Hybrid ab Initio Quantum Mechanics/Molecular Mechanics Calculations of Free Energy Surfaces for Enzymatic Reactions: The Nucleophilic Attack in Subtilisin

J. Bentzien, R. P. Muller, J. Florián, and A. Warshel*

Department of Chemistry, University of Southern California, Los Angeles, California 90089-1062

Received: October 28, 1997; In Final Form: January 12, 1998

An effective approach for ab initio calculations of activation free energies of enzymatic reactions is developed and examined. This approach uses an empirical valence bond (EVB) potential surface as a reference potential for evaluating the free energies of a hybrid ab initio quantum mechanics/molecular mechanics (QM(ai)/MM) potential surface. This procedure involves an automated calibration of the EVB potential using gas-phase ab initio calculations. In addition, strategies for treating the contact region of QM and MM atoms as well as enzyme and solvent environments are developed. Two levels of ab initio calculations are used in studying the QM atoms: the HF/4-31G method, which allows calculations on a large number of points while still giving accurate results, and the MP2/6-31+G* approach. The QM(ai)/MM method is implemented and examined by simulating the nucleophilic attack step in the catalytic reaction of subtilisin. It is found that the use of the EVB potential as a reference allows one to obtain the actual ab initio activation free energies of enzymatic reactions. Possible powerful simplifications such as the use of the ab initio intermolecular electrostatic energy are discussed, and the advantage of focusing on the difference between the reaction in protein and solution is demonstrated.

1. Introduction

Computer simulations of enzyme reactions are crucial to a quantitative understanding of enzyme activity and mechanism. The empirical valence bond (EVB) approach[1−4] has made significant progress in modeling enzymes, and semiempirical hybrid quantum mechanics/molecular mechanics (QM/MM) techniques[5−8] have provided further insight on such systems. Unfortunately, neither of these solutions is fully satisfactory: fitting EVB parameters to experimental or theoretical data is not a "black box" automatic procedure, and current semiempirical Hamiltonians do not provide accurate potential surfaces for the reacting fragments. There is consequently a great deal of interest in developing ab initio based QM/MM techniques (hereafter referred to as QM(ai)/MM to distinguish them from the semiempirical QM/MM method) for simulating reactions in enzymes.

Despite the success of QM(ai)/MM techniques in solution-phase reactions,[9−11] they have been applied to very few enzymatic reactions,[12−15] without attempting to evaluate the corresponding activation free energies. There are several reasons why enzyme reactions are significantly more complicated than solution-phase reactions: (i) Classical simulations of enzyme reactions are expensive because of the size of enzymes and the long simulations required for convergence. Ab initio simulations on the same system are even more expensive. Furthermore, accurate sampling of the ab initio potential energy surface requires Boltzmann averages over long trajectories. Techniques that average only a few ab initio energies cannot yield reliable results. (ii) In contrast to solution-phase chemical reactions, the quantum region of an enzyme is generally covalently connected to the classical region, and it is not completely clear what is the size of the subsystem that would give reliable results. (iii) Coupling the quantum region with protein electrostatics is also problematic. Simply adding MM protein charges to ab initio calculations without considering solvent screening will produce meaningless results. Correct computational models of enzymes have to include solvent screening, and this requirement represents another reason why enzyme reactions are significantly more challenging than solution-phase reactions.

This paper develops a QM(ai)/MM approach for simulating chemical reactions in enzymes. As with earlier solution-phase QM(ai)/MM methods,[9,16] the current work uses an EVB potential surface to obtain the proper statistical mechanics sampling of the ab initio energies along the reaction path. Additional requirements for extending this approach to studies of enzymatic reactions are identified, and the proper strategies are developed and examined. Special emphasis is placed on exploiting the fact that enzyme catalysis reflects the difference between the reaction in protein and in solution. This allows one to focus on the ab initio electrostatic free energies even when the intramolecular ab initio free energy does not reach perfect convergence.

Finally, this paper confirms the applicability of the EVB description to the enzyme−substrate system. The EVB and ab initio potential surfaces (and in particular their intermolecular part) are shown to be quite similar, indicating that the EVB mapping potential drives the reaction in the same way that an ab initio mapping potential would (if an ab initio mapping potential were possible to obtain) and explaining why earlier EVB studies of enzymatic reactions produced results superior to other approaches.

2. Computational Methodology

2.1. Calculating QM(ai)/MM Free Energies. The enzyme−substrate system is divided into a quantum region that contains those parts of the substrate that change during the reaction plus important enzyme residues, and a classical region that contains

S1089-5647(97)03480-9 CCC: $15.00
Published on Web 02/27/1998

the rest of the enzyme and the solvent. For the classical region the ENZYMIX force field[17] is used with the surface-constrained all-atom solvent (SCAAS) model[18−20] and the local reaction field (LRF) technique,[21] which correctly reproduces the effect of an infinite sphere of water around the protein to represent the classical region. The chemical reaction is driven from reactants to products via the EVB mapping potential,[2]

$$\epsilon_{\mathrm{m}} = (1 - \lambda_{\mathrm{m}})\epsilon_a - \lambda_{\mathrm{m}}\epsilon_b \tag{1}$$

where ϵ_a is a potential function describing the reactant, ϵ_b is a potential function describing the product (or intermediate), and λ_{m} is the mapping variable that ranges from 0 to 1 during the course of the reaction. The potential function ϵ_i ($i = a$ or b) takes the form

$$\epsilon_i = H_{ii} = \sum_j \Delta M_j^{(i)}(b_j^{(i)}) + \sum_l \gamma_l^{(i)} k_l^{(i)}(\theta_l^{(i)} - \theta_{0,l}^{(i)})^2 + V_{\mathrm{nb,rr}}^{(i)} + \alpha^{(i)} + V_{\mathrm{nb,rs}}^{(i)} + V_{\mathrm{s}} \tag{2}$$

where $\Delta M_j^{(i)}$ denotes the Morse potential relative to its minimum value for the jth bond in the ith VB structure. The second and third terms, respectively, denote the bond angle bending contribution and the nonbonded electrostatic and van der Waals interactions between the reacting groups (denoted by subscript r). The factor $\gamma_l^{(i)}$ in the second term is a coupling between bonds that are being broken or formed and those angles that depend on these bonds. The term $\alpha^{(i)}$ is the so called "gas-phase shift" of the ith state.[22] Its value is calculated with all fragments being at infinite separation. The nonbonded interaction with the surrounding protein and solvent, denoted by subscript s, is given by $V_{\mathrm{nb,rs}}^{(i)}$. The last term, V_{s}, represents the internal potential energy of the protein/solvent system.

These potential functions are also used to construct the EVB potential surface with

$$\mathbf{H}\,\mathbf{C} = E_{\mathrm{EVB}}\mathbf{C} \tag{3}$$

where $H_{ii} = \epsilon_i$, and the off-diagonal element is given by

$$H_{ab} = A \exp\{-\mu r\} \tag{4}$$

where r is the distance between atoms whose bonding is changed upon transfer from state a to state b. The parameters A and μ can be adjusted to reproduce the observed barrier for the reaction in solution or to reproduce gas-phase ab initio calculations.

The QM(ai)/MM approach computes the energy of the quantum region using the Hamiltonian

$$H = H_{\mathrm{QM}} + H_{\mathrm{QM/MM}} \tag{5}$$

where H_{QM} is the ab initio Hamiltonian for the quantum region. The Hamiltonian $H_{\mathrm{QM/MM}}$, which couples the quantum and the molecular mechanics regions, is

$$H_{\mathrm{QM/MM}} = \sum_{i,j} \frac{Z_i q_j}{|\mathbf{R}_i - \mathbf{R}_j|} - \sum_{k,j} \frac{q_j}{|\mathbf{r}_k - \mathbf{R}_j|} \tag{6}$$

Here i and j run over the QM and MM atoms, respectively, and k runs over the QM electrons. $\mathbf{r}$ and $\mathbf{R}$ designate the position vectors of the corresponding electrons and atoms, while q_j and Z_i are the residual charges of the MM atoms and the nuclear charges of the QM atoms, respectively. The contribution of this term is simply evaluated by treating the q_j as external charges in the GAUSSIAN[23] program.

The Hamiltonian of eq 5 does not include the van der Waals interactions, V_{vdW}, between the quantum region and its surrounding, the effect of the induced dipoles of the protein and the solvent, V_{ind}, and the effect of the bulk solvent, V_{bulk}. Obviously we could have incorporated the contributions of the induced dipoles and the bulk explicitly into the ab initio Hamiltonian, as was done in early semiempirical studies,[5,20] but using the corresponding EVB contributions appears to provide a reasonable approximation. Thus we introduce the correction term

$$V' = V_{\mathrm{vdW}} + V_{\mathrm{ind}} + V_{\mathrm{bulk}} \tag{7}$$

and calculate the terms of eq 7 by using the ENZYMIX force field.[17] The overall QM/MM energy is then given as

$$E_{\mathrm{QM/MM}} = E_{\mathrm{ai}} + V_{\mathrm{vdW}} + V_{\mathrm{ind}} + V_{\mathrm{bulk}} \tag{8}$$

where E_{ai} is evaluated by

$$E_{\mathrm{ai}} = \langle\psi|H_{\mathrm{QM}} + H_{\mathrm{QM/MM}}|\psi\rangle \tag{9}$$

with ψ being the ab initio wave function corresponding to the Hamiltonian of eq 5.

A second more simplified approach replaces the QM/MM intramolecular energy by the corresponding EVB contribution and retains only the ab initio electrostatic energy. This leads to

$$E_{\mathrm{QM/MM}}^{\mathrm{ele}} = E_{\mathrm{EVB}} - V_{q\mu}^{\mathrm{EVB}} + E_{\mathrm{ai}}^{\mathrm{inter}} \tag{10}$$

where $E_{\mathrm{ai}}^{\mathrm{inter}}$ is the intermolecular contribution to the ab initio energy, obtained by taking the difference between the actual E_{ai} and the gas-phase energy (corresponding to the Hamiltonian H_{QM}) calculated at the same geometry of the quantum atoms. $V_{q\mu}^{\mathrm{EVB}}$ is the EVB electrostatic energy given by

$$V_{q\mu}^{\mathrm{EVB}} = \sum_{i,j} Q_i^{\mathrm{EVB}} q_j / |\mathbf{R}_i - \mathbf{R}_j| \tag{11}$$

where Q_i^{EVB} is the EVB charge of the ith quantum atom and q_j is the charge of the jth MM atom. Free energy surfaces based on these ab initio potentials are obtained using the mapping potential and a free energy perturbation (FEP)/umbrella sampling approach[24] that gives

$$\exp[-\beta\Delta g_{\mathrm{ai}}(X')] = \exp[-\beta\Delta G_{\mathrm{map}}(\lambda_{\mathrm{I}} \rightarrow \lambda_{\mathrm{m}})] \times \langle\delta(X - X')\exp[-\beta(E_{\mathrm{ai}} - \epsilon_{\mathrm{m}})]\rangle_{\epsilon_{\mathrm{m}}} \tag{12}$$

where $\beta = 1/k_{\mathrm{B}}T$, is the Boltzmann constant, T is the absolute temperature, λ_{m} is the value of λ that keeps the system closest to X', and $\langle\ \rangle_{\epsilon_{\mathrm{m}}}$ designates an average over the trajectory with the given ϵ_{m}. $\Delta G(\lambda_{\mathrm{I}} \rightarrow \lambda_{\mathrm{m}})$ is the free energy associated with changing the mapping potential from ϵ_{I} to ϵ_{m}. This value is obtained by a standard FEP approach.[24] The reaction coordinate is usually taken as the difference between the reactant and product EVB potentials ϵ_a and ϵ_b (see refs 2, 9, 24 for more details).

For comparison purpose we also calculate the free energy surface of the ground-state EVB potential using eq 12 with E_{EVB} replacing E_{ai}. Although we are interested here in the ab initio free energy, it is also very useful to have the corresponding EVB free energy surface. This surface is obtained by

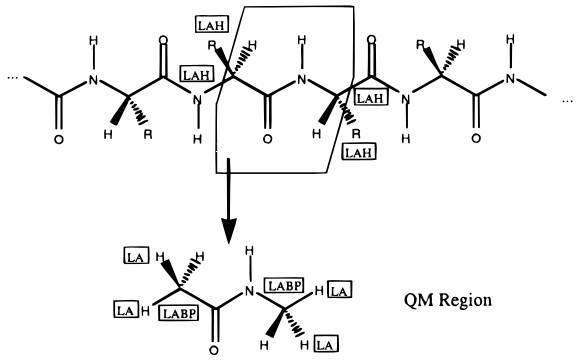

Figure 1. Division of a polypeptide into the QM and MM regions. The upper part shows the polypeptide with the QM region surrounded by a polygon. The remaining atoms are in the MM region. The lower part shows the QM fragment capped with hydrogen link atoms. LABP marks link atom bond partners, and LAH marks the link atom hosts replaced by the hydrogen link atoms LA.

$$\exp[-\beta\Delta g_{\mathrm{EVB}}(X')] = \exp[-\beta\Delta G_{\mathrm{map}}(\lambda_{\mathrm{I}} - \lambda_{\mathrm{m}})] \times \langle\delta(X - X')\exp[-\beta(E_{\mathrm{EVB}} - \epsilon_{\mathrm{m}})]\rangle_{\epsilon_{\mathrm{m}}} \quad (13)$$

2.2. Connecting the Quantum and Classical Regions. In systems where the quantum region is composed of a molecule or molecules that are not bound to the classical region, as is often the case in solution-phase reactions, separation of the two regions is rather trivial. The separation into quantum and classical regions is not as straightforward in enzyme reactions, because the quantum region is typically covalently attached to the classical region, and it is not clear how to terminate the electronic structure of the quantum region nor how to include electrostatic and van der Waals effects of the classical region into the quantum region energy expression. An effective way of connecting the two regions can be provided using hybrid orbitals[5] and related techniques.[25] Unfortunately, these approaches are difficult to implement, and because the QM(ai)/MM method aims to be an efficient method of linking standard programs (in this case GAUSSIAN[23] and ENZYMIX[17]), we chose the simpler method of using link atoms.

A link atom (LA)[10,15] is an atom inserted along the bond between the quantum and classical regions. Such an approach was introduced in ref 5, in addition to the hybrid orbital approach. Using the nomenclature of ref 15, the quantum mechanical atom to which the link atom is bonded is referred to as the link atom bond partner (LABP), and the classical atom replaced by the link atom is referred to as the link atom host (LAH). Warshel et al.[26,27] and Ostlund[28] developed methods that adjust the ionization potential of the LAs to reproduce properties of the given molecule in the presence of the actual LAHs. In our study and refs 10 and 15, the LA is a hydrogen atom; other options will be considered in the future.

Figure 1 illustrates the process of defining the quantum region and inserting LAs. This figure shows a polypeptide chain with the quantum region outlined by a polygon; the structure below the polypeptide chain shows the fragment capped with LAs to produce the quantum region. The LAs are inserted along the LAH−LABP bond at a distance determined by gas-phase ab initio energy minimization.

Despite the intense current interest in proper treatment of link atoms (e.g. refs 15, 25) we feel that this problem is not so crucial in studies of enzymatic reactions. That is, errors introduced by using a small fragment are generally similar in solution-phase and protein simulations. Thus, such errors are largely canceled when one is concerned with enzyme catalysis, which reflects the difference between the activation energy in the enzyme and in the reference solution reaction.

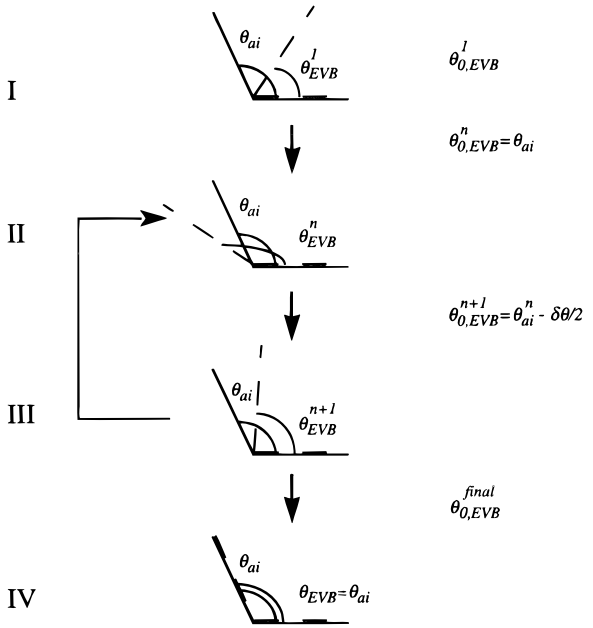

Figure 2. Automatic fitting procedure for the EVB force field parameters to reproduce the ab initio gas-phase geometries using a bond angle θ as an example. The ab initio geometry is shown in solid lines and the EVB geometry in dashed lines. On the right are shown the changes of the equilibrium force field parameters $\theta_{0,\mathrm{EVB}}$ during the fitting process. Steps II and III are iterated until the EVB and ab initio geometries are converged.

2.3. Controlling the MD Reaction Path. Obtaining converging QM(ai)/MM free energies requires the EVB potential surface to be similar to the corresponding ab initio potential surface. This allows trajectories on the EVB mapping potential to sample the most important configurations of the ab initio potential surface. In the earlier solution-phase QM(ai)/MM study[9] the quantum region was sufficiently simple and the EVB surface could be refined by a simple trial and error approach. In enzymes the substrates are too complex to obtain parameters in this fashion. Therefore the QM(ai)/MM technique includes an automatic procedure to obtain EVB parameters that reproduce the minima of the ab initio potential surface.

The first step in this automatic fitting procedure is to obtain minimized gas-phase geometries for the molecules comprising the reactant and product quantum regions. This minimization process uses Møller−Plesset perturbation theory (MP2) calculations with the 6-31+G* basis set that will later be used in the coupled QM/MM calculation.

Once the minimized ab initio structures for the reactant and product are obtained, the EVB force field parameters are calibrated in such a way that the ab initio gas-phase geometries are reproduced. This process is described in the following and illustrated in Figure 2.

In the first step of the automatic fitting procedure the EVB force field parameters are set to the values obtained from the minimized ab initio geometry.

$$b^{1}_{0,\mathrm{EVB}}(i) = b_{\mathrm{ai}}(i)$$

$$\theta^{1}_{0,\mathrm{EVB}}(i) = \theta_{\mathrm{ai}}(i)$$

$$\phi^{1}_{0,\mathrm{EVB}}(i) = \phi_{\mathrm{ai}}(i) \quad (14)$$

where i runs over the bonds b, angles θ, and torsional angles ϕ in the structures of both the reactant and the product. In addition to the force field parameters the charges for the EVB atoms are set to the ab initio values obtained from the gas-phase MP2/6-31+G* calculations:

$$Q_{\mathrm{EVB}}(a) = Q_{\mathrm{ai}}(a) \tag{15}$$

where a runs over the quantum atoms.

The geometry is now minimized using this modified EVB force field. Were there no nonbonded effects, this EVB geometry would now have already converged to the ab initio geometry. But, due to van der Waals and electrostatic effects, the nth minimized geometry will have bond lengths b^n_{EVB}, bond angles θ^n_{EVB}, and torsional angles ϕ^n_{EVB} that differ from the corresponding ab initio values by

$$b^n_{\mathrm{EVB}}(i) = b_{\mathrm{ai}}(i) + \delta b(i)$$

$$\theta^n_{\mathrm{EVB}}(i) = \theta_{\mathrm{ai}}(i) + \delta\theta(i)$$

$$\phi^n_{\mathrm{EVB}}(i) = \phi_{\mathrm{ai}}(i) + \delta\phi(i) \tag{16}$$

Therefore the force field parameters are now readjusted to

$$b^{n+1}_{0,\mathrm{EVB}}(i) = b^n_{0,\mathrm{EVB}}(i) - \delta b(i)/2$$

$$\theta^{n+1}_{0,\mathrm{EVB}}(i) = \theta^n_{0,\mathrm{EVB}}(i) - \delta\theta(i)/2$$

$$\phi^{n+1}_{0,\mathrm{EVB}}(i) = \phi^n_{0,\mathrm{EVB}}(i) - \delta\phi(i)/2 \tag{17}$$

and the geometry is minimized again. This procedure (cf. steps II and III in Figure 2) is repeated until the EVB and ab initio geometries reach a reasonable convergence. Thus a set of EVB parameters is obtained that reproduces the ab initio gas-phase geometries for the reactant and the product. In this procedure the force constants are not changed. The same procedure can be done manually focusing on key internal coordinates.

In the current fitting procedure we had to resolve a problem associated with the strong interaction between the reacting fragments in the gas phase. That is, while in the gas phase one finds a very strong sensitivity to the relative orientation of the methoxide anion and the acetamide molecule, this sensitivity almost completely disappears in solution, as the electrostatic interaction is screened by about a factor of 40. Since we are interested in potential surfaces for polar systems, we used the following procedure. The ab initio geometry of the reactant state was evaluated by optimizing the methoxide anion and acetamide separately and then bringing them to their relative orientation in the enzyme. The EVB parameters were then optimized while setting the electrostatic interaction between the fragments to zero. A better technique would involve performing both the ab initio and the EVB calculations in the presence of a Langevin dipoles (LD)[5,20,29] solvent model, thus fitting the EVB and ab initio solution rather than the gas-phase surfaces.

2.4. QM(ai)/MM Trajectories. The EVB mapping procedure involves running 11 trajectories changing the mapping parameter λ_{m} of eq 1 by 10 equal increments from 0 to 1. The trajectories are propagated at 300 K for 2 ps using a 1 fs time step, resulting in an overall simulation time of 22 ps. This calculation is preceded by an initial relaxation of 4 ps at 50 K and an additional 4 ps at 300 K. The geometry and point charges of the whole system, i.e. the quantum region, protein, and water atoms, are written to a file every 20 fs. At each point

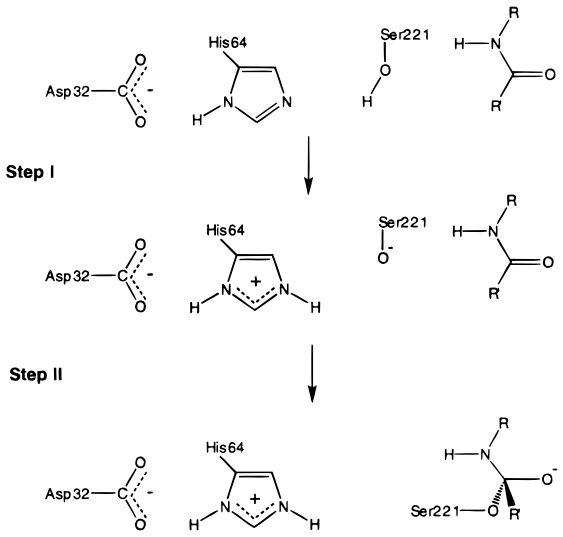

Figure 3. Schematic representation for the enzymatic reaction of subtilisin. Shown are the active site residues involved in the reaction. Step I presents the proton transfer from Ser221 to His64. Step II, which is treated in this paper, shows the nucleophilic attack of the deprotonated Ser221 on the peptide bond of the substrate.

of the reactant, transition state, and product geometries resulting from the MD simulations the ab initio potential energy for the quantum region in the field of the MM charges is calculated using the HF/4-31G and MP2/6-31+G* methods (see eqs 5, 6). The QM(ai)/MM potential is then obtained in two ways: (i) taking the ab initio ground-state energy and correcting it with the EVB van der Waals, induced, and bulk energy using eq 8 or (ii) substituting the EVB intermolecular electrostatic energy by the ab initio term according to eq 10. In the first approach the ab initio energies are additionally shifted by a common constant so that the energies of the reactants from the MM and ab initio calculation are about equal. This is not necessary in the second approach. Free energy profiles are obtained using eq 12, and the ab initio energies are collected by the mapping procedure.

3. Results and Discussion

3.1. Defining the Quantum Region and the EVB Force Field. To examine the performance of our method, we considered the catalytic reaction of subtilisin as a test case. The first step of this reaction (see ref 2 for more details) involves a proton transfer from Ser22l to His64 (see Figure 3). This is followed by the attack of the deprotonated Ser22l oxygen on the carbonyl group of the substrate peptide. The final step is the transfer of a proton back from His64 to the substrate nitrogen adjacent to the carbonyl group, followed by the breaking of the C−N peptide bond. There is some controversy over whether the final step is a stepwise or a concerted process. Here, however, we did not try to explore the degree of concertedness in the reaction but rather examined the stability of our QM-(ai)/MM method. Thus we focus only on the second step in an assumed stepwise mechanism. That is, we assume that the proton-transfer step is completed before the nucleophilic attack step by the oxygen atom of the deprotonated Ser221 on the carbonyl group of the substrate. The crystal structure of subtilisin BPN′ reported by Alden et al.[30] was used as a starting geometry of the enzyme and taken from the Brookhaven Protein Data Bank (E.C.3.4.21.14). A tyrosine−glycine dipeptide was manually fitted into the binding site as a model substrate. Since we restricted ourselves to a study of step II, we treated the

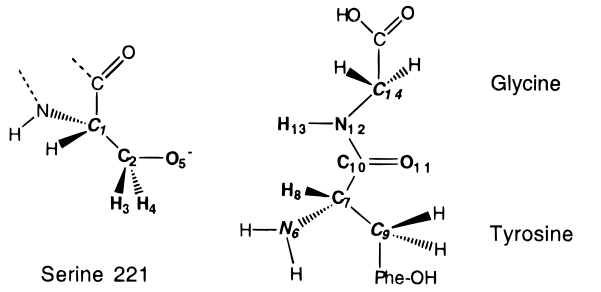

Figure 4. Representation of the quantum region for the nucleophilic attack step in subtilisin. Quantum atoms are numbered and shown in boldface. Italic symbols mark link atom hosts.

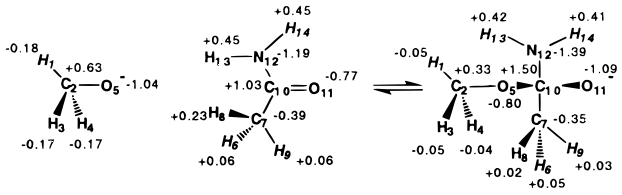

Figure 5. Atomic charges for the ab initio MP2/6-31+G* optimized gas-phase geometries of the reactants and product for the nucleophilic attack of methoxide on acetamide. Italic symbols mark link atoms.

protonated histidine classically and described quantum mechanically only the fragments shown in Figure 4, where the quantum atoms are numbered and typed in bold face. This system contains 10 quantum atoms and 4 LAs are inserted along the Ser221 C_α and C_β bond (atoms 1 and 2), along the substrate tyrosine C_α and N (atoms 6 and 7) as well as the substrate tyrosine C_α and C_β (atoms 7 and 9) and substrate glycine C_α and N (atoms 12 and 14) bonds. Thus, we represent the quantum mechanical region of the reaction by considering the attack of a methoxide ion on acetamide. The remaining atoms are treated with the ENZYMIX force field,[17] and with the exception of the hydroxyl group in tyrosine their charges were all set to zero to obtain more stable results. For the enzymatic reaction all protein atoms surrounding the EVB atoms shown in Figure 4 in an 18 Å distance from the O_γ-oxygen in Ser221 were included in the calculations. His64 was taken to be protonated and Asp32 as negatively charged. All other amino acids were taken in their un-ionized forms.

The reactant and product gas-phase geometries were optimized at the ab initio MP2/6-31+G* level, and the resulting atomic charges are shown in Figure 5.

These equilibrium geometries and the procedure described in section 2.3 were used to refine the EVB parameters for the potential functions that describe bond stretching, angle bending, and torsions of the quantum atoms. The resulting parameters are given in Table 1. The charges for the quantum atoms were taken as the corresponding gas-phase MP2/6-31+G* charges (see Figure 5). Other EVB parameters, except those mentioned below, were kept the same as in previous studies of subtilisin (e.g. ref 31).

As the van der Waals parameters were originally parametrized for using the empirical ENZYMIX charges, they had to be rescaled for the use with ab initio charges. For the analyzed reaction it turned out to be sufficient to change only the van der Waals parameters for the attacking O^-. The van der Waals energy is calculated using a Lennard-Jones potential of the form

$$V_{\mathrm{vdW}} = \frac{A_i A_j}{r_{ij}^{\,6}} - \frac{B_i B_j}{r_{ij}^{\,12}} \qquad (18)$$

and the values used for O^- are $A = 500$ kcal$^{1/2}$ mol$^{-1/2}$ Å^3 and $B = 24$ kcal$^{1/2}$ mol$^{-1/2}$ Å^6.

TABLE 1: Refined EVB Force Field Parameters for Reactant and Product Geometries[a]

bonds		reactants		product	
i	*j*	*D*	b_0	*D*	b_0
1	2	96.0	1.400	96.0	1.400
2	3	100.4	1.112	100.4	1.079
2	4	100.4	1.111	100.4	1.090
2	5	93.0	1.378	93.0	1.370
5	10	*b*	*b*	93.0	1.540
6	7	94.0	1.400	94.0	1.400
7	8	100.4	1.118	100.4	1.093
7	9	96.0	1.400	96.0	1.540
7	10	96.0	1.527	96.0	1.545
10	11	93.0	1.201	93.0	1.270
10	12	94.0	1.352	94.0	1.480
12	13	98.3	1.005	98.3	1.000
12	14	94.0	1.447	94.0	1.440

bond angles			reactants		product	
i	*j*	*k*	K_θ	θ_0	K_θ	θ_0
1	2	3	75.0	1.807	75.0	1.826
1	2	4	75.0	1.808	75.0	1.861
1	2	5	75.0	2.007	75.0	1.911
3	2	4	75.0	1.817	75.0	1.972
3	2	5	75.0	2.006	75.0	1.911
4	2	5	75.0	1.994	75.0	1.911
6	7	8	75.0	1.911	75.0	1.885
6	7	9	75.0	1.897	75.0	2.007
6	7	10	75.0	1.927	75.0	1.920
8	7	9	75.0	1.947	75.0	1.885
8	7	10	75.0	2.075	75.0	1.900
9	7	10	75.0	1.737	75.0	1.885
7	10	11	75.0	2.126	75.0	1.934
7	10	12	75.0	2.021	75.0	1.911
11	10	12	75.0	2.155	75.0	2.111
10	12	13	75.0	2.111	75.0	2.182
10	12	14	75.0	2.121	75.0	1.920
13	12	14	75.0	2.072	75.0	1.911
2	5	10	*b*	*b*	75.0	2.128
5	10	7	*b*	*b*	75.0	1.868
5	10	11	*b*	*b*	75.0	1.833
5	10	12	*b*	*b*	75.0	1.990

torsional angles				reactants			product		
i	*j*	*k*	*l*	K_ϕ	ϕ_0	*n*	K_ϕ	ϕ_0	*n*
6	7	10	11	3.0	0.0	3	3.0	0.0	3
7	10	12	14	15.0	0.0	2	3.0	0.0	3
10	5	2	1	*b*	*b*	*b*	3.0	0.0	3
2	5	10	7	*b*	*b*	*b*	3.0	0.0	3
10	7	11	12	50.0	0.0	3	50.0	0.0	3
12	10	13	14	95.0	0.0	2	95.0	0.0	2

[a] The numbers assigned to the indices *i*, *j*, *k*, and *l* are indicated in Figure 4. The bond energy is evaluated using a Morse potential of the form $\Delta M = D[1 - \exp\{-a(b - b_0)\}]$ with $a = 2.0$. Bond angle energies are calculated using $U_\theta = {}^1/_2 K_\theta(\theta - \theta_0)$. The function $({}^1/_2)\sum_i K_{\phi,i}(1 - \cos n\,\phi_i)$ is used for the torsional energy, with *n* being the multiplicity of the torsional potential. All energies are in kcal/mol. θ_0 and ϕ_0 are in radians. [b] These values do not exist in the reactant geometry.

The gas-phase equilibrium geometries of the reactants, methoxide and acetamide, and the product on the refined EVB surface were found to be quite close to the corresponding ab initio values. The deviations for the bond lengths are < 0.04 Å and for the angles are < 3° in the reactants and product state.

The value of the so-called "gas-phase shift" parameter, α of eq 2, was adjusted so that the calculated EVB free energy will reproduce the estimated reaction free energy for the nucleophilic attack of methoxide on acetamide in aqueous solution ($\Delta G \approx$ 4.1 kcal/mol). This estimate was based on using the ChemSol[29,32] program, which combines the ab initio Hamiltonian with

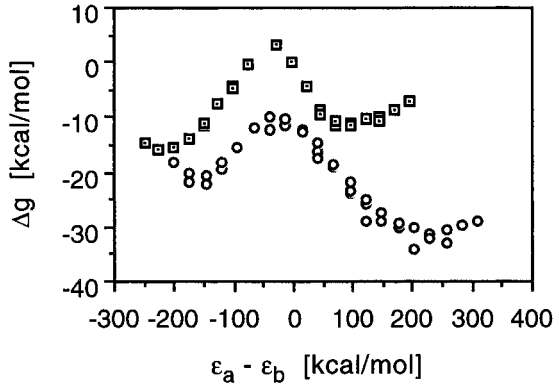

Figure 6. EVB free energy surfaces for the nucleophilic attack in water and in protein. The reaction coordinate is the energy difference between the reactant and product EVB potentials. The free energy surface for the reaction in water is marked by dotted squares (⊡), and for the reaction in protein it is marked by circles (○).

the Langevin dipoles solvent model. The experimental value of this energetics is not available, and previous estimates gave $\Delta G \approx 0$ kcal/mol.[33] For the off-diagonal element of eq 4 a constant value of $H_{ab} = 60$ kcal/mol was taken, which corresponds to $\mu = 0$ in this equation. This value was obtained to reproduce the observed activation barrier, $\Delta g^{\ddagger}$, for the OH^- attack on acetamide. This was calculated to be around $\Delta g^{\ddagger} = 12.1 \pm 5$ kcal/mol with the ChemSol approach, and experimental estimates give a value of $\Delta g^{\ddagger} = 14$ kcal/mol.[33]

3.2. FEP−EVB Results. Having established the EVB force field, MD simulations were carried out for the reaction in water and for the reaction in the protein. The results of the simulation for the reaction in water and in the enzyme active site are shown in Figure 6. The activation free energies for the reaction in water and in the protein are ~19 and ~12 kcal/mol, respectively. These results cannot be converted to the catalytic effect of the enzyme since this effect reflects the sum of the energies of the proton-transfer (PT) and nucleophilic attack steps and this work does not consider the PT step. Nevertheless, the current reduction of the activation barrier is similar to that obtained in more complete earlier studies.[31] The result for the enzymatic reaction seems to be overall too exothermic, which reflects less than perfect overall convergence. That is, previous studies indicate that the convergence of the overall PT and nucleophilic steps is much better than that of each individual step. Furthermore, in more functionally (rather than technically) oriented studies we frequently run longer trajectories and average them over several initial conditions.[34] Here, however, the focus is not on obtaining the most reliable EVB free energies but rather on developing a converging QM(ai)/MM strategy and on the examination of the relationship between the QM(ai) and EVB free energies for a limited simulation time. One might also note that we did not force $\Delta g^{\ddagger}$ for the solution reaction to reproduce the exact observed value. This could be easily achieved by adjusting the nonbonded parameters for the interaction between O_5 and C_{10} relative to their value in previous EVB studies (we changed of course the corresponding charges). However, such additional refinement procedure would not change the catalytic effect and does not seem to be justified in the present work.

3.3. QM(ai)/MM Results. After obtaining the FEP/EVB results we performed ab initio calculations on the reactant, transition-state, and ground-state FEP/EVB geometries for the reaction both in water and in the protein. The first step in these calculations involved the use of the HF/4-31G approximation. This was done as a convenient tool for demonstrating our approach since it is quite expensive to calculate the potential energy with larger basis sets for a big enough number of points, which is essential for obtaining free energies with meaningful sampling in a reasonable time. Perhaps more importantly, as discussed in section 3.3.3, we introduce in a subsequent step of our procedure a basis set correction treatment. This correction provides a good approximation for the results expected from ab initio methods that include electron correlation effects and larger basis sets.

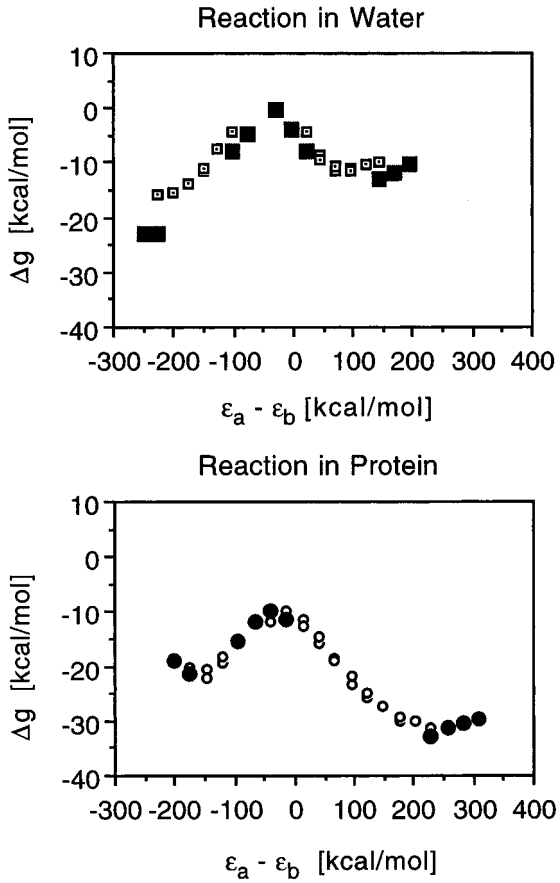

Figure 7. Electrostatic QM(ai)/MM free energy surfaces for the reaction in water (top) and the reaction in protein (bottom). The reaction coordinate is the energy difference between the reactant and product EVB potentials. EVB points in water are marked by dotted squares (⊡); ab initio points are marked by filled squares (■). The EVB points for the reaction in protein are marked with open circles (○), and ab initio points with filled circles (●).

3.3.1. QM(ai)/MM Electrostatic Free Energies. Before examining the total QM(ai)/MM free energy we considered a simpler method where the intermolecular electrostatic EVB energy was replaced by the corresponding ab initio term (see eq 10). The resulting free energy curves for the reaction in water and in protein are shown in Figure 7. In these figures only those points marked around the reactant, product, and transition state were obtained by ab initio calculations. As seen from Figure 7, the ab initio points fit very nicely along the line generated by the EVB free energy calculations. This shows that the intermolecular electrostatic energies obtained from ab initio and the EVB potentials are very similar. Comparing the catalytic effect calculated from this QM(ai)/MM analysis with the pure EVB result, we see an excellent agreement.

To further analyze the potential energies, we show in Figure 8 a plot of $E_g - \epsilon_m + \Delta G$ versus $\epsilon_a - \epsilon_b$. Here E_g is the ground-state energy, ϵ_m the mapping potential, and ΔG the mapping free energy. This plot allows us to illustrate all the raw data

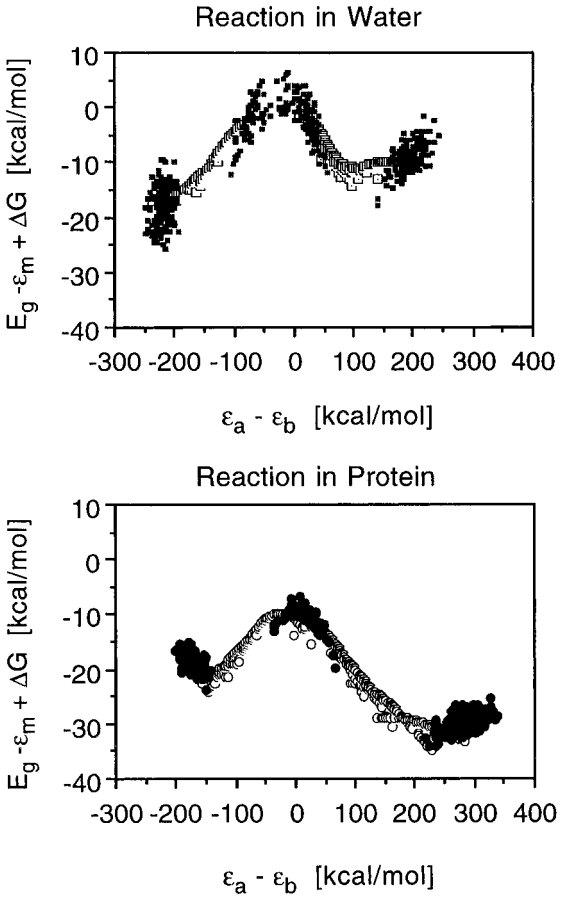

Figure 8. Distribution of the electrostatic QM(ai)/MM potential energies for the reaction in water (top) and the reaction in protein (bottom). The reaction coordinate is the energy difference between the reactant and product EVB potentials. EVB points in water are marked by dotted squares (⊡); ab initio points are marked by filled squares (■). The EVB points for the reaction in protein are marked with open circles (○), and ab initio points with filled circles (●).

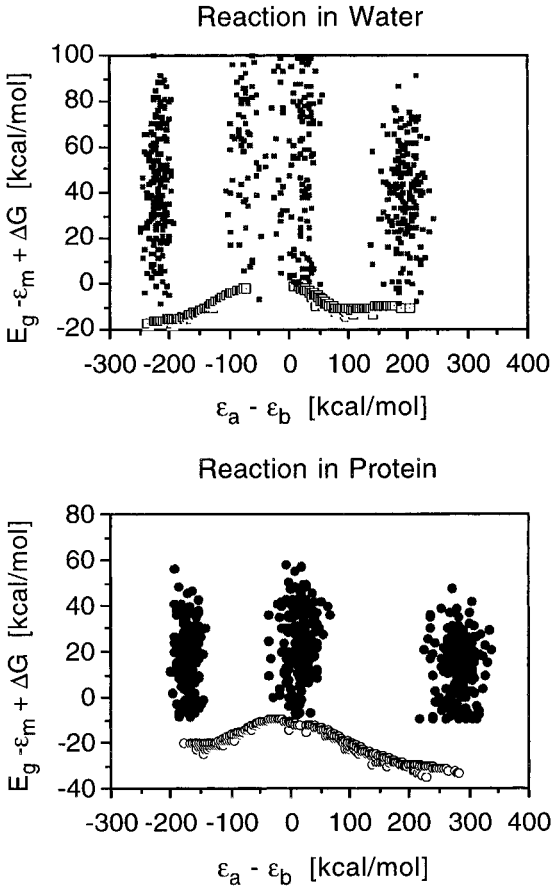

Figure 9. Distribution of the full QM(ai)/MM potential energies for the reaction in water (top) and the reaction in protein (bottom). The reaction coordinate is the energy difference between the reactant and product EVB potentials. EVB points in water are marked by dotted squares (⊡); ab initio points are marked by filled squares (■). The EVB points for the reaction in protein are marked with open circles (○), and ab initio points with filled circles (●).

that go into the evaluation of the free energy functions of the figures. The results presented in Figure 7 and in Figure 8 are quite significant. First they demonstrate that the QM(ai)/MM electrostatic free energies reach a proper convergence. Furthermore, it is demonstrated that the EVB electrostatic free energy surfaces are similar to those obtained with a properly fluctuating ab initio charge distribution. This fact indicates that the EVB method captures the main physics of enzyme catalysis and thus provides an accurate model for studying this effect.

3.3.2. Total QM(ai)/MM Free Energies. Having shown that replacing the EVB intermolecular electrostatic energy by the corresponding ab initio term gives excellent results, we now turn to the more rigorous approach where the EVB ground-state energy is replaced by the ab initio ground-state energy according to eq 8.

We start these analyses by considering the fluctuation of $E_g - \epsilon_m + \Delta G$ versus $\epsilon_a - \epsilon_b$, as was done for the electrostatic energies in Figure 8. The results summarized in Figure 9 indicate that the ab initio energies vary much more than the corresponding EVB energies. The fluctuations are clearly much larger than the corresponding variations of the electrostatic energies shown in Figure 8. Taking the fluctuating $E_g - \epsilon_m$ and analyzing it by eq 12 gives somewhat more encouraging results (Figure 10), although clearly inferior to those obtained by considering only the electrostatic contributions (see Figure 7). Overall, despite the large statistical spread we find that the average ab initio free energies agree reasonably well with the corresponding EVB results.

Both the EVB and the ab initio calculations produce large catalytic effects, although direct comparison with the experimentally observed catalysis requires simulation of the entire reaction (see section 3.2) and is out of the scope of this work.

The fact that the QM(ai) potential has larger fluctuations than the corresponding EVB potential can be due in part to having a less than optimal mapping potential. That is, trajectories around the least energy path of the EVB mapping potential would lead to larger energy fluctuations of the QM(ai) surface than the EVB surface if the least energy path of the QM(ai) surface is shifted from the corresponding EVB path. This problem can be greatly reduced by further refinement of the EVB surface. In particular one can use the QM(ai) potential energies generated along mapping trajectories to refine the EVB surface. That is, the n values of the QM(ai) energies, $E_{ai}(\mathbf{R}_i)$ collected over a given trajectory and the corresponding coordinates, $\mathbf{R}_i$, can be sent to an optimization program that would further refine the EVB parameters by requiring that

$$\sum_{i=1}^{n}(E_{ai}(\mathbf{R}_i) - E_{EVB}(\mathbf{R}_i))^2 \rightarrow \min \qquad (19)$$

We can also require that the QM(ai) and EVB forces for the selected $\mathbf{R}_i$ will be similar. The advantage of these strategies relative to the refinement approach of section 2.3 is that they

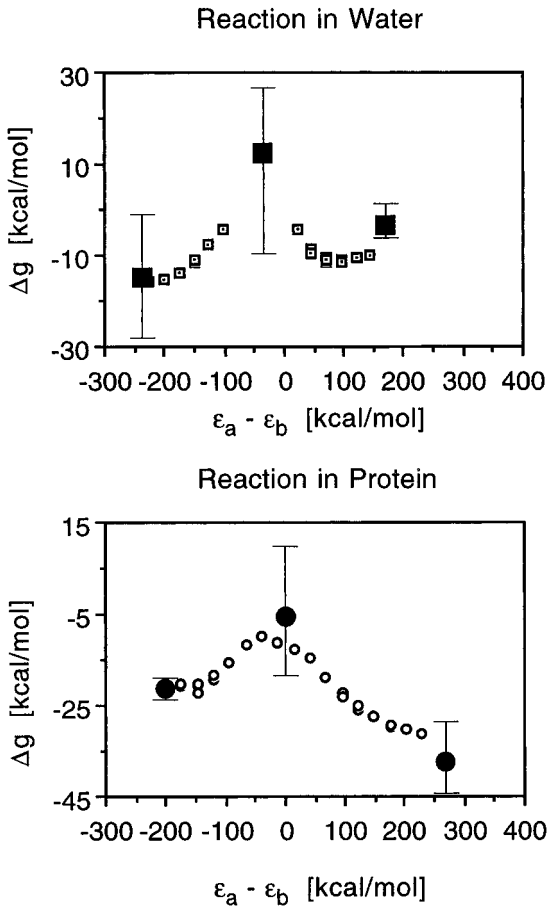

Figure 10. Full QM(ai)/MM free energy surfaces for the reaction in water (top) and the reaction in protein (bottom). The reaction coordinate is the energy difference between the reactant and product EVB potentials. EVB points in water are marked by dotted squares (⊡); ab initio points are marked by filled squares (■). The EVB points for the reaction in protein are marked with open circles (○), and ab initio points with filled circles (●). The error bars represents the maximum deviation from the average ab initio value.

deal with a solvated substrate, so that the effect of the large gas-phase electrostatic interaction is strongly attenuated. The examination and implementation of such strategies is currently under development in our lab.

3.3.3. Effect of ab initio Hamiltonian on QM(ai)/MM Results. Since it is quite expensive to perform MP2/6-31+G* calculations along the same number of points as when using the HF/4-31G method, we calculated the MP2 energies for the reaction in water for 50 points at the reactant, transition-state, and product geometries generated by the MD simulation instead of 200 points used in the HF/4-31G analysis. As can be seen from Figure 11, the MP2 curve for the potential energy has the same shape as that for the HF result but is overall lower in energy. The average gain in energy by using MP2/6-31+G* is 15.6, 18.4, and 19.3 kcal/mol at the reactants, transition state, and product, with standard deviations being 2.6, 1.9, and 2.4 respectively. From this analysis it seems justified to simulate the effect of MP2/6-31+G* by lowering the energy at the transition-state and product geometries by 2.6 and 3.7 kcal/mol relative to the energy of the reactants. More extensive studies that will involve the actual evaluation of MP2/6-31+G* free energies will be reported elsewhere.

4. Concluding Remarks

This work has developed and examined a general strategy for calculating free energies of enzymatic reactions by hybrid

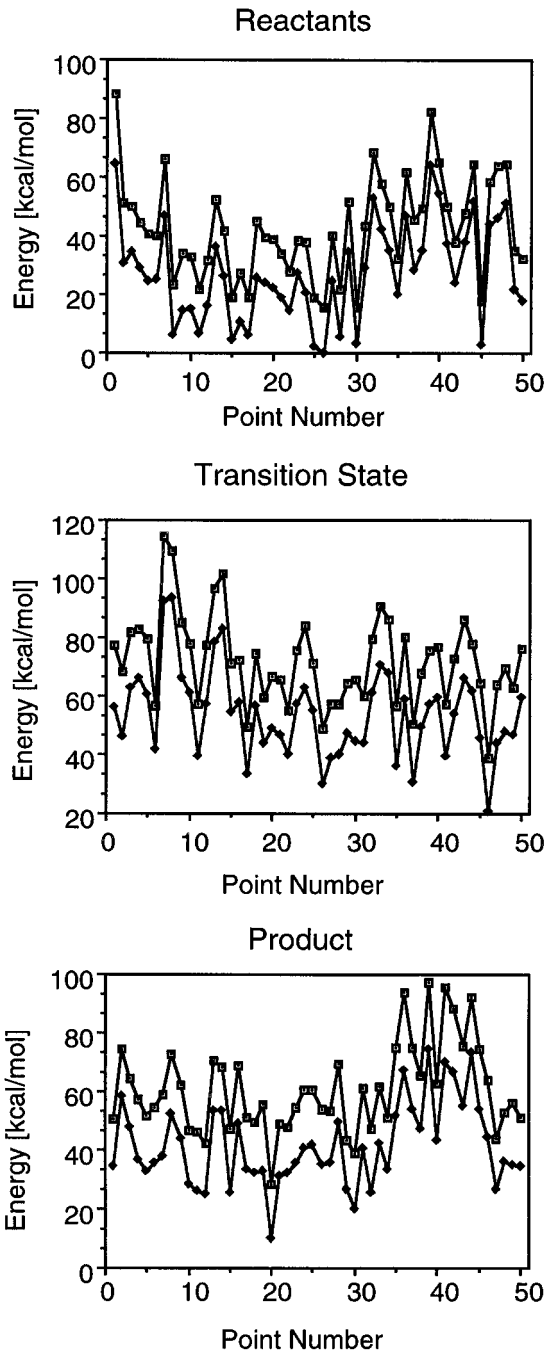

Figure 11. QM(ai) potential energy distribution for the reaction in water, comparison of HF/4-31G and MP2/6-31+G* results. Presented is the potential energy for 50 points along trajectories at the reactant (top), transition-state (middle), and product (bottom) geometries. The QM(ai) potential energy was calculated every 20 fs along a trajectory of 1 ps. HF/4-31G results are marked by (⊡) and MP2/6-31+G* results by ◆. The average difference between the HF and MP2 energies at the reactant geometry is 15.6 kcal/mol, with a standard deviation of 2.57. For the transition state the energy difference is 18.4 kcal/mol, with a standard deviation of 1.94, and for the product geometry the values are 19.3 kcal/mol and 2.43.

QM(ai)/MM methods. This strategy is based on using the EVB potential surface as a reference potential for the evaluation of the QM(ai)/MM free energy. The EVB potential surface is refined iteratively by fitting its gas-phase minima to the corresponding ab initio minima. The refined EVB surface is then used as a mapping potential for sampling the ab initio free energy. The validity and convergence of our approach is examined by considering the nucleophilic attack step in the

catalytic reaction of subtilisin. The result of the simulations can be considered encouraging or somewhat discouraging depending on one's perspective. On one hand, we establish that the electrostatic contributions to the ab initio free energy function converge in a very satisfactory way and are in excellent agreement with the corresponding EVB results. On the other hand, the full ab initio free energies display a large statistical uncertainty, although their average values provide reasonable results.

The catalytic effect of the enzyme is to a large extent associated with the change in the "solvation energy" of the reacting fragments upon moving from water to the enzyme active site.[2] Thus we find it quite promising that our approach gives converging results for the electrostatic contributions to the ab initio free energy function. In fact, if we adopt the approximation that the intramolecular contributions are similar in solution and in the enzyme active site, we do have an ab initio strategy for studying enzyme catalysis. This approach captures the change in the substrate charge distribution during the reaction and properly takes into account the effect of the surrounding environment.

As stated above, the use of the full QM(ai)/MM potential surface does not provide sufficiently stable results. The convergence can be improved by minimizing the difference between the EVB and ab initio surfaces. This can be done either by improving the geometry-based refinement procedure of section 2.3 or by using results of the mapping runs in the energy-based procedure of section 3.3.2.

This paper has also confirmed the utility of the EVB method for performing calculations in enzymes. It has demonstrated an automatic technique for obtaining EVB parameters that reproduce ab initio geometries. It has also shown that the EVB charge fluctuations during the reaction correspond very well to those calculated using ab initio techniques. Because of the great expense of ab initio calculations compared to that of EVB calculations, it is likely that the EVB method will remain the most practical method of computing free energy profiles for chemical reactions in enzymes for some time to come.

Acknowledgment. This work was supported by NIH Grant GM24492. J.B. gratefully acknowledges support by a fellowship from the Deutsche Forschungsgemeinschaft (DFG).

References and Notes

(1) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218.
(2) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons: New York, 1991.
(3) Åqvist, J.; Warshel, A. *J. Mol. Biol.* **1992**, *224*, 7.
(4) Åqvist, J.; Fothergill, M. *J. Biol. Chem.* **1996**, *271*, 10010.
(5) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(6) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826.
(7) Zheng, Y.-J.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **1992**, *114*, 10498.
(8) Monard, G.; Loss, M.; Théry, V.; Baka, K.; Rivail, J.-L. *Int. J. Quantum Chem.* **1996**, *58*, 153.
(9) Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1995**, *99*, 17516.
(10) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.
(11) Gao, J.; Freindorf, M. *J. Phys. Chem. A* **1997**, *101*, 3182.
(12) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. *J. Chem. Soc., Perkin Trans. 2* **1990**, 1259.
(13) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. *J. Chem. Soc., Perkin Trans. 2* **1991**, 225.
(14) Waszkowycz, B.; Hillier, I. H.; Gensmantel, N.; Payling, D. W. *J. Chem. Soc., Perkin Trans. 2* **1991**, 1819.
(15) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. *Int. J. Quantum Chem.* **1996**, *60*, 1189.
(16) Wesolowski, T.; Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1996**, *100*, 15444.
(17) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161.
(18) Warshel, A.; King, G. *Chem. Phys. Lett.* **1985**, *121*, 124.
(19) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647.
(20) Luzhkov, V.; Warshel, A. *J. Comput. Chem.* **1992**, *13*, 199.
(21) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100.
(22) Åqvist, J.; Fothergill, M.; Warshel, A. *J. Am. Chem. Soc.* **1993**, *115*, 631.
(23) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94, Revision D.2*; Gaussian, Inc.: Pittsburgh, PA, 1995.
(24) Hwang, J.-K.; King, G.; Creighton, S.; Warshel, A. *J. Am. Chem. Soc.* **1988**, *110*, 5297.
(25) Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269.
(26) Sharon, R.; Warshel, A. *QCFF/ALL*; Weizmann Institute of Science: Israel, 1973.
(27) Levitt, M.; Warshel, A. *ERFN/QM*; Weizmann Institute of Science: Israel, 1974.
(28) Ostlund, N. S. *Hyperchem*; Hypercube, Inc.: Cambridge, Ontario, 1985.
(29) Florián, J.; Warshel, A. *J. Phys. Chem B* **1997**, *101*, 5583.
(30) Alden, R. A.; Birktoft, J. J.; Kraut, J.; Roberts, J. D.; Wright, C. S. *Biochem. Biophys. Res. Commun.* **1971**, *45*, 337.
(31) Warshel, A.; Sussman, F.; Hwang, J.-K. *J. Mol. Biol.* **1988**, *201*, 139.
(32) Florián, J.; Warshel, A. *ChemSol*; University of Southern California: Los Angeles, 1997.
(33) Warshel, A.; Russell, S. *J. Am. Chem. Soc.* **1986**, *108*, 6569.
(34) Fuxreiter, M.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 183.