen from the figure, rather regular LFERs are obtained for a typical PT reaction in solution and in the active site of an enzyme. It can also be seen that the two approaches corresponding to parametrically changing $\Delta G°$ by varying α and to substituting the "catalytic" ion fall on the same LFER in solution. It is also interesting to note that while the slope of the LFERs in solution and in the protein are very similar, the enzyme LFER is shifted downward. This reflects what one might call a "true transition-state stabilization" since the given $\Delta G°$ in the enzyme results in a barrier that is roughly 5 kcal/mol lower than that obtained for the same ΔG_0 in solution. The origin of this interesting effect is the reduction of the reorganization energy in the enzyme active site[68,95] (note that $\Delta g^* \sim \lambda/4$ when $\Delta G° \sim 0$ so that a smaller λ results in a smaller Δg^*).

The LFERs in Figure 13 also suggest that the enzyme does not exclusively work by "true transition-state stabilization" since the $\Delta G°$'s for the Zn^{2+}-catalyzed reaction is also reduced in the enzyme. Hence, the enzyme appears to use a combination of change in λ and $\Delta G°$ to optimize its reaction.

Apparently, the calculations of Figure 13 follow the trend predicted by eq 31 since the free-energy functions Δg_1 and Δg_2 (which are depicted in Figure 12) are almost quadratic even in the protein active site. This apparent quadratic behavior is due to the fact that the protein active site responds linearly to the development of electrostatic forces (see refs 52a, 68, and 108 for related observations). The linear response of the enzyme active site is a collective property of the protein that involves many compensating effects and occurs despite the fact that intermolecular forces such as hydrogen bonds are quite anharmonic. Thus, one would expect mutations that change $\Delta G°$ while leaving λ approximately the same to provide a regular LFER. It is important to emphasize, however, that these results establish the relationship between Δg^* and $\Delta G°$ for a transfer between two well-defined resonance structures. If the reaction involves several steps then the prediction would be that eq 31 should hold for each individual step but no such prediction is made for the overall dependence of Δg^* on the free energy difference between the reactants and products. Nevertheless, if the observation that eq 31 is valid for some enzymatic reactions is found to be a ubiquituous phenomenon, then LFERs may be a general rule for elementary steps in enzymatic reactions.

8.3. Calculations of Nuclear Tunneling and Zero-Point Energy Effects

The studies discussed in the previous sections have described the motion of the reacting atoms and their surrounding classically. A more rigorous treatment should consider the quantum mechanical aspects of the nuclear motion and evaluate nuclear tunneling and zero-point energy corrections. Such corrections could, for example, change the LFER trend predicted by eq 31. The simulation of quantum mechanical nuclear effects for reactions in proteins is far from simple. Approaches that can be used to describe diabatic processes such as electron-transfer reactions (e.g. ref 108) are not directly applicable to adiabatic processes such as PT reactions. Progress has been made in describing tunneling in PT reactions in solutions (e.g. refs 109 and 110) but the development of practical approaches for calculation of quantum corrections in enzyme reactions is still in its early stages.[111] One finds again that the EVB framework can provide a convenient way for obtaining quantum mechanical corrections to the rate constants of, e.g., PT and hydride transfer (HT) reactions in solutions and proteins. In using the EVB model for calculations of quantized rate constants it is reasonable to assume[109,112,113] that the quantum mechanical corrections can be attributed primarily to the corresponding activation barriers so that

$$k_{qu} \simeq k_{cl}\exp\{-\beta(\Delta g^*_{qu} - \Delta g^*_{cl})\} \quad (32)$$

where qu and cl designate quantum mechanical and classical quantities, respectively. Several strategies for the evaluation of Δg^* have been developed and examined.[108,109,111] The first one is based on the dispersed polaron (DP) model[108] which can be implemented for adiabatic problems as described in ref 109.

Another option that seems quite promising is the evaluation of the adiabatic free energy function by using the classical partition function as the reference for the quantum mechanical calculations. The starting point of this approach is a quantum mechanical equivalent of our expression for the classical free-energy function (eq 23). The corresponding quantized free-energy function is given by the path integral expression.[111]

$$\exp[-\beta\Delta g_{qu}(X)] \simeq \exp\{-\beta\Delta G_{qu}(\theta_m)] \times \langle\delta(X-\Delta\epsilon^c)\exp\{-(\beta/p)\sum_k[E_g(x_k)-\epsilon_m(\bar{x})]\}\rangle_{\epsilon_m^{qu}} \quad (33)$$

where the potentials E_g and ϵ_m^{qu} of eq 23 are replaced here by effective potentials which are given in the simple example of a one dimensional system by

$$\epsilon_m^{qu} = \sum_k^p \left\{\frac{1}{2p}M\omega^2(x_{k+1}-x_k)^2 + \frac{1}{p}\epsilon_m(x_k)\right\} \quad (34)$$

here each classical particle is replaced by a ring of quasiparticles, where M is the mass of our particle and $\Omega = p/\hbar\beta$ (the extension to the multidmensional case is quite simple). The δ function which is used to collect the contributions to the different values of the reaction coordinate involves $\Delta\epsilon^c = \epsilon_2(\bar{x}) - \epsilon_1(\bar{x}) =$ constant, where $\bar{x} = \sum^k x_k/p$. The evaluation of eq 33 by a direct approach is extremely time consuming since it requires the replacement of each atom by p quasiparticles. This problem can be simplified considerably and eq 33 can be evaluated using standard classical MD. That is, as shown in ref 109 we can reexpress eq 33 as

$$\exp[-\beta\Delta g_{qu}(X)] \simeq \exp[-\beta\Delta G_{qu}(\theta_m)] \times \langle\langle\int(X-\Delta\epsilon^c)\exp\{-(\beta/p)\sum_k[E_g(x_k)-\epsilon_m(\bar{x})]\}\rangle_{fp}\rangle_{\epsilon_m^{qu}} \quad (35)$$

where $\langle\ \rangle_{fp}$ designates an average over a free-particle distribution function. This approach appears to converge faster than eq 33 and can be implemented conveniently in standard MD programs. The use of eq 35 can be restricted to the solute atoms or include special light enzyme atoms. It is also possible to describe the quantum mechanical aspects of the enzyme atoms by representing them with the DP model (while still treating the solute atoms by the approach of eq 35). In the recent path integral study of hydride transfer in LDH the quantum mechanical treatment was restricted to the EVB atoms.[111] The calculations of quantum corrections for the initial PT step in HCAI[95] gave an isotope effect of 2.3 while the observed isotope effect is 3.8.[5] This, however, should not yet be considered as a validation of the computational approach since the rate-determining step might involve another PT step. In addition to evaluation of isotope effects one can use the approach of ref 111 to examine whether the classical LFER remains valid when quantum mechanical effects are taken into account. Preliminary studies have indicated that the quantum corrections are similar for the reaction in the protein and the reference solvent cage so that the corresponding LFER is probably

valid.[111] However, more studies with longer simulation times are still needed to establish this point.

9. Concluding Remarks

We have tried here to give an overview of the different strategies that are available for simulating enzymatic reactions today. While each method may have its strengths and weaknesses, it has become clear that a proper representation of the entire system in which the reacting fragments find themselves is essential for approaching quantitative success. Hence, seemingly rigorous treatments that only consider a limited part of the system or that do not treat the solute–solvent coupling in an adequate way may well yield results that are irrelevant for the real enzyme reaction. Of course, by treating the entire enzyme–substrate–solvent system one is usually forced to adopt some simplifications, such as the use of hybrid quantum/classical models, but it is this type of compromise that allows any progress to be made as long as a full quantum mechanical description of the entire system is impractical.

The review has emphasized the advantages of the EVB approach for modeling enzyme reactions. It should be noted that this method is not limited to charge separation and translocation reactions (as in the above examples) but can also easily be applied to, e.g., radical reactions. It can also be applied to reactions that involve many resonance structures. One of the most important aspects of the EVB method is its unique calibration possibilities, particularly using (experimental) information about relevant solution reactions, that provide a way to eliminate the type of "model errors" often encountered in gas-phase calculations (the method can, however, also be conveniently calibrated by accurate gas-phase ab initio calculations if this is preferable). This interpolation procedure starting from simple reference reactions in solution is essential at the present stage, since quantum mechanical methods have not reached the stage where calculations of enzyme reactions can be trusted, in particular with respect to energetics. Another inherent advantage of the EVB method is its ability to provide the diabatic free-energy surfaces of the relevant resonance structures. This allows one to examine, e.g., the microscopic validity of LFERs in proteins. EVB simulations[95] as well as experimental studies[101,102] have indicated that LFERs are valid even in enzyme-active sites (provided that they refer to the energetics of the internal resonance structures rather than simply to the energetics of reactants and products). Accepting the idea of LFERs in proteins, at least as an approximate view, may allow one to classify catalytic effects in a simple and powerful way by considering the interaction of the enzyme with the pure resonance structures rather than with transition states obtained from their mixing. In this way, e.g., it is quite easy to understand how stabilization of an ionic resonance structure by the charge of a divalent metal can lead to catalysis, by its "indirect" effect on the transition state which is a mixture containing this ionic component. This also appears to provide a quantitatively more useful concept than the customary statement that the metal polarizes the substrate (which seems more difficult to translate into a well-defined amount of free energy).

Studies employing the EVB approach as well as other methods seem to accumulate more and more evidence that points toward electrostatic effects as the key factor in enzyme catalysis.[2] This probably reflects the fact that it is more effective to catalyze reactions by electrostatic forces, that vary rather slowly with distance, than with steric forces. Actually, there does not seem to exist any example today of an enzyme reaction with significant rate acceleration that does not at some stage involve charge transformation, i.e. either charge separation or movement of charges. Nevertheless, contributions from other factors (besides electrostatic effects), such as entropy etc., may still be important for the rate enhancement that enzymes achieve.[3] It is our belief that computer simulation methods can play a very important role in elucidating the relative importance of different possible sources of catalysis. They might even turn out to be inexpendable in this respect, since much of the type of information that one is interested in is difficult to obtain by experimental means.

Acknowledgments. This work was supported by grants from the Swedish Natural Science Research Council (NFR), the NIH (grant GM 24492), and the ONR (grant no. OO14-91-J-1318).

References

(1) (a) Phillips, D. C. *Sci. Am.* **1966**, *215*, 78. (b) Blow, D. M.; Briktoft, J. J.; Hartley, B. S. *Nature* **1969**, *221*, 337. (c) Kartha, G.; Bello, J.; Harker, D. *Nature* **1967**, *213*, 862. (d) Wyckoff, H. W.; Hardman, K. D.; Allewell, N. M.; Inagami, T.; Tsernoglou, D.; Johnson, L. N.; Richards, F. M. *J. Biol. Chem.* **1967**, *242*, 3749.
(2) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and in Solutions*; Wiley: New York, 1991.
(3) Page, M. I. In *Enzyme Mechanism*; Page, M. I., Williams, A., Eds.; Royal Society of Chemistry; London, 1987; p 1.
(4) (a) Wilkinson, A. J.; Fersht, A. R.; Blow, D. M.; Carter, C.; Winter, G. *Nature (London)* **1984**, *307*, 187. (b) Craik, C. S.; Largman, C.; Fletcher, T.; Roczniak, S.; Barr, P. J.; Fletterick, R.; Rutter, W. J. *Science* **1985**, *228*, 291. (c) Wells, J. A.; Cunningham, B. C.; Graycar, T. P.; Estell, D. A. *Phil. Trans. R. Soc. London Ser. A* **1986**, *317*, 415. (d) Cronin, C. N.; Malcolm, B. A.; Kirsh, J. F. *J. Am. Chem. Soc.* **1987**, *109*, 2222. (e) Knowles, J. R. *Science* **1987**, *236*, 1252.
(5) Silverman, D. N.; Lindskog, S. *Acc. Chem. Res.* **1988**, *21*, 30.
(6) Jönsson, B.; Karlström, G.; Wennerström, H. *J. Am. Chem. Soc.* **1978**, *100*, 1658.
(7) (a) Liang, J.-Y.; Lipscomb, W. N. *J. Am. Chem. Soc.* **1986**, *108*, 5051. (b) Jacob, O.; Cardenas, R.; Tapia, O. *J. Am. Chem. Soc.* **1990**, *112*, 8692. (c) Peng, Z.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1992**, *114*, 2733. (d) Merz, K. M., Jr.; Hoffmann, R.; Dewar, M. J. S. *J. Am. Chem. Soc.* **1989**, *111*, 5636.
(8) (a) Pullman, A. *Ann. N. Y. Acad. Sci.* **1981**, *367*, 340. (b) Cook, C. M.; Allen, L. C. *Ann. N. Y. Acad. Sci.* **1984**, *429*, 84. (c) Liang, J.-Y.; Lipscomb, W. N. *Biochemistry* **1988**, *27*, 8676. (d) Solà, M.; Lledòs, A.; Duran, M.; Bertràn, J. *J. Am. Chem. Soc.* **1992**, *114*, 869. (e) Krauss, M.; Garmer, D. R. *J. Am. Chem. Soc.* **1991**, *113*, 6426. (f) Tapia, O.; Jacob, O.; Colonna, F. *Theor. Chim. Acta* **1992**, in press. (g) Zheng, Y.-J.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1992**, *114*, 10498.
(9) (a) Umeyama, H.; Imamura, A.; Nagata, C.; Hanano, M. *J. Theor. Biol.* **1973**, *41*, 485. (b) Scheiner, S.; Kleier, D. A.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 2606. (c) Scheiner, S.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 432. (d) Bolis, G.; Ragazzi, M.; Salvaderi, D.; Ferro, D. R.; Clementi, E. *Gazz. Chim. Ital.* **1978**, *108*, 425. (e) van Duijnen, P. Th.; Thole, B. Th.; Hol, W. G. J. *Biophys. Chem.* **1979**, *9*, 273. (f) Kollman, P. A.; Hayes, D. M. *J. Am. Chem. Soc.* **1981**, *103*, 2955. (g) Dewar, M. J. S.; Storch, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2225. (h) Stamato, F. M. L. G.; Tapia, O. *Int. J. Quantum Chem.* **1988**, *33*, 187.
(10) (a) Scheiner, S.; Lipscomb, W. N. *J. Am. Chem. Soc.* **1977**, *99*, 3466. (b) Alex, A.; Clark, T. *J. Comput. Chem.* **1992**, *13*, 704.
(11) (a) Tapia, O.; Cardenas, R.; Andres, J.; Colonna-Cesari, F. *J. Am. Chem. Soc.* **1988**, *110*, 4046. (b) Tapia, O.; Cardenas, R.; Andres, J.; Krechl, J.; Campillo, M.; Colonna-Cesari, F. *Int. J. Quantum Chem.* **1991**, *39*, 767. (c) Wu, Y.-D.; Houk, K. N. *J. Am. Chem. Soc.* **1991**, *113*, 2353. (d) Wilkie, J.; Williams, I. H. *J. Am. Chem. Soc.* **1992**, *114*, 5423.
(12) Goldblum, A. *Biochemistry* **1988**, *27*, 1653.

(13) Alagona, G.; Desmeules, P.; Ghio, C.; Kollman, P. A. *J. Am. Chem. Soc.* **1984**, *106*, 3623.
(14) Osman, R.; Basch, H. *J. Am. Chem. Soc.* **1984**, *106*, 5710.
(15) Andrés, J.; Safont, V. S.; Tapia, O. *Chem. Phys. Lett.* **1992**, *198*, 515.
(16) Mulholland, A. J.; Grant, G. H.; Richards, W. G. *Protein Eng.* **1993**, *6*, 133.
(17) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 154.
(18) Tapia, O.; Lluch, J. M.; Cardenas, R.; Andres, J. *J. Am. Chem. Soc.* **1989**, *111*, 829.
(19) (a) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486. (b) Kirkwood, J. G. *J. Chem. Phys.* **1939**, *7*, 911.
(20) (a) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213. (b) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8552.
(21) Luzhkov, V.; Warshel, A. *J. Comput. Chem.* **1992**, *13*, 199.
(22) Krishtalik, L. I. *Theor. Biol.* **1985**, *112*, 251.
(23) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1991**, *95*, 4366.
(24) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(25) (a) Warshel, A. *Curr. Opin. Struct. Biol.* **1992**, *2*, 230. (b) Kollman, P. A. *Curr. Opin. Struct. Biol.* **1992**, *2*, 765.
(26) Vaidehi, N.; Wesolowski, T. A.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 4264.
(27) Warshel, A.; Sussman, F.; Hwang, J.-K. *J. Mol. Biol.* **1988**, *201*, 139.
(28) Dewar, M. J. S.; Haselbach, E. *J. Am. Chem. Soc.* **1970**, *92*, 590.
(29) (a) Lifson, S.; Warshel, A. *J. Chem. Phys.* **1968**, *49*, 5116. (b) Warshel, A.; Karplus, M. *J. Am. Chem. Soc.* **1974**, *96*, 5677.
(30) (a) Lushkov, V.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4491. (b) Warshel, A.; Chu, Z. T.; Hwang, J.-K. *Chem. Phys.* **1991**, *158*, 303.
(31) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
(32) (a) Weiner, S. J.; Singh, U. C.; Kollman, P. A. *J. Am. Chem. Soc.* **1985**, *107*, 2219. (b) Weiner, S. J.; Seibel, G. L.; Kollman, P. A. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 649.
(33) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. *J. Chem. Soc., Perkin Trans. 2* **1991**, 225.
(34) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
(35) (a) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700. (b) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826.
(36) (a) Tapia, O.; Goscinski, O. *Mol. Phys.* **1975**, *29*, 1653. (b) Tapia, O.; Johannin, G. *J. Chem. Phys.* **1981**, *75*, 24.
(37) (a) Rullmann, J. A. C.; Bellido, M. N.; van Duijnen, P. Th. *J. Mol. Biol.* **1989**, *206*, 101. (b) Dijkman, J. P.; Osman, R.; Weinstein, H. *Int. J. Quantum Chem.* **1989**, *35*, 241. (c) Thole, B. T.; van Duijnen, P. Th. *Biophys. Chem.* **1983**, *18*, 53.
(38) Aqvist, J.; Luecke, H.; Quiocho, F. A.; Warshel, A. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 2026.
(39) See for example: (a) Newton, M. D. *J. Phys. Chem.* **1975**, *79*, 2795. (b) Hylton-McCreery, J.; Christofferson, R. E.; Hall, G. G. *J. Am. Chem. Soc.* **1976**, *98*, 7191. (c) Rivail, J.-L.; Rinaldi, D. *Chem. Phys.* **1976**, *18*, 233. (d) Bonaccorsi, R.; Cimaraglia, R.; Tomasi, J. *J. Comput. Chem.* **1983**, *4*, 567. (e) Mikkelsen, K. V.; Dalgaard, E.; Swanstrom, P. *J. Phys. Chem.* **1987**, *91*, 3081. (f) Karlström, G. *J. Phys. Chem.* **1988**, *92*, 1315. (g) Karelson, M. M.; Katritzky, A. R.; Szafran, M.; Zerner, M. C. *J. Org. Chem.* **1989**, *54*, 6030. (h) Wong, M. W.; Frisch, M. J.; Wiberg, K. B. *J. Am. Chem. Soc.* **1991**, *113*, 4776.
(40) (a) Kim, H. J.; Hynes, J. T. *Int. J. Quantum Chem.: Quantum Chem. Symp.* **1990**, *24*, 821. (b) Kim, H. J.; Hynes, J. T. *J. Am. Chem. Soc.* **1992**, *114*, 10508; 10528.
(41) (a) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671. (b) Gilson, M. K.; Honig, B. H. *Proteins* **1988**, *3*, 32. Rashin, A. A. *J. Phys. Chem.* **1990**, *94*, 1725.
(42) See, for example: Parr, R. G. *Ann. Rev. Phys. Chem.* **1983**, *34*, 631.
(43) Bajorath, J.; Kraut, J.; Li, Z.; Kiston, D. H.; Hagler, A. T. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 6423.
(44) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218–6226.
(45) (a) Goddard, W. A., III; Dunning, T. H., Jr.; Hunt, W. J.; Hay, P. J. *Acc. Chem. Res.* **1973**, *6*, 368. (b) Epiotis, N. D.; Larson, J. R.; Eaton, H. *Unified Valence Bond Theory of Electronic Structure*, Springer-Verlag: Heidelberg, 1982. (c) Cooper, D. L.; Garratt, J.; Raimondi, M. *Nature* **1986**, *323*, 699.
(46) (a) Warshel, A.; Russell, S. *J. Am. Chem. Soc.* **1986**, *108*, 6569. (b) Chang, Y.-T.; Miller, W. H. *J. Phys. Chem.* **1990**, *94*, 5884.
(47) Coulson, C. A.; Danielson, U. *Ark. Fys.* **1954**, *8*, 239.
(48) Hwang, J.-K.; King, G.; Creighton, S.; Warshel, A. *J. Am. Chem. Soc.* **1988**, *110*, 5297.
(49) Hammond, G. S. *J. Am. Chem. Soc.* **1955**, *77*, 334.
(50) Albery, J. W.; Kreevoy, M. M. *Adv. Phys. Org. Chem.* **1978**, *16*, 87.
(51) (a) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, (b) Valleau, J. P.; Torrie, G. M. In *Modern Theoretical Chemistry*; Berne, B. J., Ed.; Plenum: New York, 1977; Vol. 5, p 169.
(52) (a) Warshel, A.; Russell, S.; Sussman, F. *Isr. J. Chem.* **1986**, *27*, 217. (b) van Gunsteren, W. F., Weiner, P. K., Eds. *Computer Simulation of Biomolecular Systems*; ESCOM: Leiden, 1988. (c) Beveridge, D. L.; DiCapua, F. M. *Ann. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431. (d) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184. (e) Kollman, P. A.; Merz, K. M., Jr. *Acc. Chem. Res.* **1990**, *23*, 246. (f) Straatsma, T. P.; McCammon, J. A. *Ann. Rev. Phys. Chem.* **1992**, *43*, 407.
(53) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.
(54) Warshel, A. *J. Phys. Chem.* **1982**, *86*, 2218.
(55) Bash, P. A.; Field, M. J.; Karplus, M. *J. Am. Chem. Soc.* **1987**, *109*, 8092.
(56) (a) Aqvist, J.; Warshel, A. *J. Mol. Biol.* **1992**, *224*, 7. (b) Aqvist, J.; Fothergill, M.; Warshel, A. *J. Am. Chem. Soc.* **1993**, *115*, 631.
(57) (a) Liljas, A.; Kannan, K. K.; Bergstén, P.-C.; Waara, I.; Fridborg, K.; Strandberg, B.; Carlbom, U.; Järup, L.; Lövgren, S.; Petef, M. *Nature New Biol.* **1972**, *235*, 131. (b) Kannan, K. K.; Ramanadham, M.; Jones, T. A. *Ann. N. Y. Acad. Sci.* **1984**, *429*, 49. (c) Eriksson, A. E.; Kylsten, P. M.; Jones, T. A.; Liljas, A. *Proteins* **1988**, *4*, 283. (d) Håkansson, K.; Carlsson, M.; Svensson, L. A.; Liljas, A. *J. Mol. Biol.* **1992**, *227*, 1192.
(58) (a) Lindskog, S.; Engberg, P.; Forsman, C.; Ibrahim, S. A.; Jonsson, B.-H.; Simonsson, I.; Tibell, L. *Ann. N. Y. Acad. Sci.* **1984**, *429*, 61. (b) Behravan, G. Studies of site-specific mutants of carbonic anhydrase. Ph.D. Thesis, University of Umeå, 1990. (c) Behravan, G.; Johnsson, B.-H.; Lindskog, S. *Eur. J. Biochem.* **1990**, *190*, 351. (d) Tu, C.; Silverman, D. N.; Forsman, C.; Jonsson, B.-H.; Lindskog, S. *Biochemistry* **1989**, *28*, 7913. (e) Campbell, I. D.; Lindskog, S.; White, A. I. *J. Mol. Biol.* **1974**, *90*, 469. (f) Simonsson, I.; Jonsson, B.-H.; Lindskog, S. *Eur. J. Biochem.* **1982**, *129*, 165.
(59) Eigen, M.; de Maeyer, L. *Z. Elektrochem.* **1955**, *59*, 986.
(60) (a) Aqvist, J. *J. Phys. Chem.* **1991**, *95*, 4587. (b) Aqvist, J. *J. Mol. Struct. (THEOCHEM)* **1992**, *256*, 135.
(61) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; pp 331–342.
(62) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647.
(63) (a) Warshel, A. *Biochemistry* **1981**, *20*, 3167. (b) Russell, S. T.; Warshel, A. *J. Mol. Biol.* **1985**, *185*, 389.
(64) (a) Magid, E.; Turbeck, B. O. *Biochim. Biophys. Acta* **1968**, *165*, 515. (b) Pinsent, B. R. W.; Pearson, L.; Roughton, F. J. W. *Trans. Faraday Soc.* **1956**, *52*, 1512.
(65) (a) Merz, K. M., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 406. (b) Merz, K. M., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 3572.
(66) Liang, J.-Y.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 3675.
(67) The experimental estimate for this quantity is about 29–31 kcal/mol. See ref 21, 7c, and the following: Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283.
(68) Yadav, A.; Jackson, R. M.; Holbrook, J. J.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4800.
(69) Xue, Y.; Vidgren, J.; Svensson, L. A.; Liljas, A.; Jonsson, B.-H.; Lindskog, S. Manuscript in preparation.
(70) Rao, S. N.; Singh, U. C.; Bash, P. A.; Kollman, P. A. *Nature (London)* **1987**, *328*, 551.
(71) Hwang, J.-K.; Warshel, A. *Biochemistry* **1987**, *26*, 2669.
(72) Hunkapillar, M. W.; Smallcombe, S. H.; Whitaker, D. R.; Richards, J. H. *Biochemistry* **1973**, *12*, 4732.
(73) (a) Umeyama, H.; Nakagawa, S.; Kudo, T. *J. Mol. Biol.* **1981**, *150*, 409. (b) Naray-Szabo, G. *Int. J. Quantum Chem.* **1982**, *22*, 582. (c) Naray-Szabo, G. *Int. J. Quantum Chem.* **1983**, *23*, 723. (d) Warshel, A.; Russell, S. *J. Am. Chem. Soc.* **1986**, *108*, 6569.
(74) (a) Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J.-K. *Biochemistry* **1989**, *28*, 3629. (b) Daggett, V.; Schröder, S.; Kollman, P. *J. Am. Chem. Soc.* **1991**, *113*, 8926.
(75) Aqvist, J.; Warshel, A. *Biochemistry* **1989**, *28*, 4680.
(76) Sepersu, E. H.; Shortle, D.; Mildvan, A. S. *Biochemistry* **1987**, *26*, 1289.
(77) Aqvist, J.; Warshel, A. *J. Am. Chem. Soc.* **1990**, *112*, 2860.
(78) Alagona, G.; Desmeules, P.; Ghio, C.; Kollman, P. A. *J. Am. Chem. Soc.* **1984**, *106*, 3623.
(79) Lodi, P. J.; Knowles, J. R. *Biochemistry* **1991**, *30*, 6948.
(80) Singh, U. C. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 4280.
(81) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* **1987**, *235*, 574.
(82) Wong, C. F.; McCammon, J. A. *J. Am. Chem. Soc.* **1986**, *108*, 3830.
(83) Brooks, C. L., III. In *Computer Simulation of Biomolecular Systems*, van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM: Leiden, 1989; p 73.
(84) Caldwell, J. W.; Agard, D. A.; Kollman, P. A. *Proteins* **1991**, *10*, 140.
(85) Merz, K. M., Jr.; Murcko, M. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1991**, *113*, 4484.
(86) (a) Werber, M. M. *J. Theor. Biol.* **1976**, *60*, 51. (b) Woolley, P. *Nature* **1975**, *258*, 677.
(87) (a) Klug, D.; Rabani, J.; Fridovich, I. *J. Biol. Chem.* **1972**, *247*, 4839. (b) Allison, S. A.; Bacquet, R. J.; McCammon, J. A. *Biopolymers* **1988**, *27*, 251. (c) Sharp, K.; Fine, R.; Honig, B. *Science* **1987**, *236*, 1460.
(88) Warshel, A.; Aqvist, J.; Creighton, S. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 5820.
(89) Page, M. I.; Jencks, W. P. *Proc. Natl. Acad. Sci. U. S. A.* **1971**, *68*, 1678.
(90) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; Dover Publications Inc.: New York, 1987.

(91) Warshel, A. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, 5250.
(92) (a) Storm, D. R.; Koshland, D. E., Jr. *Proc. Natl. Acad. Sci. U. S. A.* **1970**, *66*, 445. (b) Gorenstein, D. G. *Chem. Rev.* **1987**, *87*, 1047.
(93) Menger, F. M. *Biochemistry* **1992**, *31*, 5368.
(94) Bruice, T. C. *Ann. Rev. Biochem.* **1976**, *45*, 331.
(95) Warshel, A.; Hwang, J. K.; Aqvist, J. *Faraday Discuss.* **1992**, *93*, 225.
(96) Marcus, R. A. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155.
(97) Albery, W. J. *Annu. Rev. Phys. Chem.* **1980**, *31*, 227.
(98) (a) Kreevoy, M. M.; Kotchevar, A. T. *J. Am. Chem. Soc.* **1990**, *112*, 3579. (b) Kreevoy, M. M.; Ostovic, D.; Lee, I. H.; Binder, D. A.; King, G. W. *J. Am. Chem. Soc.* **1988**, *110*, 524.
(99) King, G. A.; Warshel, A. *J. Chem. Phys.* **1990**, *93*, 8682.
(100) Kuharski, R. A.; Bader, J. S.; Chandler, D.; Sprik, M.; Klein, M. L.; Impey, R. W. *J. Chem. Phys.* **1988**, *89*, 3248.
(101) Fersht, A. R.; Leatherbarrow, R. J.; Wells, T. N. C. *Nature* **1986**, *322*, 284.
(102) Toney, M. D.; Kirsch, J. F. *Science* **1989**, *243*, 1485.
(103) Warshel, A. *Pontif. Acad. Sci. Script. Varia* **1984**, *55*, 59.
(104) Straub, J. E.; Karplus, M. *Protein Eng.* **1990**, *3*, 673.
(105) Fersht, A. R.; Wells, T. N. C. *Protein Eng.* **1991**, *4*, 229.
(106) Aqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021.
(107) Burgess, M. A. *Metal Ions in Solution*; Ellis Horwood Ltd.: Chichester, England, 1978.
(108) Warshel, A.; Chu, Z. T.; Parson, W. W. *Science* **1989**, *246*, 112.
(109) Warshel, A.; Chu, Z. T. *J. Chem. Phys.* **1990**, *93*, 4003.
(110) Borgis, D. C.; Lee, S.; Hynes, J. T. *Chem. Phys. Lett.* **1989**, *162*, 19.
(111) Hwang, J.-K.; Chu, Z. T.; Warshel, A. *J. Phys. Chem.* **1991**, *95*, 8445.
(112) Gillan, M. J. *Phys. Rev. Lett.* **1987**, *58*, 563.
(113) Voth, G. A.; Chandler, D.; Miller, W. H. *J. Chem. Phys.* **1989**, *91*, 7749.