The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions

Shina C. L. Kamerlin and **Arieh Warshel**[*]
Department of Chemistry SGM418, University of Southern California, 3620 McClintock Ave., Los Angeles, CA-90089, USA.

Abstract

Recent years have seen dramatic improvements in computer power, allowing ever more challenging problems to be approached. In light of this, it is imperative to have a quantitative model for examining chemical reactivity, both in the condensed phase and in solution, as well as to accurately quantify physical organic chemistry (particularly as experimental approaches can often be inconclusive). Similarly, computational approaches allow for great progress in studying enzyme catalysis, as they allow for the separation of the relevant energy contributions to catalysis. Due to the complexity of the problems that need addressing, there is a need for an approach that can combine reliability with an ability to capture complex systems in order to resolve long-standing controversies in a unique way. Herein, we will demonstrate that the empirical valence bond (EVB) approach provides a powerful way to connect the classical concepts of physical organic chemistry to the actual energies of enzymatic reactions by means of computation. Additionally, we will discuss the proliferation of this approach, as well as attempts to capture its basic chemistry and repackage it under different names. We believe that the EVB approach is the most powerful tool that is currently available for studies of chemical processes in the condensed phase in general and enzymes in particular, particularly when trying to explore the different proposals about the origin of the catalytic power of enzymes.

I. Introduction

Recent years have seen dramatic improvements in computer power, allowing for ever more challenging problems to be approached. In light of this, it is imperative to have a quantitative model for examining chemical reactivity, both in the condensed phase and in solution. Additionally, in the same vein, such approaches are very important in accurately quantifying physical organic chemistry, particularly as experimental approaches can often be inconclusive.[1] Furthermore, experimental attempts to identify key factors in such fundamental problems as the origin of enzyme catalysis simply cannot lead to unique conclusions, since the relevant energy contributions cannot be separated without some computational models (see ref. 2 for further details). The problems in the field are very complex, and as such there is a need for an approach that can combine reliability with an ability to capture complex systems in order to resolve long-standing controversies in a

[*]warshel@usc.edu; Fax: +1 (213) 740 2701; Tel: +1 (213) 740 4114 .

unique way. Here, we will demonstrate that at present, the optimal tool for exploring different proposals about the origin of the catalytic power of enzymes is the empirical valence bond (EVB) approach. We will additionally discuss the proliferation of the use of the EVB, as well as attempts to capture its basic chemistry and repackage it under different names.

II. The EVB as a reliable semi-empirical QM/MM method

II.1 General

If one wants to gain a quantitative understanding of enzymatic reactions (as well as the corresponding reactions in solution), it is essential to be able to calculate the free energy profiles for these reactions. The common approach of obtaining potential surfaces for chemical reactions involves the use of quantum mechanical computational approaches, and such approaches have become quite effective when treating small molecules in the gas phase (*e.g.* ref. 3). However, here we are interested in chemical reactions in very large systems, which at present cannot be explored by *ab initio* methods. Similarly, though molecular mechanics simulations (*e.g.* ref. 4 and 5) have been proven to be very effective in exploring the protein configurational space, they cannot be used to describe bond breaking and bond making reactions in proteins or solutions. The generic solution to the above problem has been provided by the development of the hybrid quantum mechanics/molecular mechanics (QM/MM) approach,[6] which in recent years has become the key tool for calculating protein function in general and for studying chemical processes in proteins in particular. Here, we can only mention a few works in this field (*e.g.* ref. 7–30). However, despite these advances, we are not yet at the stage where one can use QM/MM approaches in fully quantitative studies of enzyme catalysis. The major problem is associated with the fact that while a quantitative evaluation of the potential surfaces for the reacting fragment should involve *ab initio* electronic structure calculations, such calculations are too expensive to allow for the extensive configurational averaging needed for proper free energy calculations, and the use of a QM/MM approach without proper sampling is not so effective.[31] This is a painful fact, despite the ever increasing popularity of using energy minimization approaches without proper sampling (*e.g.* ref. 17, 32 and 33, amongst others). Specialized approaches can help one to move towards *ab initio* QM/MM free energy calculations (*e.g.* ref. 5, 34 and 35) but, despite great progress in this direction (see ref. 5 for an overview), even these approaches are still in a development stage. Fortunately, one can use approaches that are calibrated to the energetics of the reference solution reaction to obtain reliable results with semi-empirical QM/MM studies, and the most effective way of doing so is the EVB approach described below.

As stated above, reliable studies of enzyme catalysis require accurate results for the difference between the activation barriers in enzyme and in solution. The early realization of this fact led to a search for a method that can be calibrated using experimental and theoretical information of reactions in solution. It also becomes apparent that in studies of chemical reactions, it is more physical to calibrate surfaces that reflect bond properties (*i.e.* valence bond-based, VB, surfaces) than to calibrate surfaces that reflect atomic properties (*e.g.* MO-based surfaces). Furthermore, it appears to be very advantageous to force the

potential surfaces to reproduce the experimental results of the broken fragments at infinite separation in solution. This can be easily accomplished with the VB picture. The resulting empirical valence bond (EVB) method has been discussed extensively elsewhere,[36,37] but its main features will be outlined below for readers who may be unfamiliar with this approach.

The EVB is a QM/MM method that describes reactions by mixing resonance states (or more precisely diabatic states) that correspond to classical valence-bond (VB) structures, which describe the reactant, intermediate (or intermediates), and product states. The potential energies of these diabatic states are represented by classical MM force fields of the form:

$$\varepsilon_i = \alpha_{gas}^i + U_{intra}^r(\mathbf{R}, \mathbf{Q}) + U_{Ss}^i(\mathbf{R}, \mathbf{Q}, \mathbf{r}, \mathbf{q}) + U_{ss}(\mathbf{r}, \mathbf{q}) \quad (1)$$

Here, $\mathbf{R}$ and $\mathbf{Q}$ represent the atomic coordinates and charges of the diabatic states, and $\mathbf{r}$ and $\mathbf{q}$ are those of the surrounding protein and solvent, α_{gas}^i is the gas-phase energy of the ith diabatic state (where all the fragments are taken to be at infinity), $U_{intra}(\mathbf{R},\mathbf{Q})$ is the intramolecular potential of the solute system (relative to its minimum), $U_{Ss}(\mathbf{R},\mathbf{Q},\mathbf{r},\mathbf{q})$ represents the interaction between the solute (S) atoms and the surrounding (s) solvent and protein atoms, $U_{ss}(\mathbf{r},\mathbf{q})$ represents the potential energy of the protein–solvent system ("ss" designates surrounding-surrounding). The ε_i of eqn (1) forms the diagonal elements of the EVB Hamiltonian (H_{ii}). The off-diagonal elements of the Hamiltonian, H_{ij}, are represented by simple exponential functions of the distances between the reacting atoms. The H_{ij} elements are assumed to be the same in the gas phase, in solutions and in proteins. The ground state energy, E_g, is obtained by diagonalizing the EVB Hamiltonian.

The EVB approach evaluates the relevant activation free energies ($\Delta g^{\ddagger}$) by changing the system adiabatically from one diabatic state to another. In the simple case of two diabatic states, this "mapping" potential, ε_m, can be written as a linear combination of the reactant and product potentials, ε_1 and ε_2:

$$\varepsilon_m = (1 - \lambda_m)\varepsilon_1 + \lambda\varepsilon_2, (0 \leq \lambda \leq 1) \quad (2)$$

When λ_m is changed from 0 to 1 in $n + 1$ fixed increments ($\lambda_m = 0/n, 1/n, 2/n,., n/n$), and potentials with one or more of the intermediate values of λ_m will force the system to fluctuate near the TS.

The free energy, ΔG_m, associated with changing λ_m from 0 to m/n is evaluated by the wellknown free energy perturbation (FEP) procedure (which has been described in detail in, for example, ref. 37, amongst others). However, after obtaining ΔG_m, we still need to obtain the free energy that corresponds to the adiabatic ground state surface along the reaction coordinate, x. This free energy (referred to as a "free energy functional") is obtained by the FEP-umbrella sampling (FEP/US) method, which is described elsewhere.[37,38] The main point for the purposes of this work is that the FEP/US approach may be also used to obtain the free energy functional of the diabatic states by:

$$\Delta g_1\left(x^{'}\right) = \Delta G_m - \beta^{-1}ln\left\langle\delta\left(x - x^{'}\right)exp\left[-\beta\left(\varepsilon_1\left(x\right) - \varepsilon_m\left(x\right)\right)\right]\right\rangle_{\varepsilon_m} \quad (3)$$

where ε_m is the mapping potential that keeps x in the region of x'. If the changes in ε_m are sufficiently gradual, the free energy functional $g(x')$ obtained with several values of m overlap over a range of x', and patching together the full set of $g(x')$ gives the complete free energy curve for the reaction. The generated reaction coordinate, x, is usually taken to define the energy gap ($x = \varepsilon_1 - \varepsilon_2$). This selection[37,39] is particularly powerful when one tries to represent the entire many dimensional solvent space by a single coordinate (see ref. 38).

The diabatic free energy profiles of the reactant and product states represent a microscopic equivalent of the Marcus parabolae.[40,41] For example, in the case of the ($Cl^- + CH_3$–$Cl \rightarrow ClCH_3 + Cl^-$) S_N2 reaction, one obtains[38] the results shown in Fig. 1. Thus, the EVB provides the unique ability to correctly capture the linear relationship between activation free energies and reaction energies (LFER) that is observed in many important reactions (see, for example, ref. 37). Furthermore, the EVB benefits from the aforementioned ability to consistently and conveniently treat the solute–solvent coupling. This feature is essential, not only in allowing one to properly model charge-separation reactions, but also in allowing for reliable and convenient calibration. Calibrating EVB surfaces using *ab initio* calculations was found to provide quite reliable potential surfaces.

II.2 The proliferation of the EVB

It should be noted that the seemingly simple appearance of the EVB approach may have led to the initial impression that this is an oversimplified qualitative model (rather than a powerful quantitative approach). This attitude is nicely illustrated in the discussions section of a previous *Faraday Discussion* on the structure and activity of enzymes.[42] Nevertheless, the EVB model has eventually been widely adopted as a general model for studies of reactions in large molecules and in the condensed phase (*e.g.* ref. 43–46), despite the continued misunderstandings about its theory and applications (see below for further discussion). Similarly, several very closely related versions have been put forward that basically have the same ingredients as in the EVB approach (see the discussion in ref. 47 and 48).

Since we will be dealing with proton transport processes as key examples, it might be useful to clarify that the EVB and the so-called MS-EVB[49,50] (that was so effective in studies of proton transport in water) are more or less identical. More specifically, the so-called MS-EVB typically includes 6 EVB states in the solute quantum mechanical (QM) region and the location of this QM region changes if the proton moves. The QM region is surrounded by classical water molecules (the molecular mechanics (MM) part), whose effect is sometimes included incorrectly by solvating the charges of the gas phase QM region instead of solvating the diabatic charges (this leads to inconsistent QM/MM coupling with the solute charges, as explained in, for example, ref. 4 and 51). More recently the coupling was consistently introduced by adding the interaction with the MM water in the diagonal solute Hamiltonian. Now the multi-state idea is not new, as our EVB studies were performed repeatedly with a multi-state treatment (*e.g.* 5 states in ref. 52), and this has always been done with consistent coupling to the MM region. Thus, the only difference that we can find between the two versions is that our EVB studies did not update the location of the QM

region during the simulations, since they dealt with processes in proteins where the barrier is high, rather than with low barrier transport processes (so that the identity of the reacting region has not changed during the simulations). Also, note that the MS-EVB simulations in proteins, where we have a limited number of quantum sites, do not have to change the QM region (*e.g.* ref. 53) during the simulations. Finally, in cases of high barriers, the change of the QM region is not useful, since the main issue is the ability to obtain a proper evaluation of the free energy associated with climbing the barrier. Thus, we conclude that the EVB and MS-EVB are basically identical methods, although we appreciate the elegant treatment of changing the position of the QM region during simulations, which is a very useful advance in EVB treatments of processes with a low activation barrier.

Perhaps the best example of the realization of the potential of the EVB approach by the wider scientific community (and thus attempts to exploit this potential) is a recent article comparing different diabatic models of chemical reactivity[54] that severely criticizes the EVB approach, based on scientifically incorrect information, while overlooking the fact that one of the authors is using it himself for both gasphase[55] and solution studies.[56] This issue has already been discussed in great length in ref. 57, so here we will only highlight the most problematic points. Namely, the most serious incorrect criticisms of the EVB approach in this work can be grouped into five categories: (i) an attempt to reproduce our EVB gas-phase surface (in order to "show" that this is a problematic surface) while using incorrect parameters, which resulted in an incorrect surface that has little to do with the actual EVB surface as discussed in section III.1 (though we should point out that the provision of computational material demonstrating why this EVB surface is erroneous on our website, see section "EVB Verification" on http://futura.usc.edu, has resulted in the retraction of all data pertaining to this EVB surface); (ii) claims that the EVB solution results are irreproducible when the approach used to try to reproduce them is actually incorrect, as was discussed in detail in ref. 57; (iii) claims that the EVB cannot calculate "detailed rate quantities such as kinetic isotope effects that require an accurate treatment of the potential energy surface, zero-point energy (ZPE), and quantum mechanical tunneling", whereas there are examples of the accurate calculation of all these properties in the literature that the authors neglect to cite; this issue was discussed in detail in ref. 57; (iv) claims that the EVB has never been parameterized to the *ab initio* surface, where this was done repeatedly (starting in 1988), and has been clarified in our papers in addressing the authors' allegations; and (v) Questioning the EVB usage of orthogonalized diabatic states and solvent-independent off-diagonal elements, while ignoring the fact that a CDFT study[58] has established the validity of such an approximation. It should be noted that this review has been presented at a time when the most senior author of ref. 54 has recently introduced an approach (which he calls the electrostatically embedded multi-configurational molecular mechanics (MCMM) approach[56]), which, for all intents and purposes, is identical to the EVB approach, as was discussed in our recent Centennial Article for *The Journal of Physical Chemistry*[5] and in ref. 57. This instructive issue will also be touched on in section IX.

Overall, we believe that despite the fact that ref. 54 is factually incorrect on numerous grounds (as we have discussed in detail elsewhere[57]), nevertheless the fact that such a

perspective was even written provides an instructive example of the growing appreciation of the power of the EVB, as well as misunderstanding of the fact that the foundations of the EVB are quite rigorous. In light of the significant errors in this perspective, we have highlighted some of the key problems below, in order to prevent confusion for readers who are unfamiliar with the field.

III. Highlighting the key features of the EVB

This section will focus on the key features of the EVB, and explain why it is such an effective yet rigorous approach.

III.1 The reliability of the EVB parameterization and EVB surfaces

The EVB is obviously a semi-empirical approach that relies on proper parameterization, and our works have focused on searching for an effective and reliable parameterization approach from day one. Here, it is useful to clarify what effective parameterization *is*, in particular due to the fact that a large part of the community may believe that reproducing the exact gas-phase *ab initio* result is the most productive way of moving in semi-empirical models. In fact, it is easy to tell those who are not familiar with the reliability of studies of enzymatic reactions that the best approach should involve calibration on gas-phase *ab initio* results.[56] However, such a proposition demonstrates major inexperience in the case of the study of enzyme catalysis. That is, in our extensive experience of actually *doing* such calculations, we have realized that the best way to parameterize reliable EVB surfaces for studies in condensed phases is to calibrate to *ab initio* calculations, using either implicit or explicit solvation models.

The reason for this (which was already discussed at length in ref. 34) is that, in the gas phase, electrostatic interactions between ionized groups lead to major polarization effects that are difficult to quantify (even when using large basis sets), and are also hard to parameterize. On the other hand, these interactions are strongly screened in both proteins and in solution, and thus one obtains a far more stable set of parameters as well as more stable and reliable differences between the enzymatic and solution results.

The problematic assumption that the key to the reliability of the EVB is gas-phase calibration is in part responsible for the misleading work of, for example, ref. 54, and thus it is useful to go through this issue in some detail. That is, one of the factually incorrect criticisms made in ref. 54 was the misleading claim that our EVB approach has presumably never parameterized the off-diagonal elements (*i.e.* the H_{ij} term) by fitting them to *ab initio* calculations. However, as we have already repeatedly clarified since the first time that this allegation was made,[59] this is simply incorrect. The EVB has been parameterized to reproduce *ab initio* results starting as early as 1988 (see Fig. 3 in ref. 38), even before the important work of Chang and Miller,[60] and this was done professionally in very careful studies[61,62] long before even the gas-phase adaptation of the so called MCMM approach. This was done even earlier[63] by the Miller approach,[60] and by approaches that parameterized the whole surface and not only the reactants and products.[61] We have even reported EVB parameterization to *ab initio* results for the challenging case of the

autodissociation of water in water[35] and other studies.[62,64] Thus, stating that the EVB has not been fitted to *ab initio* calculations in simply incorrect.

Next, the authors of ref. 54 present an incorrect gas-phase EVB surface as an example of the poor performance of the EVB approach (though this has been since retracted after we pointed out their error on our website, see the Correction to ref. 54). There are unfortunately a number of problems with how this surface was generated. First and foremost, even though the EVB formulation and treatment has in fact been reproduced by several high profile workers in the field (some even without our help, see ref. 65–69 as just a few examples), in order to ensure that our approach has been properly implemented, it is essential to validate the similarity between the model being used in the new implementation and the one that was used in the original work being examined (in this particular case that of ref. 70). Such a validation should start by using the same program (which is readily available on request, and would have been to the authors had they actually asked for it), or by trying to reproduce the reported solution profile (which would require familiarity with the corresponding mapping). This is *crucial*, in particular when one is trying to discredit a given model without consulting the main author. More specifically, we have evaluated the EVB adiabatic potential energy (rather than the usual free energy) surface for the same system as examined in ref. 54 using our own MOLARIS[71] software package for the same reaction, we have obtained our EVB free energy surfaces using two different gas-phase shifts ($a = -23.0$ and $+23.0$), and the corresponding surfaces are shown in Fig. 2a and b.. Additionally, for comparison, we have evaluated the *ab initio* gas-phase energy surface using the B3LYP density functional which is shown in Fig. 2c. As seen from the figure, when we use an incorrect gas-phase shift without thorough mapping (*i.e.* mapping in all directions both backwards and forwards to check for hysteresis and other potential problems, and by using careful reaction coordinate pushing), we obtain a very bad EVB surface that is qualitatively very similar to that present in ref. 54. However, when using the correct gas-phase shift we obtain an energy surface that shows a clear S_N2 pathway, where $V^{\ddagger}$ 9 kcal mol^{-1}. It can also be seen that we obtain good agreement between the surface obtained using the correct gas-phase shift (Fig. 2b) and that obtained using *ab initio* in the region pertaining to the S_N2 reaction.

Going back to our original subject of the reliability of the EVB surfaces that were calibrated to solution information, it is useful to use as an example an issue that has been a key unresolved question in the study of the enzyme dehalogenase (DhlA), namely the barrier height in water.[72] When examining this issue, we of course tried to consider *ab initio* calculations, but found out that we are dealing with the very challenging problem of the energetics of negatively charged ions, and decided to first try to look for relevant experimental information, where we performed a careful analysis of the LFER results (see also the discussion in footnote 68 in ref. 72). Subsequently[34] (see section III.2), we performed a quantitative *ab initio* FEP calculation on this system (a challenge not yet accomplished by other workers), and established the fact that our previous estimates were indeed accurate.

III.2 The EVB as a powerful reference potential for *ab initio* QM/MM studies

In order to obtain reasonable convergence when using free energy perturbation to evaluate activation barriers, it is essential to force the system to spend significant time at the transition state.[73] When doing this, one *must* use a mapping potential that can polarize the solvent to the desired configuration, which can be effectively achieved by the EVB approach, using the mapping potential described above. Additionally, an advantage of using the EVB approach is that it includes an enormous amount of chemical information. As such, this makes the EVB ideal as a reference potential for QM/MM calculations (particularly in light of the fact that the computationally expensive mapping from the reference potential to the *ab initio* potential need only be done at the reactant and transition states), where the free energy of transferring from the EVB to the QM/MM surface is evaluated using a linear response approximation (LRA). Such an approach has been successfully applied to the study of activation barriers in solution and in proteins (*e.g.* ref. 34 and 73–75). The most effective use of the EVB as a reference potential has been implemented in the recent work of ref. 34 (see Fig. 3). This work used the EVB to successfully resolve the highly controversial question of the energetics of the reaction in solution.[72] This study actually indicated that our EVB estimate of the catalytic effect and the effect of the enzyme are quantitatively correct. Apparently, the work of Rosta *et al*. is at present the only true *ab initio* QM/MM study that considered the free energy surface of the DhlA reaction in the gas phase and solution.

III.3 The justification for the EVB diabatic representation and the off-diagonal term

The power of the EVB is largely due to its "simple" orthogonal diabatic representation and the assumption that the off-diagonal elements of the EVB Hamiltonian do not change significantly when transferring the reacting system from one phase to another. Thus, it is important to examine what the justifications are for this representation, and what has been learned about it in recent studies. A good start can be provided by the discussion of ref. 54, which argued against the use of a solvent independent H_{12} and the orthogonal diabatic representation (while overlooking the fact that the MCMM and any other EVB-type work uses orthogonal diabatic surfaces). The criticism is based on the perspective that since the so-called MOVB[76] and other *ab initio* VB approaches have a very large (and complex) overlap effect, it must be included in all diabatic treatments. This view is fundamentally flawed, and reflects a major misunderstanding about the meaning of diabatic representations. That is, the diabatic representation does not reflect absolute "reality", but is rather a powerful mathematical tool, and the best tool is the one that produces the most accurate adiabatic QM/MM free energy surface. Here, we are not aware of any careful MOVB study that has achieved this.

Fortunately, our arsenal includes the frozen DFT (FDFT) and constrained DFT (CDFT) approaches,[73,75] which are particularly useful for exploring this issue. These approaches split the system into two regions: a region comprising the solute (and any other key residues or solvent molecules, denoted region I), and the rest of the system (region II). Here, the entire system is treated by means of *ab initio* DFT, but the electron densities of the groups in region II are frozen, and the coupling between the two regions is then evaluated by means of a non-additive kinetic energy functional. Using the EVB as a reference potential in this approach has allowed for the accurate evaluation of the free energies in the condensed phase

and in proteins, while simultaneously treating the solute–solvent interaction by means of an *ab initio* quantum mechanical approach.[73]

Now the CDFT studies[77,78] involved an *ab initio* constrained DFT (CDFT) diabatic treatment of an S_N2 reaction as well as proton transfer reactions (see also recent related studies such as that in ref. 79). The CDFT approach constructed two diabatic states with formal Löwdin orthogonality, as is done in all EVB studies. These surfaces (ε_1 and ε_2) are constructed by mixing the effective off-diagonal terms, and the ground state energy (E_g) is obtained as the lowest eigenvalue of the 2×2 EVB Hamiltonian. Using both the diabatic surfaces and the adiabatic surface obtained by treating the whole reaction system with a regular DFT treatment will give us per definition the rigorous off-diagonal element by the relationship below:

$$H_{12} = \sqrt{(\varepsilon_1 - E_g)(\varepsilon_2 - E_g)} \quad (4)$$

This relationship is shown schematically in Fig. 4. The study of ref. 58 established that the EVB approximation of a solvent independent H_{12} is an excellent one (an issue that has also been established by others[80]). Thus, we have demonstrated that using formally orthogonal diabatic surfaces (where the solvation of each diabatic state is included in the diabatic energies) is an excellent approximation since it reproduces the correct solvated adiabatic QM/MM results.

Since ref. 54 claims that including the overlap integral, S_{12}, in the specification of H_{12} is a source of (additional) error, we would like to remind the reader that this is a formally orthogonalized model without "errors", since the selection of the diabatic states is the way the model is defined, and for each diabatic treatment we can have a unique H_{12} that reproduces the exact adiabatic surface. Different adiabatic states require different off-diagonal terms, and a model without overlap (which is what is also done in the so called embedded MCMM approach discussed below) has different off-diagonal terms than the one with overlap. Obviously such different models should be very different from the H_{12} with overlap.

In view of the above it is unjustified to accept the assertion[54] that using an environmentally-independent H_{12} is a source of major error. This would have possibly been a reasonable criticism if (a) the authors were not aware of the recent CDFT calculations[58] that proved that the EVB approximation of a phase-independent H_{12} is actually valid (see below), (b) the authors had any estimate of the proposed error, and (c) if the authors of ref. 54 were not unknowingly reproducing an environmentally-independent H_{12} themselves (see the discussion in ref. 5 and below). Furthermore, one would have expected a realization that H_{12} is different for different representations (*i.e.* with and without overlap). Obviously, we have explained this issue clearly since 1988.[38]

III.4 Solvation energy, energy gap, reorganization energy and free energy surfaces in solution

One of the most powerful aspects of the EVB is its ability to rigorously assess solvation effects, and to quantify key concepts such as the reorganization energy and non-equilibrium

solvation. Here, it is useful to start by addressing ref. 54, which may lead some workers to assume that the EVB treatment of solvation effects (*i.e.* eqn (3)) is somehow problematic. In fact, the view of ref. 54 probably reflects major unfamiliarity with the EVB mapping approach (for a careful discussion of this issue, see ref. 81).

Here, it is important to realize that the EVB US-FEP mapping is a particularly effective mapping that uses the energy gap as a generalized reaction coordinate. This guarantees accelerated convergence for processes in condensed phases, since it captures the main physics of the solvent response, and is the most physically accurate way of moving from the reactant to the product state. The effectiveness of the EVB mapping approach is now widely appreciated by some scientific communities (for example, see ref. 65–69).

Another key issue is the ability of the EVB to accurately and effectively capture the reorganization energy, λ. Here, it is useful to start with the problematic assertion of ref. 54, which defined the reorganization energy as "the energy difference between the product and reactant diabatic states at the reactant geometry". This definition (which perhaps stems from unfamiliarity with the reorganization concept and from not using the EVB or related strategy for this purpose) is simply inappropriate. In fact, the reorganization energy can be estimated by use of the expression:

$$\lambda = 0.5 \left(\langle \Delta\varepsilon \rangle_b - \langle \Delta\varepsilon \rangle_a = \langle \Delta\varepsilon \rangle_a + \Delta G \right) \quad (5)$$

where $\langle \varepsilon \rangle$ is the difference between ε_a and ε_b and $\langle \delta\varepsilon \rangle_a$ designates average over trajectories on ε_a, or the value obtained by evaluating the microscopic Marcus parabolae, which were evaluated professionally in our works in a way used now by all workers in the field (*e.g.* ref. 67–69, amongst others). This has little to do with the evaluation of the energy along the least energy MOVB path shown in, for example, ref. 54. Furthermore, calculations of λ by MOVB will be meaningless since it should be defined for pure non-mixed diabatic states. It is also an irrelevant point, as the real issue is the change in the charges of the diabatic states.

The EVB provides an extremely powerful way of consistently capturing solvent effects. In discussing this issue it is again useful to consider the unjustified criticism of ref. 54 as a "primer" for this discussion. We are told that the EVB keeps constant charges, and thus that the diabatic charges do not change along the reaction coordinate which is claimed to be a really major problem. However, this is simply incorrect, since many EVB studies (including the paper that introduced the EVB) have considered the polarization of the diabatic state (equation 29 in ref. 82). The same was done in our 1988 papers (see equation 11 in ref. 38 and equation 7 in ref. 83), and even in our book,[37] where Chapter 3 describes gas-phase surfaces for S_N2 reactions with polarization of the diabatic states.

In an analysis of the solvated EVB, it was argued that including solvation effects in the diagonal and calibrating on the isolated fragment in solution "are mutually incompatible". This shows a major unfamiliarity with diabatic concepts. Of course, the isolated fragments have the same solvation as the diabatic states, since the mixing term has no effect when the fragments are at infinite separation. The EVB diabatic state at infinite separation is the isolated fragment, and it is identical to the adiabatic state. This problem is compounded by

hiding the fact that the embedded MCMM also includes the solvation effect in the diagonal element. Of course, the EVB puts the solvent into the diagonal Hamiltonian. Trying to solvate the adiabatic charges rather than the diabatic states leads to bad physics, particularly in the case of S_N1-type and proton transfer reactions (the problems with S_N1 reactions were already clarified in the 1988 paper).

It has been stated[54] that calculations using the EVB surface (presumably those of ref. 54) "have been unable to reproduce the free energy of activation obtained by molecular dynamics simulations (24.9 kcal mol^{-1}). These findings cast doubts on the reported EVB results and suggest that a careful parameterization of the gas-phase reaction might have been useful in order to obtain more meaningful results for the condensed-phase reaction by the EVB method". However, this is an incorrect allegation, as the presumed calculations of the EVB solvation energies (see Table 4 of ref. 54) are incorrect, and have no similarity at all to the values reported in our paper,[72] or to the values expected from any reasonable solvation model. Additionally, to claim that the EVB gives more solvation in the TS than in the ground state in an S_N2 reaction, where this was never obtained in any EVB calculations by us or any other EVB user, is quite worrying. In any case, comparing the results shown in Table 4 of ref. 54 to our previous work should quickly demonstrate to the reader that the model used in ref. 54 has *nothing* to do with the solvation model used by the EVB calculations, or by proper microscopic free energy calculations.

At any rate, in order to clarify the misleading nature of the allegation that the EVB solution results are irreproducible, we have performed our EVB runs in both the gasphase and in solution using the same parameters (*i.e.* those presented in the supplementary material of ref. 70, obtained with the same MOLARIS program). The corresponding energy and free energy surfaces obtained using both $\alpha = -23.0$ (incorrect) and $+23.0$ (correct) are shown in Fig. 2a and b, respectively. In order for our results (both in the gas phase and in solution) to be reproducible to any interested reader, we have made all the required input files, our output files, a guest binary for the MOLARIS software package, and instructions on how to run the calculations available on our website (http://futura.usc.edu), under the section "EVB Verification".

IV. The EVB as a basis for evaluating catalytic effects

In recent years, the EVB has been used in studies of many key enzymatic reactions, and has provided what is, at present, the most effective way to analyze the catalytic effects of enzymes. An extensive review of the effectiveness of EVB studies has already been provided in ref. 2, and here we will only mention the most recent major advances in the field.

As an example, a popular belief is that strain contributes significantly to catalysis (in that the enzyme destabilizes the ground state of the reacting system, thus reducing the activation barrier for the chemical step),[84,85] and one of the last bastions of this belief has been the catalytic power of enzymes that contain the coenzyme B_{12} cofactor. One advantage of the EVB approach is that it overcomes the sampling difficulties that are inherent to energy minimization studies, and a recent study of the strain hypothesis in enzymes that do contain

the coenzyme B_{12} cofactor[86] demonstrated that the major part of the catalytic effect is not actually due to strain but, rather, it is due to the electrostatic interaction between the ribose and the protein and the strain contribution is in fact very small. Thus, it appears that enzymes can use electrostatic effects even in a radical process, where the charge distribution of the reacting fragments is more or less constant during the reaction, by attaching a polar group to the leaving fragment and designing an active site that interacts more strongly with this group in the product state rather than in the reactant state. The fact that evolution has had to use this trick provides further evidence that it is extremely hard to catalyze enzymatic reactions by non-electrostatic factors, a finding that can be manipulated in enzyme design (see below for more on computer-aided enzyme design).

The EVB has also been used by Åqvist and coworkers[87] to study cold-adapted enzymes. These workers performed extensive computer simulations of the ketoenol(ate) isomerization steps in differently adapted citrate synthases,[87] and reproduced the absolute rates of both the psychrophilic and mesophilic enzymes at 300 K, as well as both the lower enthalpy and more negative activation entropy of the cold-adapted enzyme. Here, it was found that the overall catalytic effect is stemming from the electrostatic stabilization of the transition state and enolate, as well as a reduction in the re-organization free energies. The lower activation enthalpy and more negative activation entropy observed for the cold-adapted enzymes was found to be associated with decreased protein stiffness that is not localized to the active site, but rather spread over several regions of the protein structure. The main point for the purpose of the present work is that the extensive entropy calculations would have been simply impossible without the effectives of the EVB.

A related EVB study that carefully explored the relationship between thermal stability and catalysis was also reported in ref. 88. This study explored the idea that the fact that thermophiles are less flexible than mesophiles (and also less active) proves that flexibility is related to catalysis. In fact, it was shown that while the free energy surface is shallower in the mesophiles, the surface along the chemical coordinate has similar curvature in both cases, and the difference in reactivity is entirely due to the difference in the chemical reorganization energy.

The EVB approach has also been used to elucidate the relationship between the folding landscape of enzymes and their catalytic power,[89] by using a simplified folding model in order to generate the free energy landscape of an enzyme, and subsequently using this to evaluate the activation barriers for the chemical step in the different regions of the landscape. The aim of this study was to investigate experimental findings[90,91] that an engineered monomeric chorismate mutase (CM) has a catalytic efficiency that is similar to that of the naturally occurring dimer, even though it has the structural properties of an intrinsically disordered molten globule. Examining the landscape demonstrated that the CM monomer (that behaves like a molten globule in the absence of the substrate) has low activation barriers even in regions that are not exactly at the native configuration.

Finally, the ability to design effective enzymes presents one of the most fundamental challenges in biotechnology, and such ability would be one of the most convincing manifestations of a complete understanding of the origin of enzyme catalysis. A recent

study[92] explored the reliability of different simulation approaches in terms of their ability to rank different possible active site constructs, and demonstrated that the EVB model serves as a practical and reliable tool in the final stages of computer-aided enzyme design. Also, other approaches examined in this work were found to be comparatively less accurate, and mainly useful for the qualitative screening of ionized residues. Part of the problem arises from the fact that current design approaches[93–111] are not based on modeling the chemical process in the enzyme active site. In fact some approaches (*e.g.* ref. 93) use gas-phase or small cluster calculations which then estimate the interaction between the enzyme and the transition state model, rather than the transition state-binding free energy (or the relevant activation free energy), and the accurate ranking of the different options for enzyme design cannot be accomplished by approaches that cannot capture the electrostatic preorganization effect. Clearly, the ability of the EVB model to act as a quantitative tool in the final stages of computer-aided enzyme design is a major step towards the design of enzymes whose catalytic power is closer to native enzymes than the current generation of designer enzymes.

V. LFER in solutions and enzymes

One of the most important concepts in physical organic chemistry is the so-called linear free energy relationship (LFER), which goes back to the early work of Hammett[112–114] (who quantified the effect of *meta* and *para* benzene substituents on ester hydrolysis), which was subsequently modified by Taft[115–119] (following Ingold) to also describe the steric effects of a substituent. Despite the popularity and power of this concept, it had not been quantified by computational studies of solution reactions until the emergence of the EVB free energy functionals (the g of eqn (3)) that were originally developed in ref. 39 in order to provide the microscopic equivalent of Marcus' theory for electron transfer reactions.[40] This approach allows one to explore the validity of the Marcus formula and the underlying linear response approximation on a microscopic molecular level.[120] While this point is now widely accepted by the electron transfer (ET)[121] community, the validity of the EVB as perhaps the most general tool in microscopic LFER studies is less appreciated. This issue will be addressed below.

In order to explore the molecular basis of LFER, we will consider a one-step chemical reaction and describe this reaction in terms of two diabatic studies that correspond to the reactant and product states. In this case the ground state adiabatic surface is given by:

$$E_g = \frac{1}{2}\left[(\varepsilon_1 + \varepsilon_2) - \sqrt{(\varepsilon_1 - \varepsilon_2)^2 + 4H_{12}^2}\right] \quad (6)$$

With this well-defined adiabatic surface, we can explore the correlation between $g^{\ddagger}$ and G°. Now the EVB-umbrella sampling procedure (*e.g.* ref. 37) allows one to obtain the rigorous profile of the free energy function, $\Delta \bar{g}$, that corresponds to E_g, and the free energy functions g_1 and g_2, that correspond to ε_1 and ε_2, respectively (see Fig. 5). It is important to point out here that such profiles have been quantitatively evaluated in many EVB simulations of chemical reactions in solutions and proteins (for reviews, see ref. 36 and 122). The corresponding profiles provide the activation free energy $g^{\ddagger}$ for the given

chemical step. The calculated activation barrier can then be converted (for example, see ref. 37) to the corresponding rate constant using transition state theory (TST):

$$k_{i\to j} \cong (RT/h)\, exp\left\{-\Delta g_{i\to j}^{\ddagger}/RT\right\} \quad (7)$$

A more rigorous expression for $k_{i\to j}$ can be obtained by multiplying the TST expression by a transmission factor that can be easily calculated by running downhill trajectories.[37] However, the corresponding correction is usually small.[51,123] At this point, it is useful to consider the approximated expressions for $\Delta \bar{g}$ and $g^{\ddagger}$. Here we note that, with the simple two-state model of eqn (6), we can obtain a very useful approximation to the $\Delta \bar{g}$ curve. That is, using the aforementioned free energy EVB-umbrella sampling formulation yields the $\Delta \bar{g}$ that corresponds to the E_g and the free energy functions g_i that correspond to the ε_i surfaces. This leads to the approximated expression:

$$\Delta \bar{g}(x) = \frac{1}{2}\left[(\Delta g_1(x) + \Delta g_2(x)) - \sqrt{(\Delta g_1(x) - \Delta g_2(x))^2 + 4H_{12}^2(x)}\right] \quad (8)$$

This relationship can be verified in the case of small H_{12} by considering our ET studies,[120] while for larger H_{12} one should use a perturbation treatment. Now we can exploit the fact that the g_i curves can be approximated by parabolae of equal curvature (this approximated relationship was found to be valid by many microscopic simulations (*e.g.* ref. 36). This approximation can be expressed as:

$$\Delta g_i(x) = \lambda\left(\frac{x - x_\circ^{(i)}}{x_\circ^{(j)} - x_\circ^{(i)}}\right)^2 \quad (9)$$

where λ is the so-called "solvent reorganization energy" (which is illustrated in Fig. 5).

From eqn (8) and (9), one obtains the Hwang–Åqvist–Warshel (HAW) equation, which is given in the general case by:

$$\Delta g_{i\to j}^{\ddagger} = (\Delta G^{\circ}{}_{i\to j} + \lambda_{i\to j})^2/4\lambda - H_{ij}\left(x^{\ddagger}\right) + H_{ij}^2\left(x_0^{(i)}\right)/(\Delta G^{\circ}{}_{i\to j} + \lambda_{i\to j}) + \Gamma_{ij} \quad (10)$$

where $G^{\circ}{}_{i\to j}$ is the free energy of the reaction and H_{ij} is the off-diagonal term that mixes the two relevant states with average value at the transition state, $x^{\ddagger}$, and at the reactant state, $x_0^{(j)} \cdot \Gamma_{ij}$ is a correction that reflects the effect of tunneling and zero point energy corrections in cases of light atom transfer reactions.

Repeated quantitative EVB studies of reactions in solutions and proteins (*e.g.* ref. 36 and 124) established the quantitative validity of eqn (10). With this fact in mind, we can take these equations as a quantitative correlation between $\Delta g_{i\to j}^{\ddagger}$ and G°. Basically, when the changes in G° are small, we obtain a linear relationship between $\Delta g_{i\to j}^{\ddagger}$ and G°. This

linear relationship, which can be obtained by simply differentiating the $\Delta g^{\ddagger}_{i \to j}$ of eqn (10) with respect to $G^{\circ}_{i \to j}$, can be expressed in the form:

$$\Delta\Delta g^{\ddagger}_{i \to j} = \theta \Delta\Delta G^{\circ}_{i \to j} \quad (11)$$

where $\theta = (G^{\circ}_{i \to} + \lambda)/2\lambda$, and where the contribution from the last term of eqn (10) is neglected. The linear correlation coefficient depends on the magnitude of G° and λ. At any rate, more details about this linear free energy relationship (LFER) or free energy relationship (FER) and its performance in studies of chemical and biochemical problems are given elsewhere.[36–38,125–127]

The main point of eqn (11) and Fig. 5 is that $G^{\circ}_{i \to j}$, which determines the corresponding $\Delta g^{\ddagger}_{i \to j}$, is correlated with the difference between the two minima of the $\Delta \bar{g}$ profile that correspond to states *i* and *j*.

While our ability to reproduce the observed LFER might not look like a conceptual advance, the fact that the EVB provides a rigorous basis for FER in condensed phases leads to a different picture than that which has been assumed in traditional LFER studies. That is, as is clear from the HAW relationship, it is essential to take into account the effect of H_{ij} in LFER studies that involve actual chemical reactions (rather than ET reactions). In such cases, H_{ij} is frequently very significant and its neglect leads to an incorrect estimate of the relevant reorganization energy. This point has not been widely appreciated due to the fact that the correlation between $\Delta \bar{g}$ and $g^{\ddagger}$ does not depend so critically on H_{ij}. Thus, as long as one fits the experimentally observed relationship by phenomenological parameters, it is hard to realize that the relevant reorganization energies are underestimated in a drastic way. A case in point is the systematic analysis of hydride transfer reactions by Kong and Warshel,[125] and S$_N$2 reactions,[38] which are summarized in Fig. 6.

The use of the EVB and eqn (10) in studies of reactions in solutions has been extended to studies of LFER in enzymes. The successes of this approach have been demonstrated in studies of carbonic anhydrase,[42] p21 Ras,[126,128] tyrosyl-tRNA synthetase,[127] aspartic proteases[129] and DNA polymerase β.[130,131] At present, we view these studies as the most quantitative LFER studies of enzymes. It is also useful to point out the successes of our approach in LFER studies of electron transport in proteins (*e.g.* ref. 121 and 132).

A recent attempt to derive an LFER for PT reactions has been reported by Kiefer and Hynes,[133] who basically used an EVB formulation, with a continuum treatment of the solvent. Unfortunately, they assumed that a "Marcus relation was never actually derived for PT reactions", apparently choosing to ignore all the works above. Furthermore, their derivation ignored the crucial effect of H_{ij}. Nevertheless, it is encouraging to see the effectiveness of the EVB in providing a molecular basis for LFER treatments again.

VI. Proton transport in carbonic anhydrase as an example of the difference between microscopic and phenomenological LFER

In order to illustrate our point about the difficulties associated with phenomenological LFER treatments of reactions in solutions and in enzymes, it is instructive to consider the studies of human carbonic anhydrase III (which will be referred to here as CA III).[134] Studies of this system[134,135] demonstrated that the rate of PT in mutants of CA III is correlated with the pK_a difference between the donor and acceptor. It was found that the observed LFER follows a Marcus' type relationship. Although this study provided an excellent benchmark for studies of PT in proteins, it also raised the question about uniqueness of the parameters deduced from phenomenological LFER studies. This issue will be explored below.

The catalytic reaction of CA III can be described in terms of two steps. The first is attack of a zinc-bound hydroxide on CO_2[136]

$$CO_2 + EZn^{+2}\left(OH^-\right) + H_2O \leftrightarrows HCO^-_3 + EZn^{+2}\left(H_2O\right) \quad (12)$$
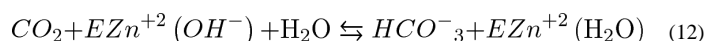
The reversal of this reaction is called the "dehydration step". The second step involves the regeneration of the OH^- by a series of PT steps:[137,138]

$$EZn^{+2}\left(H_2O\right) + B \underset{k_B}{\overset{k_{-B}}{\rightleftharpoons}} EZn^{+2}\left(HO^-\right) + BH^+ \quad (13)$$
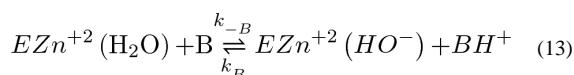
where $K_{-B} = K_{-B}/k_B$ (in the notation of ref. 134) and BH^+ can be water, buffer in solution or the protonated form of Lys64 (other CAs have His in position 64). Previous experimental studies[134] have established a LFER that was fitted to Marcus' equation using:

$$\Delta g^{\ddagger} = w^r + \left\{1 + \Delta G^{\circ}/4\Delta G_0^{\ddagger}\right\}^2 \Delta G_0^{\ddagger} \quad (14)$$

where the observed reaction free energy is given by G°_{obs} w^r + G° − w^p, where w^r is the work of bringing the reactants to their reacting configuration and w^p is the corresponding work for the reverse reaction. G° is the free energy of the reaction when the donor and acceptor are at their optimal distance. $\Delta G_0^{\ddagger}$ is the so-called intrinsic activation barrier which is actually ¼ of the corresponding reorganization energy, λ. Here we use $g^{\ddagger}$ rather than $G^{\ddagger}$ for the activation barrier following the consideration of ref. 37. Eqn (13) can also be written in the well-known form:

$$\Delta g^{\ddagger} = w^r + (\Delta G^{\circ} + \lambda)^2/4\lambda \quad (15)$$

The phenomenological fitting processes yielded $\lambda = 5.6$ kcal mol^{-1} and $w^r \approx 10.0$ kcal mol^{-1}. The estimated value of λ appears to be in conflict with the value deduced from computer simulation studies ($\lambda \approx 80$ kcal mol^{-1} in ref. 42). Furthermore, the large value of w^r is hard to rationalize since the reaction involves a proton transfer between a relatively fixed donor and acceptor (residue 64 and the zinc bound hydroxide). The very small values of λ obtained by fitting eqn (13) to experiment are not exclusive to CA III. Similarly, small

values were obtained in analysis of other enzymes and are drastically different from the values obtained by actual microscopic computer simulations (note in this respect that λ cannot be measured directly).

As pointed out before,[139–141] the above discrepancies reflect the following problems: first, the reaction under study may involve more than two intersecting parabolae and thus cannot be described by eqn (14). Second, although eqn (14) gives a proper description for electron transfer (ET) reactions where the mixing between the reactant and product state (H_{12}) is small, it cannot be used for describing proton transfer or other bond breaking reactions where H_{12} is large. In such cases, one should use the HAW expression (see Section V).

In order to obtain a proper molecular description of LFER, it is essential to represent each reactant, product or intermediate by a parabolic free energy function.[37] In the case of CA III, we describe the proton transfer from residue 64 (Lys or His) to the zinc-bound hydroxyl *via* a bridging water molecule (and alternatively two water molecules), by considering the three states:

$$\psi_1 = BH^+(H_2O)_b(OH^-)_a Zn^{+2}$$
$$\psi_2 = B(H_3O^+)_b(OH^-)_a Zn^{+2} \quad (16)$$
$$\psi_3 = B(H_2O)_b(H_2O)_a Zn^{+2}$$

where we denote the base at residue 64 by "B", and where ψ_1 and ψ_2 correspond, respectively, to the right and left sides of eqn (13) (see also Fig. 7 for a graphic description of the three states). The relative free energy of these states can be estimated from the corresponding pK_as, where the pK_as of $(H_2O)_a$ and B are known from different mutations,[134] and the pK_a of $(H_2O)_b$ can be calculated by the PDLD/S-LRA approach. Note that our three state systems can be easily extended to include one more water molecule and one more state.

In order to set out the problem in a clear way for further considerations, we show the potential surface for the wild type enzyme using pK_as of 5.0 and 9.0 for $(H_2O)_a$ and Lys64 (taken from ref. 134) and p$K_a = 0$ for $(H_2O)_b$, evaluated by the PDLD/S-LRA method as described in the previous section, in Fig. 8.

With the model of eqn (16) and with a reasonable estimate of the free energies G°_{12} and G°_{23}, we can start to evaluate the apparent activation barrier. Before doing so, we must clarify several points: (i) a Marcus-type relationship and the corresponding LFER is *only* valid for a two state system (1 → 2), *i.e.* for a reaction with a single step. We have, however, a three state process that involves a two-step mechanism (1 → 2 → 3). Fitting such a system to a Marcus type formula can lead to non-physical parameters (*e.g.* a value for λ that is too small). (ii) In order to use the HAW approach in a three state system (or in a four state system), we must consider the elementary rate constants and then consider the pre-equilibrium concentrations. That is, for the reaction:

$$1 \underset{}{\overset{k_{12}}{\rightleftarrows}} 2 \underset{}{\overset{k_{23}}{\rightleftarrows}} 3 \quad (17)$$
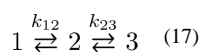
where $K_{12} = k_{12}/k_{12}$ involves the forward and backward rate constants k_{12} and k_{21}, while $K_{23} = k_{23}/k_{32}$ involves the rate constants k_{23} and k_{32} (note that k_{13} corresponds to k_{-B} in eqn (13)), we can fit the HAW equation to $\log k_{13}$ (or $\log k_B$), but if we want to obtain the LFER for the $1 \rightarrow 3$ step we must take the free energy of state 2 into account. After some manipulations,[142] we obtain:

$$
\begin{aligned}
\Delta g_{13}^{\ddagger} &\cong \Delta G^{\circ}{}_{12} + \left[(\Delta G^{\circ}{}_{23} + \lambda_{23})^2/4\lambda_{23} - H_{23} \right] + w \\
&= \Delta G^{\circ}{}_{12} + \left(\Delta_{23}^{\ddagger} \right)_{di} - H_{23} + w
\end{aligned}
\tag{18}
$$

If the rate limiting step is $1 \rightarrow 2$, then k_{12} determines k_{13} and we can write:

$$
\begin{aligned}
\Delta g_{13}^{\ddagger} &= \Delta g_{12}^{\ddagger} = (\Delta G^{\circ}{}_{12} + \lambda_{12})^2/4\lambda_{12} - H_{12} + w \\
&= \left(\Delta g_{12}^{\ddagger} \right)_{di} - H_{12} + w
\end{aligned}
\tag{19}
$$

Here, we neglected the $(H_{12})^2$ term in eqn (10) for convenience, since the corresponding correction is small in our case. Note also that the subscript "di" here denotes "diabatic".

In light of the complexity of eqn (18) and (19), we find it more convenient to try to reproduce the observed LFER by the direct evaluation of these equations, using the relevant calculated and observed parameters, with an only minimal adjustment procedure. Our starting point is the estimate of $G^{\circ}{}_{12}$ and $DG^{\circ}{}_{23}$ for the different mutants studied in ref. 134. The relevant free energy values were obtained using the observed pK_as of $(H_2O)_a$ and B_{64} and the calculated pK_a of $(H_2O)_b$. We used the value $\lambda_{23} \approx 80$ kcal mol^{-1} obtained from EVB simulations[42] as a generic value for both λ_{12} and λ_{23}. The value of $\bar{H}_{ij}\left(x^{\ddagger}\right)$ was taken to be 18 kcal mol^{-1} (which reflects a minor adjustment from the value found in ref. 42) and $\bar{H}_{ij}\left(x_0\right)$ was taken to be 10 kcal mol^{-1} from an EVB study of a proton transfer from $(H_2O)_a$ to $(H_2O)_b$. With the estimated G_0, λ, and $\bar{H}_{ij}$ we evaluated the diabatic free energy functions and used them to obtain the corresponding $(g^{\ddagger})_{di}$ and $\delta g^{\ddagger}$ for several mutants. The corresponding graphical analysis is given in Fig. 9a–c, and the resulting dependence of $g^{\ddagger}$ on G_{13} in Table 1. As seen from the table and the figures, our model reproduced the observed trend in a satisfactory way. However, the origin of the trend is very different to that deduced from the two state Marcus equation. That is, the flattening of the LFER at $pK_a > 0$, which would be considered in a phenomenological analysis of a two state model as the beginning of the Marcus inverted region (where $G_0 = -\lambda$), is due to the behavior of the three state system (see ref. 142).

The extraction of λ from fitting eqn (14) to the observed LFER requires that $\lambda = -G^{\circ}$ so that $G^{\circ} < 0$ in the point where the LFER becomes flat. This means that we must have data from regions where $G^{\circ} < 0$. However, at least for the cases when $\Delta g_{12}^{\ddagger}$ is rate limiting, $G^{\circ}{}_{12}$ cannot be negative and the observation of a beginning of a flat LFER is actually due to other factors. It is also important to realize that λ_{out} cannot become too small and never approaches zero, which is the continuum limit for a completely non-polar environment (see discussion in ref. 143). The reason is quite simple; the protein cannot use a non-polar environment, since this will drastically decrease the pK_a of $(H_2O_b{}^+)$. Instead, proteins use

polar environments with partially fixed dipoles. However, no protein can keep its dipoles completely fixed (the protein is flexible), and this thus gives a non-negligible λ_{out} (where "out" denotes "outer sphere"). Of course, this reorganization energy is still smaller than the corresponding value for proton transfer in solutions but it never approaches the low value obtained from the fitting in a two state Marcus' formula,[144] without considering the effect of the off-diagonal element.

As long as we obtain the value of H_{ij} from fitting to the observed LFER, it is possible to argue that both eqn (13) and (10) reflect a phenomenological fitting with a free parameter (w and λ in the case of eqn (13) and H_{ij} in the case of eqn (10)). The difference, however, is that eqn (10) and the use of three free energy functionals reflects much more realistic physics. This is evident, for example, from the fact that with eqn (10) we do not obtain an unrealistically large w^r. Note in this respect that ΔG°_{12} in eqn (18) might look like w in a phenomenological fitting to Marcus' equation. Of course, we can now obtain the H_{ij} in the protein by the FDFT approach, as was discussed in section III.3, but this is not essential for the present discussion. It is also important to emphasize at this point that the present treatment is not a phenomenological treatment with many free parameters as might be concluded by those who are unfamiliar with molecular simulations. That is, our approach is based on realistic molecular parameters obtained while starting from the X-ray structure of the protein and reproducing the relevant pK_as and reorganization energy. Reproducing the observed LFER by such an approach without adjusting the key parameters is fundamentally different to an approach that takes the observed LFER and adjusts free parameters in a given model to reproduce it. In such a case, one could reproduce any experiment by almost any model.

In order to further explore the validity of the stepwise modified Marcus' model, we recently developed[145] a simplified EVB model which represents the given conduction chain by explicit EVB, while representing the rest of the environment (protein and solvent) implicitly. The implicit treatment forces the minima of the free energy parabolae of the simplified model to coincide with those of the full model. The dynamics of the system are then studied by Langevin dynamics simulations. The simulations established that the rate of the PTR process is determined by the energetics of the proton along the conduction chain, once the energy of the proton in the two successive sites is significantly higher than the energy of the proton in the bulk water. The model was also applied to PTR in the K64H-F198D mutant of CA III and reproduced the observed rate constant. Typical simulations for the case where the energy of the proton on His64 is raised by 1.2 kcal mol^{-1} in order to accelerate the calculations are described in Fig. 10. The calculated average time for PTR from His64 to the Zn-bound hydroxide is about 5×10^{-6} s. Correcting this result for the energy shift and the effect of using an overdamped rather than underdamped BD simulation gives a result that is close to the observed k_B($k_B = 3 \times 10^{-6}$ s). The simulation provides additional major support for the use of the=HAW model.

We would also like to address the validity of the general use of eqn (10) and the multi-state procedure used for studies of the proton transport (PT) in CA. The use of eqn (10) for subsequent PT steps might look to some as an *ad hoc* approach, considering the assumption that PTR processes involve the Grotthuss mechanism which is not sensitive to the ΔG°_{ij}

values for the sequential transfer process. However, the assumption that the Grotthuss mechanism is a key factor recently underwent a major paradigm shift, where those who supported this idea started to attribute major importance to the electrostatic barrier,[146] in agreement with our early view.[147–149] A major part of this realization has been "forced" on workers in the field by the analysis of the blockage of PT in aquaporin (for a review, see ref. 149). The aquaporin study also established that the delocalization effects are relatively small, and that in cases of a high barrier in the electrostatic profile, it is possible to use this profile with a small modification instead of the EVB profile (see the validation studies in Fig. 8 of ref. 149). Interestingly, the idea that the electrostatic profile is the key factor has also been recently independently reached by other workers.[150] The close relationship between the full EVB free energy profile and the corresponding electrostatic profile has allowed for major progress in studies of exceedingly complex PT processes in biology. In particular, we have been able to explore the long-time pumping mechanism in cytocrome c oxidase (CcO) while using Langevin dynamics and even Monte Carlo approaches,[151] as well as by careful EVB free energy calculations.[152] Our studies are at present the only studies that actually explore the relevant PT barriers rather than just the energetics of rotating water molecules or protein groups, or the energetics of transferring protons between water molecules without considering the transfer to and from the protein groups (see the discussion in ref. 152). All of this has been mainly possible due to the development of the EVB approach. Finally, combining EVB with a QM/MM approach has allowed us to explore the primary PT in bacteriorhodopsin and its relationship to the change of the protein conformational changes.[153]

VII. Using the EVB to assess the protein reorganization energy and the preorganization concept

The previous sections established the validity of the EVB as a general tool for analyzing activation barriers in enzymes and for correlating them with different factors. Since the HAW equation reproduces the EVB trend, we may now ask which parameters contribute to the catalytic effect, and what the molecular origin of the changes in these parameters actually is. As shown in Fig. 11, the enzyme can reduce $g^{\ddagger}$ by reducing G° or λ in the Marcus formula or in the HAW equation (where it can also increase H_{12}). In fact, the question of whether the catalysis is due to the reduction of G° or λ has attracted significant interest.[144,154] Unfortunately, the real question is what the molecular origin of the reduction of G and λ actually *is*. Superficial considerations seem to suggest an almost trivial rationale for the reduction of λ and the corresponding catalytic effect. All that is needed, according to the dielectric continuum theory, is a reduction in the dielectric constant, and thus it is tempting to suggest that the protein reduces λ by providing a non-polar active site and therefore a low dielectric constant.[154,155] Unfortunately, non-polar active sites increase rather than decrease the energy of polar transition states (relative to the corresponding energy in water). Similarly, non-polar active sites do not help reduce the energy of charge transfer reactions where the RS is more polar than the TS, despite the reduction in reorganization energy (see Fig. 11). Apparently, enzyme active sites are polar rather than non-polar and the origin of the reduction of $G^{\ddagger}$ must be more complex.

Our starting point in exploring the origin of the reduction in $g^{\ddagger}$ is what was learned from computer simulation of enzymatic reactions. Such studies repeatedly showed that the difference between $\Delta g^{\ddagger}_{cat}$ and $\Delta g^{\ddagger}_{cage}$ is mainly due to electrostatic effects. The calculations indicated that enzymes "solvate" their TSs more than the corresponding TSs in the reference solution reactions.[2,156] The nature of this "solvation" effect appeared to be far from obvious. That is, the calculated interaction energy between the TS charges of the reacting atoms and the enzyme were found to be similar to the corresponding interaction energies in solution. This finding indicated that the "strength" of the interactions could not explain the catalytic effect. However, since the calculated electrostatic energy does account for the catalytic effect, it is clear that we must examine the entire electrostatic energy associated with the formation of the TS[37] rather than only the interaction energy at the TS. This includes the penalty for the reorganization of the environment upon "charging" the TS.

In order to analyze and quantify the overall electrostatic contribution associated with the binding of the TS, it is useful to evaluate the free energy of forming the TS charges in the enzyme and in solution. Performing such calculations by a FEP approach is very demanding, but fortunately they can be conveniently approximated by using the linear response approximation (LRA)[157] shown in eqn (20):

$$
\begin{aligned}
\Delta G\left(Q^{\ddagger}\right) = 0.5 &\left(\left\langle U\left(Q{=}Q^{\ddagger}\right) - U\left(Q{=}0\right) \right\rangle_{Q-Q^{\ddagger}} \right.\\
&\left. + \left\langle U\left(Q{=}Q^{\ddagger}\right) - U\left(Q{=}0\right) \right\rangle_{Q-0} \right) = 0.5\left(\left\langle \Delta U \right\rangle_{Q-Q^{\ddagger}} + \left\langle \Delta U \right\rangle_{Q-0} \right)
\end{aligned}
\tag{20}
$$

where U is the solute–solvent interaction potential, Q designates the residual charges of the solute atoms with $Q^{\ddagger}$ indicating the TS charges and $<U>_Q$ designates an average over configurations obtained from an MD run with the given solute charge distribution. The first term in eqn (20) is the aforementioned interaction energy at the TS, where $Q = Q^{\ddagger}$, which is similar in the enzyme and in solution. The second term expresses the effect of the preorganization of the environment. If the environment is randomly oriented toward the TS in the absence of a charge (as is the case in water), then the second term is zero and we obtain:

$$
\Delta G\left(Q^{\ddagger}\right)_{\mathrm{w}} = 1/2 \left\langle \Delta U \right\rangle Q^{\ddagger}
\tag{21}
$$

where the electrostatic free energy is half of the average electrostatic potential.[140] However, in the preorganized environment of an enzyme, we obtain a significant contribution from the second term and the overall $G(Q^{\ddagger})$ is more negative than in water. This extra stabilization is responsible for the catalytic effect of the enzyme.

Another way to look at the above TS stabilization is to realize that in water (where the solvent dipoles are randomly oriented around the uncharged form of the TS), the activation free energy includes the free energy needed to reorganize the solvent dipoles towards the charged TS. On the other hand, the reaction in the protein costs less reorganization energy since the active site dipoles (associated with polar groups, charged groups and water

molecules) are already partially preorganized toward the TS charges,[37,156] although this is associated with polar rather than non-polar environments.

It is important to note that the reorganization energy associated with the binding of the transition state charges is given by:

$$\lambda^{\ddagger}=0.5\left(\langle\Delta U\rangle_{Q=Q^{\ddagger}} - \langle\Delta U\rangle_{Q=0}\right) \quad (22)$$

This reorganization energy is related to the well-known Marcus reorganization energy, but it is not equal to it. More specifically, the Marcus reorganization energy[158] is related to the transfer from the reactant to the product state, while here we deal with charging the TS (see the discussion in ref. 51). Nevertheless, calculations of the Marcus reorganization energy in enzymes and in solution are also consistent with the above idea, and it has been repeatedly found that λ_p is smaller than λ_w (see ref. 51 and 159–161), although this is associated with polar rather than non-polar environments.

It is also useful to point out that the nature of the reorganization energies and the way to evaluate this quantity is still not widely recognized in the computational chemistry community (see also section III.4). This is to a large extent due to the unfamiliarity with the EVB-based evaluation of the reorganization energy and the difficulties associated with obtaining this quantity by standard molecular orbital approaches. Some workers still assume (based on experimental analysis and *ab initio* calculations) that the reorganization energy is very small in contrast to the EVB results of ref. 138. The reason for this is associated with the assumption that the reorganization energy can be obtained from the $G^{\ddagger}$ of eqn (14) using the relationship $G^{\ddagger} = \lambda/4$. Unfortunately, $G^{\ddagger}$ (the so-called intrinsic barrier) corresponds to the adiabatic barrier, which reflects the effect of H_{12}. A proper calculation of λ by using the functional of eqn (3) gives unique and stable results with large λ_p.[138]

Perhaps one of the best examples of the crucial role of the preorganizaton effect has emerged from recent studies of ketosteroid isomerase (KSI). A recent study of the binding of different phenolate transition state analogues (TSAs) to ketosteroid isomerase (KSI)[162] found a small change in the binding energy, accompanied by a change in the charge delocalization of the TSAs, which was taken as proof that electrostatic effects do not make a major contribution to catalysis. However, an in-depth EVB study[163] of the relationship between the binding of the TSAs and the chemical catalysis by KSI, as well as the binding of the TS not only reproduced all relevant experimental results (which can be used to quantify the different contributions to the observed effects), but also demonstrated that the binding of the TSAs and chemical catalysis represent different thermodynamic cycles, with very different electrostatic contributions. That is, while the binding of the TSA involves a small electrostatic contribution, chemical catalysis involves a charge transfer process with major electrostatic contribution, due to the preorganization of the active site. It was also found that electrostatic preorganization contributions to the binding of the enolate intermediate of KSI and the TS are much larger than the corresponding effect for the binding of the TSAs. This reflects the dependence of the preorganization on the orientation of the non-polar form of the TSAs, relative to the oxyanion hole. This work not only demonstrated the need for computational studies in order to analyze key experimental findings about

enzyme catalysis, but also illustrates the fact that is becoming clearer and clearer, that a deeper understanding of this issue is crucial for enzyme design.[92]

Although it is becoming clear that the most important catalytic effect is due to electrostatic preorganization,[2] alternative proposals also have to be explored. Such explorations have been considered in several of our reviews (*e.g.* ref. 121–123 and 160, amongst others), but here we will only consider a few issues that are directly related to the use of the EVB approach.

VII.1 The low barrier hydrogen bond proposal can be best examined by using VB concepts

The catalytic role of hydrogen bonds (HBs) has been an issue of significant controversy since the identification of the oxyanion hole in subtilisin.[164] This structural observation was clearly consistent with the idea that HBs can stabilize the tetrahedral intermediate, but could not provide any estimate of the relevant catalytic energy. Subsequent theoretical studies[83,156,165] have established the idea that the overall electrostatic effect of the preorganized hydrogen bonds contributes to enzyme catalysis in a major way. These theoretical predictions were confirmed by mutation experiments, which clearly demonstrated that a single hydrogen bond can contribute around 5 kcal mol^{-1} to an ionic transition state.[166,167] The results of some specific mutation experiments were subsequently reproduced by free energy perturbation/umbrella sampling (FEP/US) calculations.[83]

After the experimental demonstration of TS stabilization by HBs, it has been proposed by several workers that HBs stabilize TSs in a special non-electrostatic way, which they called a "low barrier hydrogen bond" (LBHB).[168–170] The LBHB proposal has suggested that catalytic HBs involve a flat minimum rather than a double minimum. Unfortunately, this suggestion (which is sometimes true) does not allow one to distinguish the LBHB proposal from the previous proposal of ionic HBs (and thus does not provide a testable definition).

In order to distinguish between ionic HBs and LBHBs, it is essential to first define the LBHB proposal in a way that reflects the energetics of the system and can be used to determine the actual catalytic contribution associated with this proposal. At present, the best way to define the LBHB proposal is to use the valence-bond (VB) representation. This representation can be treated in a simplified two-state version of the three-state model of Coulson and Danielsson,[171,172] augmented by the EVB solvent effect.[37] Here, we consider the $[X^- \ H–Y \leftrightharpoons X–H \ Y^-]$ system as an example, but the same considerations will be applicable to the $[X^- \ H–B^+ \leftrightharpoons X–H \ B]$ system. At any rate, we can describe the total wave function in the two state VB representation by:

$$\Psi = C_1\Phi_1 + C_2\Phi_2 \quad (23)$$

where $\Phi_1 = [X^- \ H–Y]$ and $\Phi_2 = [X–H \ Y^-]$ are the diabatic wave functions whose energies are E_1 and E_2, respectively. The coefficients C_1 and C_2 and the ground state free energy, E_g, are obtained by solving the two-state VB secular equation in its EVB representation, where Φ_1 and Φ_2 are assumed to be orthogonal wave functions whose off-diagonal resonance integral is the mixing term H_{12} (see ref. 37 and 173 for this description). Now we can

approximate the adiabatic ground state free energy by eqn (24), which is written for the present case as:

$$\bar{g}\,(r_1, R) \cong \left[(g_1 + g_2) - \left((g_1 - g_2)^2 + 4H_{12}^2\right)^{1/2}\right]/2 \quad (24)$$

where r_1 is the H–Y bond length and R is the X$\cdots$Y distance, while g_1 and g_2 are the diabatic free energy functionals that correspond to the diabatic potential surfaces E_1 and E_2, respectively. This expression is obtained by converting the EVB ground state potential energy, E_g, to the corresponding ground state free energy functional, $\bar{g}$, by the EVB umbrella sampling approach. We can also approximate the adiabatic free energy barrier by:

$$\Delta\bar{g}' = \Delta g_{dia}' - H_{12} + H_{12}^2/(\lambda + \Delta G_{PT}) \quad (25)$$

where g_{dia}' is the diabatic free energy barrier obtained from the intersection of g_1 and g_2 and where the second and third terms reflect the differences between the diabatic and adiabatic energies at r' and x_1^0. Eqn (23) allows us to immediately define the limits of a LBHB and ionic HB. That is, since g_{dia}' can be approximated by Marcus' formula and be expressed as g_{dia}' $(\lambda + G_{PT})^2/4\lambda$, we obtain g_{dia}' $\simeq \lambda/4$ when $G_{PT} \simeq 0$. Thus, we will have a single minimum or a very small barrier at $r_1 \simeq r$' for g_{dia}' $\simeq |H_{12}|$ and $G_{PT} \simeq 0$ (see Fig. 12). On the other hand, when $\lambda/4 \gg |H_{12}|$, we will have a double minima system which cannot be classified as a LBHB. We also note that the LBHB proponents distinguish between a single minimum and a small barrier at $r_1 \simeq r$', but this does not change any of our conclusions with regards to the interplay between λ and H_{12}. The situation becomes much clearer when $|G_{PT}| > 0$. In this case, we have an ionic HB (Φ_1 or Φ_2, depending on the sign of G_{PT}) and we cannot describe the system as an LBHB.

The transition between the LBHB and HB limits can be further quantified by considering the behavior of $\Delta\bar{g}'$ and asking when this barrier becomes small. This can be formulated by defining a parameter θ, as described in ref. 173.

$$\theta = H_{12}/\left[\Delta g_{dia}' + H_{12}^2/(\lambda + \Delta G_{PT})\right] \quad (26)$$

This equation satisfies the relationship $\Delta\bar{g}' = H_{12}(1 - \theta)/\theta$ where $\Delta\bar{g}'$ is the adiabatic free energy barrier of eqn (23). Now, when θ 1, we have $\Delta\bar{g}' \leq 0$ and the system can be classified as an LBHB.

The above analysis allows one to see how the interplay between the covalent mixing H_{12} and the electrostatic (solvation) effects are reflected by λ and G_{PT}, which determines the nature of ionic HBs. It is also important to note that G_{PT} is linearly correlated with the pK_a difference between the donor and acceptor, and thus we have a clear relationship between the LBHB character and G_{PT}. Now, when θ 1 and when the minimum of the adiabatic ground state is near x_1^0 we can use perturbation theory and write:

$$C_2^2 \left(r \simeq r_1^0 \right) = \left(H_{12}{}^{/} \left(g_2 - g_1 \right) \right)_{r \simeq r_1^0}{}^2 = \left(H_{12}{}^{/} \left[\lambda + \Delta G_{PT} + H_{12}{}^{2/} \left(\lambda + \Delta G_{PT} \right) \right] \right)^2 \quad (27)$$

Note that $g_2 - g_1$ is rigorously equal to $E_2 - E_1$.[120] The magnitude of C_2^2 can tell us how much delocalization we still have in the given HB. At any rate, we established here that the existence of an LBHB is defined in terms of the competition between H_{12} and ($\lambda + \Delta G_{PT}$). In other words, we are dealing here with a competition between the localized [X–H Y$^-$], [X$^-$ H–Y] pictures and delocalized [X$^{-\frac{1}{2}}\cdots$H$\cdots$Y$^{-\frac{1}{2}}$] picture. In the gas-phase, the delocalized picture tends to dominate, while in solution the localized picture is more important.

With the above limiting cases in mind we can ask what is new in the LBHB proposal. Obviously, the idea that HBs (which are preorganized to stabilize ionic TSs) contribute to catalysis is not new (see above). Thus, the only new element in the LBHB proposal is the idea that the covalent delocalized character, which leads to the single energy minimum, is the origin of the catalytic effect. In this respect, it should be clear that HBs in solution have a significant covalent character (for an early demonstration, see ref. 82). Furthermore, for the LBHB proposal to be valid, the covalent character must be larger in the enzyme than in solution, and the corresponding difference must be the source of the HB catalytic effect. Obviously, these issues cannot be examined without evaluating the relevant energies.

At this point it is important to clarify that the entire issue of the validity of the LBHB proposal is related to the interaction between the environment and the VB states of the given ionic HB (in the gas phase we will frequently have LBHBs). The proponents (who originally assigned the enormous energy of ~20 kcal mol^{-1} to LBHB in enzymes) of gas-phase LBHB (*e.g.* ref. 168), argued that × the enzyme environment is non-polar and thus should presumably lead to gas-phase-like LBHB. However, such desolvation arguments are not useful without actual calculations of the relevant polarity and the corresponding solvation effect (in fact, all consistent studies demonstrate that enzyme sites are very polar, *e.g.* ref. 37). Performing such calculations in a reliable way is the best way to examine the LBHB proposal.

This work will not examine specific LBHB proposals since this was done in many of our works including a very recent one.[174] Instead we would only like to emphasize the importance of a well-defined, testable definition. We also take exception to more traditional definitions that unfortunately led the LBHB proponents into circular logic. The landmark work of Hibbert and Emsley[175] classified HBs according to what they called weak, intermediate and strong HBs. Now, while the review of ref. 175 is very instructive, it does not address the effect of the environment on the nature of HBs and thus, in contrast to the EVB approach, cannot be used to analyze this crucial effect or to examine the LBHB proposal. Furthermore, and more importantly, the notion of the strength of HBs, which is reasonable when one deals with HBs in a single phase (*e.g.* gas or solution), becomes extremely problematic when one deals with HBs in proteins. Here, what counts is the energy relative to the corresponding energy in water (stability rather than force). The best way to see this fact is to realize that an ionic HB is very strong in the gas phase, but much less stable than the "weak" HBs in water (see ref. 173 for a clear demonstration of this issue).

Once we avoid the ill-defined concept of "strength" we are back to the above VB definition of the relative energy of the two VB states.

One of the most crucial ingredients for following the LBHB "controversy" is the ability to judge different arguments at their face value. Thus, it is useful to examine the arguments that the LBHB proposal also applies to asymmetric HBs (*e.g.* ref. 176). One can perhaps trace this proposal to the statement in ref. 169 that when the fractional charge (δ) on the donor is between 50% to almost 100%, we can have an LBHB. Unfortunately, if we allow the fractional charge to be close to 100% we clearly *cannot* have an LBHB, since this contradicts the assumption of equal pK_a ($\Delta G_{PT} \simeq 0$), which is shared by all the LBHB proponents, including those of ref. 169.

Having $\delta \sim 1$ corresponds exactly to the localized ionic HB concept, which means that assigning such a system as an LBHB cannot be a new proposal (see above). Apparently, the suggestion that an asymmetric HB is compatible with the LBHB proposal is simply inconsistent with the requirement of having a new proposal or having Δp$K_a \sim 0$. In other words, the common case of asymmetric single-minimum ionic HBs is not a LBHB but a clear case where the ΔpK_a is large. Note in this respect that the idea that LBHBs may involve asymmetric charge distribution is in contrast with the molecular figures presented by the LBHB proponents (*e.g.* Scheme 1 in ref. 177).

It is also important to point out that the LBHBs cannot be defined by such terms as "short strong HB" (SSHB), since an ionic HB can also be short. Furthermore, in contrast to statements (*e.g.* ref. 176) that the LBHB proposal does not imply that the proton is found at an equal distance from the donor and acceptor, this is precisely the requirement for a consistently defined LBHB model. That is, if we have a large ΔG_{PT} we can have an asymmetric HB with [X$^-$ H–Y] as a dominant form so that the proton will be attached to Y. Since this will clearly be an ionic HB, we conclude that the identification of an LBHB with a single minimum system is *only* valid when the minimum is at the center of the X···Y vector.

As far as the definition of the LBHB proposal is concerned, it is important to address the repeated attempts to use experimental observations as operational definitions of this proposal (*e.g.* ref. 170 and 177). Apparently, such a definition confuses the interpretation of experiments with experimental facts. Apparently, most experimental-based definitions of the LBHB proposal are equally consistent with the existence of an ionic HB. Several experiments (*e.g.* studies of the N–H distance) are far more consistent with the ionic HB than the LBHB picture. However, the most crucial issue is the relative energy of the VB states or the ΔpK_a (or ΔG_{PT}) in the protein active site at the TS. Now Δp$K_a > 0$, but this is inconclusive since we have no experimental assignment of all the relevant protonation states (see below) of the states involved. Thus, it is crucial to use theoretical calculations to resolve the LBHB issue.

VIII. Using the EVB to explore dynamical proposals

The proposal that special "dynamical" effects play a major role in enzyme catalysis (*e.g.* ref. 178 and 179) has become quite popular in recent years (*e.g.* ref. 180–190). However, a

significant part of this popularity is a reflection of confusion with regards to the nature of dynamical effects, and the requirement for catalytic contributions, which must be related to the reference reaction in solution. Apparently, many workers overlooked the difference between the well-known fact that all chemical and biological processes involve atomic motions to the requirement for these to be true dynamical contributions to catalysis.

The dynamical proposal has been analyzed at great length in several recent reviews.[51,121,123] These studies used the EVB approach, and were able to show that enzyme catalysis is not due to dynamical effects, regardless of the definition used. The reader is encouraged to look at the above works themselves, and here we will only mention one recent point. That is, instructive NMR experiments (*e.g.* ref. 186 and 191–194) have demonstrated the involvement of different motions in enzymatic reactions. The obvious existence of motions that have components along the reaction coordinate does not in itself constitute a dynamical effect *unless* these motions are shown to be coherent. Probably, all the motional effects identified so far are related to entropic factors (*i.e.* to changes in the available configurational space), rather than being real dynamical effects. Unfortunately, there is no current experiment that can separate the conformational and chemical motions. On the other hand, our recently developed approach[195] has allowed us to use a reduced coarse grained (CG) model to simulate effective millisecond trajectories in the conformational and chemical coordinates. This study established that the energy of the conformational coordinate is fully randomized before it can be transferred to the chemical coordinate,[195] illustrating that dynamical effects cannot be used to accelerate enzymatic reactions.

IX. Recent attempts to adopt the EVB under a different name

As discussed in section II.4, the EVB has been widely adapted in recent years, sometimes under a different name. One of the most high profile recent attempts at this was initially started with gas-phase studies[55,196] under the name "multi-configurational molecular mechanics" (MCMM), which is effectively an identical approach to the EVB, as has already been discussed in ref. 47,48. The more recent attempt to extend the EVB to studies in solution[56] under the name "electrostatically embedded MCMM based on the combined density functional and molecular mechanical method"[56] is more problematic, since to a superficial reader, it may appear to be a novel and effective innovation. However, the electrostatically embedded MCMM (which we refer to here as the EE-MCMM(EVB)) is basically identical to regular EVB, with the exception of one minor modification that is in itself problematic (except for cases where it has negligible effect). That is, this method has the same diagonal EVB elements ε_1 and ε_2 (called V_{11} and V_{22} in ref. 56), and the same crucial embedding by adding the interaction of the diabatic charges with the solvent potential to ε_1 and ε_2 as in the original EVB, and the ground state energy (E_g) is again obtained by mixing the effective off diagonal terms from eqn (4) above. Now (see ref. 57) the solvent is incorporated into the EE-MCMM diagonal elements in the same way as is done by the standard EVB treatment. Even the addition of the solute polarization in the MCMM states has long been implemented in some EVB studies. In fact, our $\varepsilon_i{}^{ind}$ has been treated more consistently than in the treatment of ref. 56, being evaluated consistently (see eqn (7) of ref. 83) with the effect of the solvent permanent and induced dipoles included

(which are not yet considered by the EE-MCMM), and not by trying to do so by some expansion treatment that might or might not provide correct results.

The off-diagonal element of the EE-MCMM is not written explicitly in ref. 56, and we are told[56] that H_{12} (V_{12}) is based on the Shepard interpolation scheme, where it is really based on the EVB equation (eqn (4)), a fact which the authors seem to hide by never mentioning this fact. Of course, many interpolation approaches could be used to fit to the *ab initio* and some of them could be more effective where all the discussion is about the technical interpolation (perhaps in order to draw attention away from the basic EVB treatment). At any rate it *is* stated that V_{12} is evaluated as in ref. 196, which uses precisely the EVB equation (eqn (4)) above. Now the attempt is to imply that we have here a new approach to obtain energy surfaces in aqueous solution with electronic structure information obtained entirely in the gas phase using a solvent dependent V, based on the gas-phase charges:[56]

$$V_{12} = \sqrt{(V_{11} - V_g)(V_{22} - V_g)} \quad (28)$$

(where we used the notation of ref. 56 for eqn (4)).

However, this treatment does not provide a quantitative way to obtain the solvated V_{12} or the solvated ground state. That is, if the expansion treatment of ref. 56 were able to reproduce a correct description of the ground state adiabatic surface in solution, there would be no need for any EVB treatment or any diabatic V_{ii}, and the expansion would have provided the long awaited solution to the general QM/MM-FEP problem by gas-phase expansion so that no-one would need an EVB type formulation. As to the problems with evaluating V_g (and thus H_{12}) by the expansion approach, we can return to our standard example of S_N1 reactions.[37,38] In this case, the gas-phase system in the large separation range is a biradical, with zero charge on the separated atoms. The first term in the expansion will be zero since this term is the gas-phase charge, thus the expansion will give zero solvent effect on V_g (in contrast to the enormous effect obtained with the correct solvation treatment). In fact, the success of the approach of ref. 56 in the case of S_N2 reactions is to be expected, since even the full gas-phase charge distribution (in the Jorgensen's QM-FE treatments) gives reasonable results.[197] However, the same results would be obtained with the EVB and constant H_{12}, and this type of EVB also works extremely well in the case of S_N2 reactions.[37,198] (For more discussion, see ref. 57.)

Therefore, the EE-MCMM method is practically identical to the EVB approach once H_{12} is taken to be solvent independent (otherwise we have an inconsistent attempt to obtain solvent dependence (see ref. 57). In this case, the consistent embedding is entirely due to the effect of the solvent on the diagonal EVB elements. The argument that we have a new method here is particularly alarming when it is presented as the development of a "novel approach" that can be used for studies of enzymatic reactions.

It appears that the MCMM approach is exactly the same as the EVB approach. In other words, the MCMM approach is not equal to the diatomic-in-molecule approach or any other method that is presented as an alternative in ref. 54, but rather it is precisely identical to the EVB. Additionally, the embedded MCMM approach is identical to the solvated EVB. The

diabatic EVB is now used by many major players who have all carefully established its validity (*e.g.* ref. 198 and 199). Finally, the authors of ref. 54 imply through the text that this work will pave the way for the study of reactions in enzymes. We would like to point out that we have already studied not only the model reaction in solution but also the DhlA reaction itself by means of *ab initio* QM/MM with extensive configurational sampling using the EVB as a reference potential[34] with highly promising results. In fact, our EVB approach has already been successfully applied to studies of enzymatic reactions from as early as 1980, and has since also been successfully implemented to do the same thing by several other key workers (see ref. 65–69 for just a few examples).

X. Concluding remarks

In recent years, hybrid quantum mechanical/molecular mechanical approaches have exploded in popularity, and have rapidly become the key tool for studies of chemical processes in proteins.[4,6,8–10,13–16,200–202] However, despite this, we have not yet reached the stage where we can use QM/MM approaches for fully quantitative studies of enzyme catalysis. This is mainly due to the fact that a quantitative evaluation of the potential surface for the reacting fragment should involve *ab initio* electronic structure calculations, which are too expensive to allow for the configurational averaging needed for proper free energy calculations. While specialized approaches have helped move towards *ab inito* free energy calculations, such approaches are still in the development stage.[5,34,35,203] As we have demonstrated here, the EVB is a proper QM/MM method that describes reactivity by mixing diabatic states that correspond to the classical valence bond structures that describe the reactant, intermediate (or intermediates) and product states. Amongst its many advantages are included the fact that this approach facilitates proper configurational sampling and convergent free energy calculations (which includes the inherent ability to evaluate non-equilibrium solvation effects[51]). Additionally, the EVB approach can consistently and conveniently treat the solute–solvent coupling, which is not only essential for the proper modeling of charge-separation reactions but also in allowing for reliable and convenient calibration to *ab initio* calculations. Finally, the EVB is the ideal reference potential for *ab initio* QM/MM calculations, as it stores a tremendous amount of chemical information.

In this work, we have demonstrated that the EVB approach provides a powerful way to connect the classical concepts of physical organic chemistry to the actual energetics of enzymatic reactions by means of computation. That is, when concepts such as Marcus' parabolae are formulated in a consistent microscopic way, they allow us to obtain quantitative LFER in enzymes and in solution, which in turn allows us to quantify catalytic effects and to define them in terms of the relevant reaction free energies, reorganization energies and the preorganization of the enzyme active sites. Thus, we believe that the EVB approach is probably the most powerful current strategy as far as studies of chemical processes in the condensed phase in general and in enzymes in particular are involved, especially when trying to explore the origin of enzyme catalysis.

Acknowledgments

References

1. Åqvist J, Kolmodin K, Florián J, Warshel A. Chem. Biol. 1999; 6:R71–R80. [PubMed: 10074472]

2. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM. Chem. Rev. 2006; 106:3210–3235. [PubMed: 16895325]

3. Pople JA. Angew. Chem., Int. Ed. 1999; 38:1894–1902.

4. Shurki A, Warshel A. Adv. Protein Chem. 2003; 66:249–313. [PubMed: 14631821]

5. Kamerlin SCL, Haranczyk M, Warshel A. J. Phys. Chem. B. 2009; 113:1253–1272. [PubMed: 19055405]

6. Warshel A, Levitt M. J. Mol. Biol. 1976; 103:227–249. [PubMed: 985660]

7. Sakane S, Yezdimer EM, Liu WB, Barriocanal JA, Doren DJ, Wood RH. J. Chem. Phys. 2000; 113:2583–2593.

8. Garcia-Viloca M, Gonzalez-Lafont A, Lluch JM. J. Am. Chem. Soc. 2001; 123:709–721. [PubMed: 11456585]

9. Friesner RA, Beachy MD. Curr. Opin. Struct. Biol. 1998; 8:257–262. [PubMed: 9631302]

10. Bakowies D, Thiel W. J. Phys. Chem. 1996; 100:10580–10594.

11. Thery V, Rinaldi D, Rivail JL, Maigret B, Ferenczy GG. J. Comput. Chem. 1994; 15:269–282.

12. Zhang Y, Liu H, Yang W. J. Chem. Phys. 2000; 112:3483–3492.

13. Gao J. Acc. Chem. Res. 1996; 29:298–305.

14. Field MJ, Bash PA, Karplus M. J. Comput. Chem. 1990; 11:700–733.

15. Monard G, Merz KM. Acc. Chem. Res. 1999; 32:904–911.

16. Field MJ. J. Comput. Chem. 2002; 23:48–58. [PubMed: 11913389]

17. Mulholland AJ, Lyne PD, Karplus M. J. Am. Chem. Soc. 2000; 122:534–535.

18. Kanaan N, Marti S, Moliner V, Kohen AA. Biochemistry. 2007; 46:3704–3713. [PubMed: 17328531]

19. Bowman AL, Grant IM, Mullholland AJ. Chem. Commun. 2008:4425–4427.

20. Wang Y, Schlick T. J. Am. Chem. Soc. 2008; 130:13240–13250. [PubMed: 18785738]

21. Hu H, Lu Z, Yang W. J. Chem. Theory Comput. 2007; 3:390–406. [PubMed: 19079734]

22. Stanton CL, Kuo I-FW, Munday CJ, Laino TD, Houk KN. J. Phys. Chem. B. 2007; 111:12573–12581. [PubMed: 17927240]

23. Pradipta B. J. Chem. Phys. 2005; 122:091102. [PubMed: 15836102]

24. Rod TH, Ryde U. Phys. Rev. Lett. 2005; 94:138302. [PubMed: 15904045]

25. Liu WB, Wood RH, Doren DJ. J. Phys. Chem. B. 2003; 107:9505–9513.

26. Iftimie R, Salahub D, Schofield J. J. Chem. Phys. 2003; 119:11285–11297.

27. Crespo A, Marti MA, Estrin DA, Roitberg AE. J. Am. Chem. Soc. 2005; 127:6940–6941. [PubMed: 15884923]

28. Peräkylä M, Kollman PA. J. Am. Chem. Soc. 1997; 119:1189–1196.

29. Eurenius KP, Chatfield DC, Brooks BR, Hodoscek M. Int. J. Quantum Chem. 1996; 60:1189–1200.

30. Waszkowycz B, Hillier IH, Gensmantel N, Payling DW. J. Chem. Soc., Perkin Trans. 1991; 2:225–231.

31. Klähn M, Braun-Sand S, Rosta E, Warshel A. J. Phys. Chem. B. 2005; 109:15645–15650. [PubMed: 16852982]

32. Alberts IL, Wang YA, Schlick T. J. Am. Chem. Soc. 2007; 129:11100–11110. [PubMed: 17696533]

33. Grigorenko BL, Nemukhin AV, Topol IA, Cachau RE, Burt SK. Proteins: Struct., Funct., Bioinf. 2005; 60:495–503.

34. Rosta E, Klähn M, Warshel A. J. Phys. Chem. B. 2006; 110:2934–2941. [PubMed: 16471904]

35. štrajbl M, Hong G, Warshel A. J. Phys. Chem. B. 2002; 106:13333–13343.

36. Åqvist J, Warshel A. Chem. Rev. 1993; 93:2523–2544.

37. Warshel, A. Computer Modeling of Chemical Reactions in Enzymes and Solutions. John Wiley & Sons; New York: 1991.

38. Hwang J-K, King G, Creighton S, Warshel A. J. Am. Chem. Soc. 1988; 110:5297–5311.

39. Warshel A. J. Phys. Chem. 1982; 86:2218–2224.

40. Marcus RA. Annu. Rev. Phys. Chem. 1964; 15:155–196.

41. Marcus RA. Angew. Chem., Int. Ed. Engl. 1993; 32:1111–1121.

42. Warshel A, Hwang JK, Åqvist J. Faraday Discuss. 1992; 93:225–238. [PubMed: 1337846]

43. Vuilleumier R, Borgis D. J. Phys. Chem. B. 1998; 102:4261–4264.

44. Smondyrev AM, Voth GA. Biophys. J. 2002; 82:1460–1468. [PubMed: 11867461]

45. Billeter SR, Webb SP, Agarwal PK, Iordanov T, Hammes-Schiffer S. J. Am. Chem. Soc. 2001; 123:11262–11272. [PubMed: 11697969]

46. Billeter SR, Webb SP, Iordanov T, Agarwal PK, Hammes-Schiffer S. J. Chem. Phys. 2001; 114:6925–6936.

47. Florián J. J. Phys. Chem. A. 2002; 106:5046–5047.

48. Warshel, A.; Florián, J. The Encyclopedia of Computational Chemistry. von Ragué Schleyer, P.; Allinger, NL.; Clark, T.; Gasteiger, J.; Kollman, PA.; Schaefer, HF., III; Schreiner, PR., editors. John Wiley & Sons; Chichester, UK: 2004.

49. Vuilleumier R, Borgis D. Chem. Phys. Lett. 1998; 284:71–77.

50. Schmitt UW, Voth GA. J. Phys. Chem. B. 1998; 102:5547–5551.

51. Villà J, Warshel A. J. Phys. Chem. B. 2001; 105:7887–7907.

52. Warshel A, Russell S. J. Am. Chem. Soc. 1986; 108:6569–6579.

53. Cuma M, Schmitt UW, Voth GA. J. Phys. Chem. A. 2001; 105:2814–2823.

54. Valero R, Song L, Gao J, Truhlar DG. J. Chem. Theory Comput. 2009; 5:1–22. [PubMed: 20047005]

55. Albu TV, Corchado JC, Truhlar DG. J. Phys. Chem. A. 2001; 105:8465–8487.

56. Higashi M, Truhlar DG. J. Chem. Theory Comput. 2008; 4:790–803.

57. Kamerlin SCL, Cao J, Rosta E, Warshel A. J. Phys. Chem. B. 2009; 113:10905–10915. [PubMed: 19606825]

58. Hong G, Rosta E, Warshel A. J. Phys. Chem. B. 2006; 110:19570–19574. [PubMed: 17004821]

59. Truhlar DG. J. Phys. Chem. A. 2002; 106:5048–5050.

60. Chang Y-T, Miller WH. J. Phys. Chem. 1990; 94:5884–5888.

61. Muller RP, Warshel A. J. Phys. Chem. 1995; 99:17516–17524.

62. Bentzien J, Muller RP, Florian J, Warshel A. J. Phys. Chem. B. 1998; 102:2293–2301.

63. Hwang J-K, Chu ZT, Yadav A, Warshel A. J. Phys. Chem. 1991; 95:8445–8448.

64. Villa J, Bentzien J, Gonzalez-Lafon A, Lluch JM, Bertran J, Warshel A. J. Comput. Chem. 2000; 21:607–625.

65. Watney JB, Agarwal PK, Hammes-Schiffer S. J. Am. Chem. Soc. 2003; 125:3745–3750. [PubMed: 12656604]

66. Kuharski RA, Bader JS, Chandler D, Sprik M, Klein ML, Impey RW. J. Chem. Phys. 1988; 89:3248–3257.

67. Hatcher E, Soudackov AV, Hammes-Schiffer S. J. Am. Chem. Soc. 2004; 126:5763–5775. [PubMed: 15125669]

68. Cascella M, Magistrato A, Tavernelli I, Carloni P, Rothlisberger U. Proc. Natl. Acad. Sci. U. S. A. 2006; 103:19641–19646. [PubMed: 17179046]

69. Blumberger J, Bernasconi L, Tavernelli I, Vuilleumier R, Sprik M. J. Am. Chem. Soc. 2004; 126:3928–3938. [PubMed: 15038747]

70. Shurki A, Štrajbl M, Villa J, Warshel A. J. Am. Chem. Soc. 2002; 124:4097–4107. [PubMed: 11942849]

71. Lee FS, Chu ZT, Warshel A. J. Comput. Chem. 1993; 14:161–185.

72. Olsson MHM, Warshel A. J. Am. Chem. Soc. 2004; 126:15167–15179. [PubMed: 15548014]

73. Wesolowski T, Muller RP, Warshel A. J. Phys. Chem. 1996; 100:15444–15449.

74. Luzhkov V, Warshel A. J. Comput. Chem. 1992; 13:199–213.

75. Wesolowski T, Warshel A. J. Phys. Chem. 1994; 98:5183–5187.

76. Mo Y, Gao J. J. Phys. Chem. A. 2000; 104:3012–3020.

77. Olsson MHM, Hong G, Warshel A. J. Am. Chem. Soc. 2003; 125:5025–5039. [PubMed: 12708852]

78. Xiang Y, Warshel A. J. Phys. Chem. B. 2008; 112:1007–1015. [PubMed: 18166038]

79. Wu Q, Cheng CL, Van Voorhis T. J. Chem. Phys. 2007; 127:164119. [PubMed: 17979331]

80. Lappe J, Cave RJ, Newton MD, Rostov IV. J. Phys. Chem. B. 2005; 109:6610–6619. [PubMed: 16851742]

81. Liu H, Warshel A. Biochemistry. 2007; 46:6011–6025. [PubMed: 17469852]

82. Warshel A, Weiss RM. J. Am. Chem. Soc. 1980; 102:6218–6226.

83. Warshel A, Sussman F, Hwang J-K. J. Mol. Biol. 1988; 201:139–159. [PubMed: 3047396]

84. Jencks, WP. Catalysis in Chemistry and Enzymology. Dover; New York: 1987.

85. Ford LO, Johnson LN, Machin PA, Phillips DC, Tijian R. J. Mol. Biol. 1974; 88:349–371. [PubMed: 4453000]

86. Sharma PK, Chu ZT, Olsson MHM, Warshel A. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:9661–9666. [PubMed: 17517615]

87. Bjelic S, Brandsdal BO, Åqvist J. Biochemistry. 2008; 47:10049–10057. [PubMed: 18759500]

88. Roca M, Liu H, Messer B, Warshel A. Biochemistry. 2007; 46:15076–15088. [PubMed: 18052079]

89. Roca M, Messer B, Hilvert D, Warshel A. Proc. Natl. Acad. Sci. U. S. A. 2008; 105:13877–13882. [PubMed: 18779576]

90. Vamvaca K, Vögeli B, Kast P, Pervushin K, Hilvert D. Proc. Natl. Acad. Sci. U. S. A. 2004; 101:12860–12864. [PubMed: 15322276]

91. Pervushin K, Vamvaca K, Vögeli B, Hilvert D. Nat. Struct. Mol. Biol. 2007; 14:1202–1206. [PubMed: 17994104]

92. Roca M, Vardi-Kilshtain A, Warshel A. Biochemistry. 2009; 48:3046–3056. [PubMed: 19161327]

93. Jiang L, Althoff EA, Clemente FR, Doyle L, Roethlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas I, F. C, Hilvert D, Houk KN, Stoddard BL, Baker D. Science. 2008; 319:1387–1391. [PubMed: 18323453]

94. Bolon DN, Mayo SL. Proc. Natl. Acad. Sci. U. S. A. 2001; 98:14274–14279. [PubMed: 11724958]

95. Lippow SM, Tidor B. Curr. Opin. Biotechnol. 2007; 18:305–311. [PubMed: 17644370]

96. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D. Nature. 2008; 453:164–166. [PubMed: 18464727]

97. Dantas G, Kuhlman B, Callender D, Wong M, Baker D. J. Mol. Biol. 2003; 332:449–460. [PubMed: 12948494]

98. Street AG, Mayo S. Structure. 1999; 7:R105–R109. [PubMed: 10378265]

99. Dahiyat BI, Sarisky CA, Mayo S. J. Mol. Biol. 1997; 273:789–796. [PubMed: 9367772]

100. Chowdry AB, Reynolds KA, Hanes MS, Voorhies M, Pokola N, Handel TM. J. Comput. Chem. 2007; 28:2378–2388. [PubMed: 17471459]

101. Hellinga HW. Nat. Struct. Biol. 1998; 5:525–527. [PubMed: 9665160]

102. Kuhlman B, Baker D. Curr. Opin. Struct. Biol. 2004; 14:89–95. [PubMed: 15102454]

103. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Science. 2003; 302:1364–1368. [PubMed: 14631033]

104. Bolon DN, Grant RA, Baker TA, Sauer RT. Proc. Natl. Acad. Sci. U. S. A. 2005; 102:12724–12729. [PubMed: 16129838]

105. Jamarillo A, Wodak SK. Biophys. J. 2005; 88:156–171. [PubMed: 15377512]

106. Kortemme T, Baker D. Curr. Opin. Chem. Biol. 2004; 8:91–97. [PubMed: 15036162]

107. Jin W, Kambara O, Sasakawa H, Tamura A, Takada S. Structure. 2003; 11:581–590. [PubMed: 12737823]

108. Nagai T, Sawano A, Park ES, Miyawaki A. Proc. Natl. Acad. Sci. U. S. A. 2001; 98:3197–3202. [PubMed: 11248055]

109. Looger LL, Dwyer MA, Smith JJ, Hellinga HW. Nature. 2003; 423:185–190. [PubMed: 12736688]

110. Pokala N, Handel TM. J. Mol. Biol. 2005; 347:203–227. [PubMed: 15733929]

111. Boas FE, Harbury PB. Curr. Opin. Struct. Biol. 2007; 17:199–204. [PubMed: 17387014]

112. Hammett LP. Chem. Rev. 1935; 17:125–136.

113. Hammett LP. J. Am. Chem. Soc. 1937; 59:96–103.

114. Hammett, LP. Physical Organic Chemistry; Reaction Rates, Equilibria, and Mechanisms. McGraw-Hill Book Company; New York: 1970.

115. Taft RW. J. Am. Chem. Soc. 1952; 74:2729–2732.

116. Taft RW. J. Am. Chem. Soc. 1952; 74:3120–3128.

117. Taft RW. J. Am. Chem. Soc. 1953; 75:4538–4539.

118. Newman, MS. Steric effects in organic chemistry. John Wiley & Sons Inc.; New York: 1956.

119. Shorter, J. Plenum Press; London: 1972.

120. King G, Warshel A. J. Chem. Phys. 1990; 93:8682–8692.

121. Warshel A, Parson WW. Q. Rev. Biophys. 2001; 34:563–670. [PubMed: 11852595]

122. Warshel A. Annu. Rev. Biophys. Biomol. Struct. 2003; 32:425–443. [PubMed: 12574064]

123. Olsson MHM, Parson WW, Warshel A. Chem. Rev. 2006; 106:1737–1756. [PubMed: 16683752]

124. Warshel A. Simulating the Energetics and Dynamics of Enzymatic Reactions. Pontificiae Academiae Scientiarum Scripta Varia. 1984

125. Kong Y, Warshel A. J. Am. Chem. Soc. 1995; 117:6234–6242.

126. Schweins T, Warshel A. Biochemistry. 1996; 35:14232–14243. [PubMed: 8916908]

127. Warshel A, Schweins T, Fothergill M. J. Am. Chem. Soc. 1994; 116:8437–8442.

128. Schweins T, Geyer M, Kalbitzer HR, Wittinghofer A, Warshel A. Biochemistry. 1996; 35:14225–14231. [PubMed: 8916907]

129. Bjelic S, Åqvist J. Biochemistry. 2006; 45:7709–7723. [PubMed: 16784222]

130. Sucato CA, Upton TG, Kashemirov BA, Batra VK, Martinek V, Xiang Y, Beard WA, Pedersen LC, Wilson SH, McKenna CE, Florian J, Warshel A, Goodman MF. Biochemistry. 2007; 46:461–471. [PubMed: 17209556]

131. Sucato CA, Upton TG, Kashemirov BA, Osuna J, Oertell K, Beard WA, Wilson SH, Florian J, Warshel A, McKenna CE, Goodman MF. Biochemistry. 2008; 47:870–879. [PubMed: 18161950]

132. Warshel A, Chu ZT, Parson WW. Science. 1989; 246:112–116. [PubMed: 2675313]

133. Kiefer PM, Hynes JT. J. Phys. Chem. A. 2002; 106:1834–1849.

134. Silverman DN, Tu C, Chen X, Tanhauser SM, Kresge AJ, Laipis PJ. Biochemistry. 1993; 32:10757–10762. [PubMed: 8399223]

135. Silverman DN. Biochim. Biophys. Acta, Bioenerg. 2000; 1458:88–103.

136. Åqvist J, Fothergill M, Warshel A. J. Am. Chem. Soc. 1993; 115:631–635.

137. Silverman DN, Lindskog S. Acc. Chem. Res. 1988; 21:30–36.

138. Åqvist J, Warshel A. J. Mol. Biol. 1992; 224:7–14. [PubMed: 1312606]

139. Schweins T, Geyer M, Scheffzek K, Warshel A, Kalbitzer HR, Wittinghofer A. Nat. Struct. Biol. 1995; 2:36–44. [PubMed: 7719852]

140. Warshel A, Russell ST. Q. Rev. Biophys. 1984; 17:283–421. [PubMed: 6098916]

141. Hwang J-K, Warshel A. J. Am. Chem. Soc. 1996; 118:11745–11751.

142. Schutz CN, Warshel A. J. Phys. Chem. B. 2004; 108:2066–2075.

143. Muegge I, Qi PX, Wand AJ, Chu ZT, Warshel A. J. Phys. Chem. B. 1997; 101:825–836.

144. Gerlt JA, Gassman PG. J. Am. Chem. Soc. 1993; 115:11552–11568.

145. Braun-Sand S, Strajbl M, Warshel A. Biophys. J. 2004; 87:2221–2239. [PubMed: 15454425]

146. Yarnell A. Chem. Eng. News. 2004; 82:42–44.

147. Warshel A. Photochem. Photobiol. 1979; 30:285–290. [PubMed: 504352]

148. Burykin A, Warshel A. Biophys. J. 2003; 85:3696–3706. [PubMed: 14645061]

149. Kato M, Pisliakov AV, Warshel A. Proteins: Struct., Funct., Bioinf. 2006; 64:829–844.

150. Riccardi D, Konig P, Guo H, Cui Q. Biochemistry. 2008; 47:2369–2378. [PubMed: 18247480]

151. Olsson MHM, Warshel A. Proc. Natl. Acad. Sci. U. S. A. 2006; 103:6500–6505. [PubMed: 16614069]

152. Pisliakov AV, Sharma PK, Chu ZT, Haranczyk M, Warshel A. Proc. Natl. Acad. Sci. U. S. A. 2008 in press.

153. Braun-Sand S, Sharma PK, Chu ZT, Pisliakov AV, Warshel A. Biochim. Biophys. Acta, Bioenerg. 2008; 1777:441–452.

154. Albery WJ, Knowles JR. Biochemistry. 1976; 15:5627–5631. [PubMed: 999838]

155. Krishtalik LI. J. Theor. Biol. 1980; 86:757–771. [PubMed: 7253670]

156. Warshel A. Proc. Natl. Acad. Sci. U. S. A. 1978; 75:5250–5254. [PubMed: 281676]

157. Lee FS, Chu ZT, Bolger MB, Warshel A. Protein Eng., Des. Sel. 1992; 5:215–228.

158. Marcus RA. J. Chem. Phys. 1956; 24:966–978.

159. Yadav A, Jackson RM, Holbrook JJ, Warshel A. J. Am. Chem. Soc. 1991; 113:4800–4805.

160. Warshel A. J. Biol. Chem. 1998; 273:27035–27038. [PubMed: 9765214]

161. Åqvist J, Fothergill M. J. Biol. Chem. 1996; 271:10010–10016. [PubMed: 8626554]

162. Kraut DA, Sigala PA, Pybus B, Liu CW, Ringe D, Petsko GA, Herschlag D. PLoS Biol. 2006; 4:e99. [PubMed: 16602823]

163. Warshel A, Sharma PK, Chu ZT, Åqvist J. Biochemistry. 2007; 46:1466–1476. [PubMed: 17279612]

164. Alden RA, Birktoft JJ, Kraut J, Robertus JD, Wright CS. Biochem. Biophys. Res. Commun. 1971; 45:337. [PubMed: 5160720]

165. Rao SN, Singh UC, Bash PA, Kollman PA. Nature. 1987; 328:551–554. [PubMed: 3302725]

166. Leatherbarrow RJ, Fersht AR, Winter G. Proc. Natl. Acad. Sci. U. S. A. 1985; 82:7840–7844. [PubMed: 3865201]

167. Carter P, Wells JA. Proteins: Struct., Funct., Bioinf. 1989; 6:240–248.

168. Cleland WW, Kreevoy MM. Science. 1994; 264:1887–1890. [PubMed: 8009219]

169. Frey PA, Whitt SA, Tobin JB. Science. 1994; 264:1927–1930. [PubMed: 7661899]

170. Cleland WW, Frey PA, Gerlt JA. J. Biol. Chem. 1998; 273:25529–25532. [PubMed: 9748211]

171. Coulson CA, Danielsson U. Ark. Fys. 1954; 8:239–244.

172. Coulson CA, Danielsson U. Ark. Fys. 1954; 8:245–255.

173. Warshel A, Papazyan A. Proc. Natl. Acad. Sci. U. S. A. 1996; 93:13665–13670. [PubMed: 8942991]

174. Schutz CN, Warshel A. Proteins: Struct., Funct., Bioinf. 2004; 55:711–723.

175. Hibbert F, Emsley J. Adv. Phys. Org. Chem. 1991; 26:255–379.

176. Halkides CJ, Wu YQ, Murray CJ. Biochemistry. 1996; 35:15941–15948. [PubMed: 8961961]

177. Lin J, Westler WM, Cleland WW, Markley JL, Frey PA. Proc. Natl. Acad. Sci. U. S. A. 1998; 95:14664–14668. [PubMed: 9843946]

178. Careri G, Fasella P, Gratton E. Annu. Rev. Biophys. Bioeng. 1979; 8:69–97. [PubMed: 383008]

179. Karplus M, McCammon JA. Annu. Rev. Biochem. 1983; 52:263–300. [PubMed: 6351724]

180. Kohen A, Cannio R, Bartolucci S, Klinman JP. Nature. 1999; 399:496–499. [PubMed: 10365965]

181. Basran J, Sutcliffe MJ, Scrutton NS. Biochemistry. 1999; 38:3218–3222. [PubMed: 10074378]

182. Radkiewicz JL, Brooks CL III. J. Am. Chem. Soc. 2000; 122:225–231.

183. Cameron CE, Benkovic SJ. Biochemistry. 1997; 36:15792–15800. [PubMed: 9398309]

184. Berendsen HJC, Hayward S. Curr. Opin. Struct. Biol. 2000; 10:165–169. [PubMed: 10753809]

185. Neria E, Karplus M. Chem. Phys. Lett. 1997; 267:23–30.

186. Eisenmesser EZ, Bosco DA, Akke M, Kern D. Science. 2002; 295:1520–1523. [PubMed: 11859194]

187. Kohen A, Klinman JP. Chem. Biol. 1999; 6:R191–R198. [PubMed: 10381408]

188. Sutcliffe MJ, Scrutton NS. Trends Biochem. Sci. 2000; 25:405–408. [PubMed: 10973049]

189. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. Nature. 2007; 450:913–916. [PubMed: 18026087]

190. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE. Nature. 2005; 438:117–121. [PubMed: 16267559]

191. Akke M. Curr. Opin. Struct. Biol. 2002; 12:642–647. [PubMed: 12464317]

192. Wang LC, Pang YX, Holder T, Brender JR, Kurochkin AV, Zuiderweg ERP. Proc. Natl. Acad. Sci. U. S. A. 2001; 98:7684–7689. [PubMed: 11438724]

193. Palmer AG, Kroenke CD, Loria JP. Methods Enzymol. 2001; 339:204–238. [PubMed: 11462813]

194. Palmer AG. Chem. Rev. 2004; 104:3623–3640. [PubMed: 15303831]

195. Pisliakov AV, Cao J, Kamerlin SCL, Warshel A. Proc. Natl. Acad. Sci. U. S. A. 2009 in press.

196. Kim Y, Corchado JC, Villà J, Xing J, Truhlar DG. J. Chem. Phys. 2000; 112:2718–2735.

197. Chandrasekhar J, Jorgensen WL. J. Am. Chem. Soc. 1984; 106:3049–3059.

198. Kim HJ, Hynes JT. J. Am. Chem. Soc. 1992; 114:10508–10528.

199. Benjamin I. J. Chem. Phys. 2008; 129:074508. [PubMed: 19044784]

200. Lyne PD, Mulholland AJ, Richards WG. J. Am. Chem. Soc. 1995; 117:11345–11350.

201. Martí S, Andrés J, Moliner V, Silla E, Tunon I, Bertran J. Theor. Chem. Acc. 2001; 105:207–212.

202. Cui Q, Elstner M, Kaxiras E, Frauenheim T, Karplus M. J. Phys. Chem. B. 2001; 105:569–585.

203. Rosta E, Haranczyk M, Chu ZT, Warshel A. J. Phys. Chem. B. 2008; 112:5680–5692. [PubMed: 18412414]

204. Braun-Sand S, Olsson MHM, Warshel A. Adv. Phys. Org. Chem. 2005; 40:201–245.
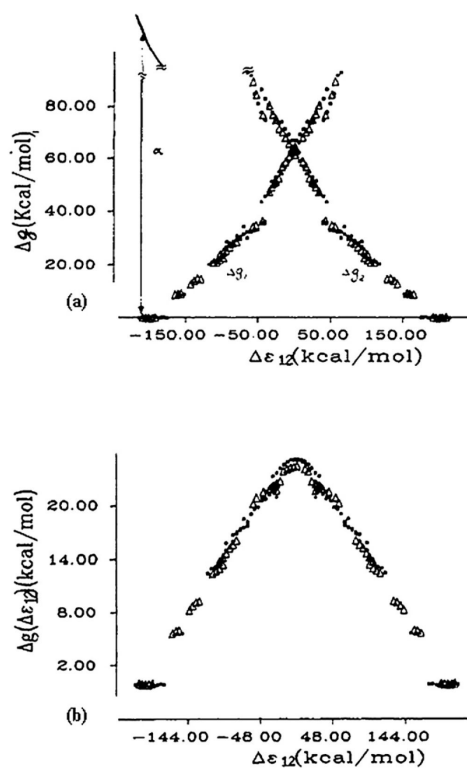
Fig. 1.
a) Free energy functionals ($g1$ and $g2$) for the SN2 reference reaction in solution as a function of the energy gap ε. (b) The actual adiabatic free energy function for the above system (taken from ref. 38).
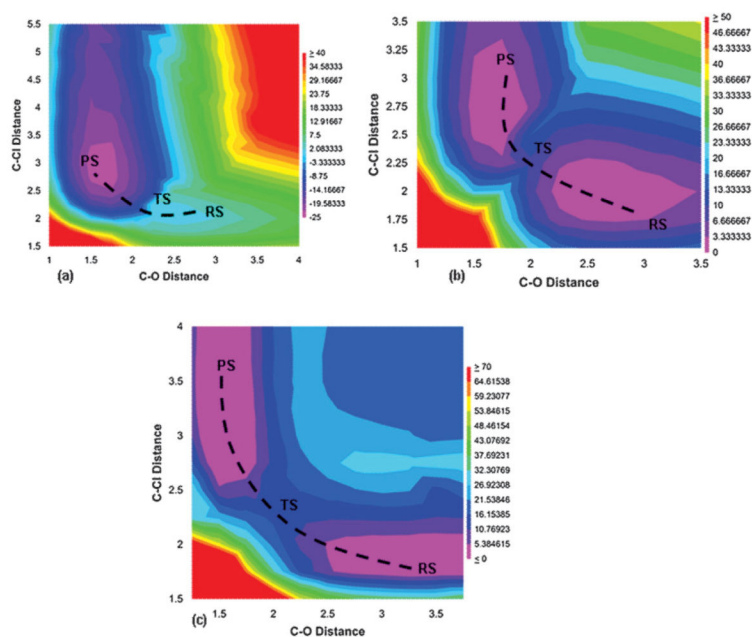
Fig. 2.
Contour plot of the ground-state adiabatic potential surfaces obtained by (a) EVB with a gas-phase shift of −23, (b) EVB with a gas-phase shift of +23.4 and (c) *ab initio* (B3LYP/6-311++G**). The reaction has been defined in terms of C–O (*x* axis) and C–Cl (*y* axis) distances. All energies are given relative to that of the reactant state, and all energies are given in kcal mol^{-1}.
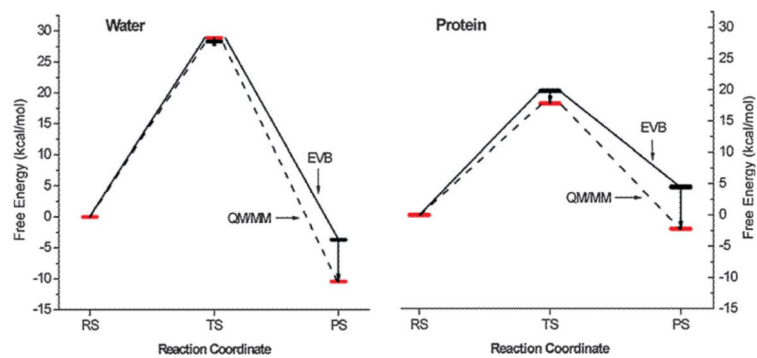
Fig. 3.
QM/MM activation free energies obtained by moving from the EVB to the QM/MM surfaces. This figure was originally presented in ref. 34.
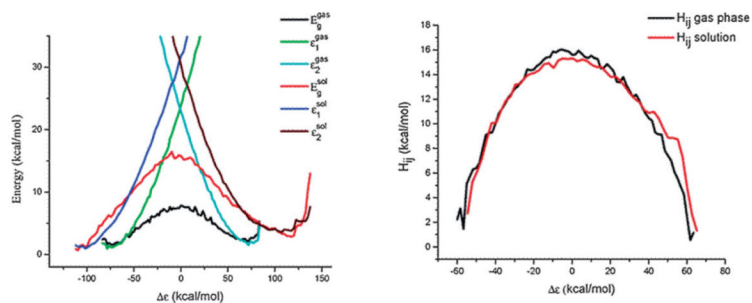
Fig. 4.

(Left) Diabatic and adiabatic FDFT energy profiles for the reaction, Cl⁻ + CH3Cl → ClCH3 + Cl⁻, in the gas phase and in solution, where the reaction coordinate is defined as the energy difference between the diabatic surfaces, $\varepsilon = \varepsilon_1 - \varepsilon_2$. (Right) Plot of the H_{ij} of the reaction, Cl⁻ + CH$_3$Cl → ClCH$_3$ + Cl⁻, both in the gas phase and in solution. The data are obtained from the diabatic and the adiabatic curves of Fig. 1, using

$$H_{12} = \sqrt{(\varepsilon_1 - E_g)(\varepsilon_2 - E_g)}$$ (taken from ref. 58).
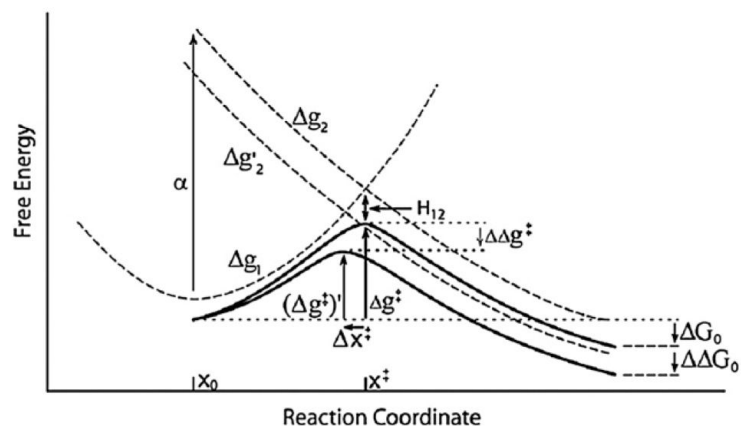
Fig. 5.
A schematic description of the relationship between the free-energy difference ΔG_0 and the activation free energy $\Delta g^{\ddagger}$. The figure illustrates how a shift of Δg_2 by ΔG_0 (that changes Δg_2 to ΔG_2 and ΔG_0 to $\Delta G_0 + \Delta G_0$) changes $\Delta G^{\ddagger}$ by a similar amount (taken from ref. 204).
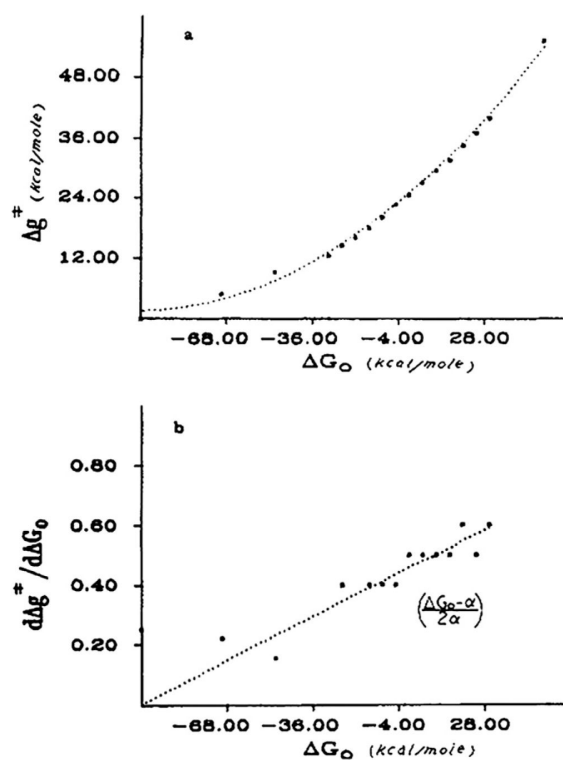
Fig. 6.
a) Calculated relationship between $\Delta g^{\ddagger}$ and ΔG_0 for a series of SN2 reactions. (b) Dependence of the correlation coefficient $\delta \Delta g^{\ddagger}/\delta \Delta G_0$ on ΔG_0. (Taken from ref. 38.)
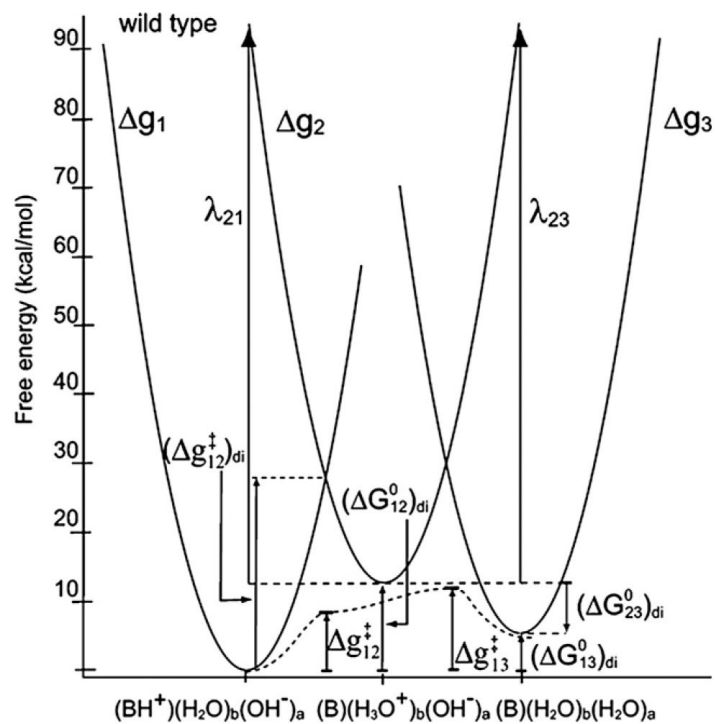
Fig. 7.
Three-state description of the PT in CA III for a case where the transfer involves two water molecules (taken from ref. 142).
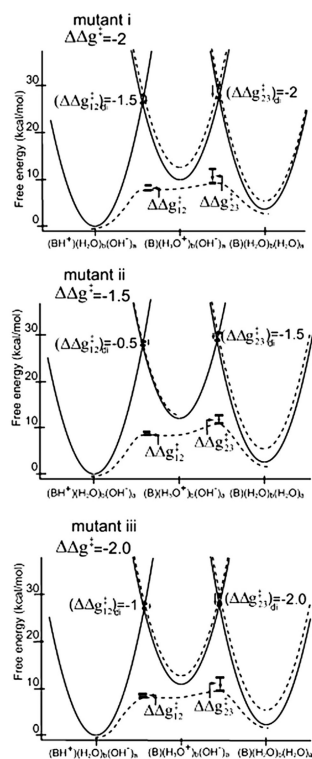
Fig. 8.
Analysis of the energetics of PT in the i, ii, and iii mutants of CA III. The figure describes the three states of eqn (16) and considers their change in each of the indicated mutants (relative to the native enzyme). The figure displays the changes in the diabatic potential surfaces and the corresponding changes in the adiabatic activation barriers. The figure also gives the changes in the diabatic activation energies. The final activation barrier is taken in each case as the highest adiabatic barriers (taken from ref. 142).
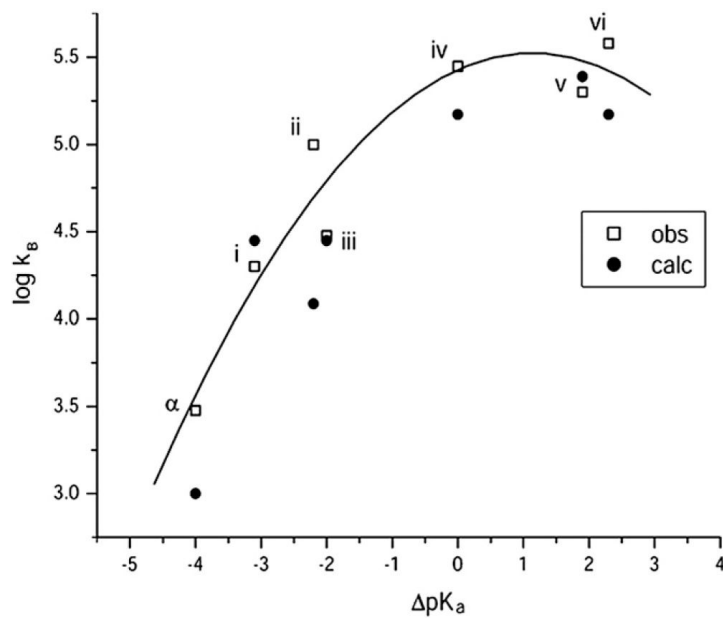
Fig. 9.
Calculated and observed FER for CA III. The different systems are marked according to the notation of ref. 142. The term pK_a corresponds to the pK_a difference between the zincbound water and the pK_a of the given donor group (DpK_a = − g_{13}/2.3RT) (taken from ref. 142).
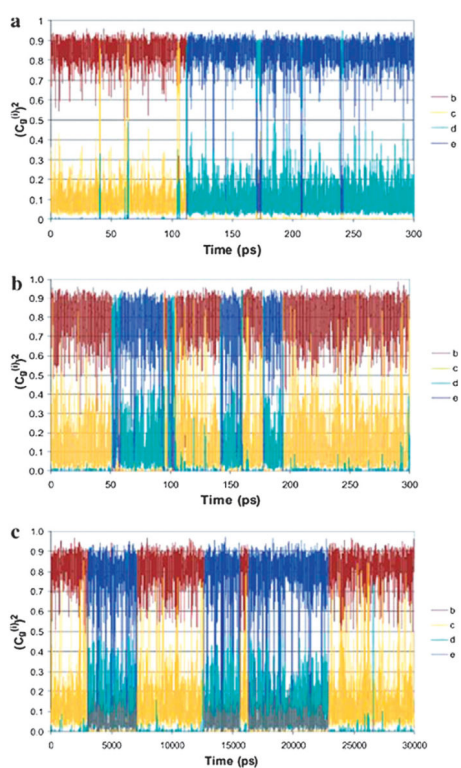
Fig. 10.
The time dependence of the probability amplitude of the transferred proton for a LD trajectory for a PTR that starts at His64 and ends at OH⁻ in the overdamped version of model S/A of the K64H-F198D mutant of CA III. The calculations were accelerated by considering a case where the minimum at site d is raised by 1.2 kcal mol^{-1} (taken from ref. 145).
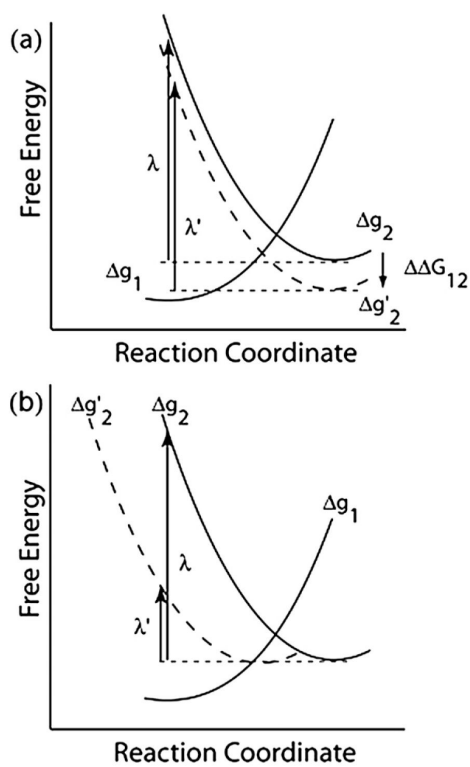
Fig. 11.
Different models for reducing $\Delta g^{\ddagger}$. (a) ΔG_{12} is reduced while keeping the positions of the free energy functionals unchanged. (b) The minimum of Δg_2 is shifted, thus changing the reorganization energy (taken from ref. 204).
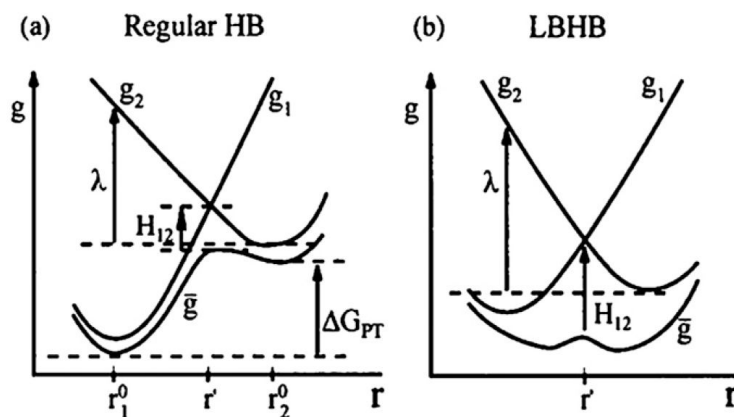
Fig. 12.
a) A two-state VB model for an ionic hydrogen bonded system (see ref. 174). The free energies g_1 and g_2 correspond to the states [X⁻ H–Y] and [X–H Y⁻]. The ground state surface E_g (with a corresponding free energy surface and $\overline{g}$) is obtained from the mixing of the two states. The donor and the acceptor are held at a distance R. The equilibrium distances for isolated X–H and H–Y fragments are designated by r_1^0 and r_2^0. λ and G_{PT} designate the reorganization energy and proton transfer energies, respectively. (b) A two-state VB model in the LBHB limit. In this case, $G_{PT} \approx 0$ and H_{12} $\lambda/4$.

Table 1

FER analysis for different mutants of CA III[a]

	System	$G°_{13}$	pK_a	$(g^\ddagger)_{di}$	$(g^\ddagger)_{calc}$	$(g^\ddagger)_{obs}$
α	Wild-type	5.5	−4.0		0	0
I	K64H	4.4	−3.1	−2.0	−2	−1.8
II	K64H-R67 N	3.0	−2.2	−1.5	−1.5	−2.6
III	K64H-F198L	2.7	−2.0	−2.0	−2.0	−1.9
IV	K64H-R67 N-F198L	0	0	−3.0	−3.0	−3.4
V	K64H-F198D	−2.6	1.9	−3.3	−3.3	−3.2
VI	K64H-R67 N-F198D	−3.2	2.3	−3.0	−3.0	−3.5

[a]Energies in kcal mol^{-1}. This Table was originally presented in ref. 142.