Ab Initio QM/MM Simulation with Proper Sampling: "First Principle" Calculations of the Free Energy of the Autodissociation of Water in Aqueous Solution

Marek Štrajbl, Gongyi Hong, and Arieh Warshel*

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

Received: July 17, 2002; In Final Form: September 26, 2002

Quantum mechanical calculations of activation free energies of chemical reactions in condensed phases present a major challenge for computational chemistry. On one hand, it is important to use high-level ab initio methods to obtain reliable results. On the other hand, it is essential to perform sufficient configurational sampling to obtain meaningful free energies. Although the advance of quantum mechanical/molecular mechanics (QM/MM) approaches has made this problem tractable, it still requires an enormous amount of computer time. The present work advances several strategies that allow one to perform practical ab initio QM/MM calculations of free energy profiles in solutions and proteins. The basic idea is the use of a simple reference potential for the ab initio calculations (e.g., Bentzien; et al. *J. Phys. Chem. B* **1998**, *102*, 2293). One version of this approach evaluates the free energy of transfer from the reference potential to the ab initio potential by a single step free energy perturbation (FEP) approach. A new version evaluates this free energy by the linear response approximation (LRA), which involves running trajectories on both the reference and the ab initio potentials. The performance of both approaches is examined by calculating the potential of mean force for the autodissociation reaction of water in solution. It is found that the LRA approach allows one to obtain reasonable results even in cases where the ab initio and reference potentials are significantly different. The present work also explores options for increasing the size of the quantum mechanical region. Here it is shown that the constrained DFT (CDFT) method provides a promising strategy. Finally, the general issue of modeling the autodissociation reaction by quantum mechanical approaches is briefly considered. It is pointed out that the use of the empirical valence bond (EVB) approach in the sampling process should provide a way for evaluating the elusive nonequilibrium solvation effect.

1. Introduction

Reliable simulations of chemical processes in solutions and in proteins present a major challenge. Hybrid quantum-mechanical/molecular mechanics (QM/MM) approaches[1−11] offer a promising option. However, at present there are major problems with the reliability of QM/MM approaches. That is, to obtain accurate potential surfaces one has to use high-level ab initio approaches. However, such approaches require significant computer time. This problem becomes very serious when one tries to face the challenge of evaluating activation free energies. Performing such calculations must involve very extensive configurational average that can make accurate studies computationally prohibitive. This paper develops strategies for ab initio calculations of free energies and activation free energies.

The main idea of our approach is related to previous works that used empirical potential surfaces as reference potentials for the ab initio calculation.[1,2] These works encountered difficulties in getting converging results in cases with large differences between the ab initio surface and the reference potential. However, here we introduce a major modification that allows one to use empirical reference potentials even if they are quite different from the ab initio potential.

In order to examine our approach we considered as a test case the autodissociation of water in aqueous solution. Simulating this fundamental process, which involves a creation of two separated charges in an originally neutral environment, is much more difficult than simulating proton migration in water.[12,13] Here one has to capture the compensation between the very large solvation contribution and the Coulombic energy of the charge separation process. The earliest molecular orbital QM/MM simulation of this process represented the solvent by the surface constraints soft sphere dipoles (SCSSD) model.[14] Problems with the asymptotic behavior of this QM/MM model led to the development of the empirical valence bond (EVB) model.[15] The EVB has been used very effectively in studies of proton transfer (PT) processes in water including the autodissociation reaction.[16] Using the EVB methodology has allowed us to reproduce the experimental results of Eigen and Mayer,[17] where the free energy for full proton transfer at infinite separation is around 21.4 kcal/mol and the activation barrier around 23.8 kcal/mol (see the analysis of ref 16).

Recent studies[18,19] that used the so-called first principle MD (FPMD)[18] have provided an instructive insight about the PMF of the autodissociation reaction and the related formation of the solvent separated ion pair.[19] However, these studies have been limited to the use of periodic boundary conditions with a relatively small unit cell. Such simulations might not be able to reproduce the solvent response in the actual bulk reaction. Furthermore, the calculated PMF does not reflect nonequilibrium solvation effects that might play an important role in determining the activation free energy.

Unfortunately, alternative ab initio QM/MM (QM(ai)/MM) approaches are not yet able to provide a practical tool for evaluating activation free energies of chemical processes in

* Corresponding author. E-mail: warshel@usc.edu.

10.1021/jp021625h CCC: $22.00
Published on Web 12/03/2002

solution. The main problem is the need for a sufficiently fast configuration average and thus for extremely long computational time. This problem is drastically reduced by the present work, where we upgrade the performance of QM(ai)/MM approaches by using the above-mentioned idea of a simple reference potential.

Section 2 describes our computational method, outlining the two strategies used to evaluate the free energy difference between the EVB and QM(ai)/MM surfaces. Section 3 examines the performance of our approaches in studies of the autodissociation reaction. This section also presents preliminary results of the CDFT calculations. Finally, we discuss in section 4 the general potential of our approach.

2. Methods

The goal of this work is to treat charge-transfer reactions in aqueous solution by a QM(ai)/MM approach while obtaining sufficient sampling for meaningful FEP calculations. In addition, we also explore ways for treating a relatively large part of the simulation system by an ab initio Hamiltonian. The methodology developed to meet these aims is described below.

2.1. QM(ai)/MM Surface and Direct PMF Calculations. The QM(ai)/MM approach expresses the total energy of the system as

$$E_{\mathrm{ai/MM}} = E^{0}_{\mathrm{ai}} + E_{\mathrm{int}} + E_{\mathrm{MM}} \tag{1}$$

where ai/MM replaces, for convenience, the notation QM(ai)/MM. The first term in eq 1 represents the energy of the quantum region (region I). The second term is the coupling between the quantum and classical regions. The third term corresponds to the energy of the classical region (region II) and is calculated by empirical potential functions. The energy of the first two terms is obtained using corresponding Hamiltonian by

$$E^{0}_{\mathrm{ai}} + E_{\mathrm{int}} = \langle \Psi | H^{0}_{\mathrm{ai}} + H_{\mathrm{int}} | \Psi \rangle \tag{2}$$

where H^{0}_{ai} is the Hamiltonian of the isolated quantum region (the gas phase Hamiltonian), H_{int} is the Hamiltonian describing interaction of the quantum system with the environment (region II), and ψ is the wave function of the combined Hamiltonians. H_{int} is given by

$$H_{\mathrm{int}} = H'_{\mathrm{int}} + V_{\mathrm{int}} = \sum_{i,j} \frac{Z_i q_j}{|\mathbf{R}_i - \mathbf{R}_j|} - \sum_{k,j} \frac{q_j}{|\mathbf{r}_k - \mathbf{R}_j|} + V_{\mathrm{int}} \tag{3}$$

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, and 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 point charges interacting with the QM system.

The interaction potential V_{int} is expressed as

$$V_{\mathrm{int}} = V_{\mathrm{vdW}} + V_{\mathrm{ind}} + V_{\mathrm{bulk}} \tag{4}$$

where V_{vdW} is the van der Waals interaction between the quantum region and its surrounding, V_{ind} represents the effect of the induced dipoles of the atoms in region II and V_{bulk} describes the effect of the solvent in the bulk region (the implicit continuum region around region II).

In principle, it is quite simple to include the induced dipoles of region II in the ab initio Hamiltonian in a self-consistent way (e.g., see ref 3). Here, however, we only considered the energy contributions from the polarization of the induced dipoles of the solvent, while neglecting their relatively small effect on the solute Hamiltonian.Thus, we use the following expression for the overall QM(ai)/MM energy

$$E_{\mathrm{ai/MM}} = \langle \Psi | H^{0}_{\mathrm{ai}} + H'_{\mathrm{int}} | \Psi \rangle + V_{\mathrm{vdW}} + V_{\mathrm{ind}} + V_{\mathrm{bulk}} + E_{\mathrm{MM}} \tag{5}$$

where the first term is calculated by standard ab initio programs (see below) and all other terms are evaluated using the ENZYMIX force field.[20]

In the general case the atoms in the quantum and classical regions may be bonded to each other. In such cases one has to use a linked atom (LA) or a related approach (see discussion in ref 1). In the present case, however, we are dealing with reaction in water and have no chemical bonding between the QM and MM regions. Thus, no LA treatment is required.

In principle and with sufficient computer time one can calculate a PMF profile directly for the QM/MM surfaces. This is usually accomplished by using some of the solute coordinates (e.g., the distance between the oxygen of the donor and the departing proton in our case) as a mapping parameter. In this way we can write (see also ref 2)

$$\exp\{-\beta \Delta G(R_i \rightarrow R_j)\} = \frac{Z(R_j)}{Z(R_i)} = \int \exp\{-\beta E(R_j,\gamma)\, \mathrm{d}\gamma\} / \int \exp\{-\beta E(R_i,\gamma)\, \mathrm{d}\gamma\} = \langle \exp\{-\beta [E(R_j) - E(R_i)]\} \rangle_{E(R_i)} \tag{6}$$

where R_i and R_j are two close values of the reaction coordinate, γ are the coordinates perpendicular to the reaction coordinate, and $\langle\ \rangle_{E(R_i)}$ designates and MD average over the potential surface with R fixed at R_i.

The QM/MM treatment has been implemented in the program MOLARIS.[20,21] The implementation involves a very convenient "weak coupling" philosophy where any QM package can be called by MOLARIS. Request for a QM calculation results in creating a file containing a snapshot of the system configuration at a given step. External QM program (e.g., Gaussian) reads this information and provides, in return, quantum mechanical energies, charges and forces. These charges and forces are then read back to MOLARIS and used to update the classical force field.

In this work, the QM region is treated by the DFT method using the B3LYP functional and a standard 6-31+G* basis set. The Gaussian 98[22] program is used to carry out the DFT calculations. The MM part (the water molecules in the classical region) is taken as the regular ENZYMIX force field.[20] The simulation model includes an 18 Å explicit simulation sphere, surrounded by a surface region whose average polarization and radial distribution are determined by the surface constrained all-atom solvent (SCAAS) model.[23] The surface region is embedded in a bulk continuum region with $\epsilon_{\mathrm{w}} = 80$. Long-range electrostatic effects are treated by the local reaction field (LRF) model.[24]

2.2. Using the EVB Surface as a Reference Potential for QM(ai)/MM-FEP Calculations. Our basic idea[2] is to have a reference potential that is as close as possible to the QM(ai)/MM surface and to use this reference potential for accelerating the convergence of the QM(ai)/MM FEP calculations. At present the best reference potential is provided by the EVB approach.[15,25]

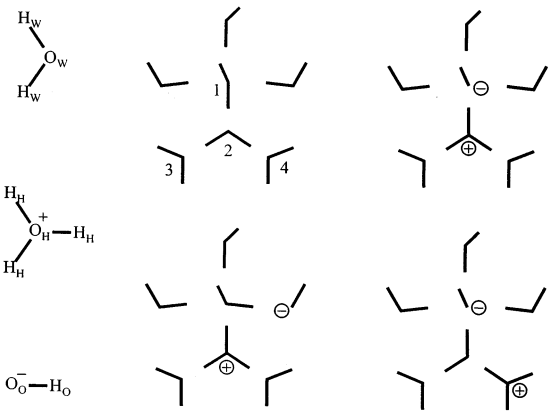

Figure 1. Illustrating the assignment of the EVB atom types and resonance forms for the 7 water molecules cluster. The atomic indices on the left part of the figure are used for the corresponding molecules and ions in the rest of the figure. Not all resonance forms are present and only the generic resonance structures are listed (e.g., the 7 waters system involved 7 resonance forms). Model systems with quantum regions of the 2, 3, and 4 water molecules comprised the (1, 2), (1, 2, 3), and (1, 2, 3, 4) subsets, respectively.

Because EVB was described extensively elsewhere (e.g., see refs 15, 16, and 25), we give here only the main points. The EVB ground state surface is obtained by mixing resonance structures (or more precisely diabatic states). Each resonance structure (RS) represents a specific bonding and charge configuration that has a well-defined interaction with the surrounding environment. Figure 1 summarizes the resonance structures used in this work.

The ground-state surface is obtained by solving the secular equation

$$\mathbf{HC}_g = E_g\mathbf{C}_g \tag{7}$$

where **H** is the EVB Hamiltonian of the system and $\mathbf{C}_g$ is the ground-state eigenvector. The matrix elements of **H** are described by analytical functions that can be calibrated by using experimental and theoretical data[25] and/or by fitting the E_g and ground-state charge distribution to ab initio results.[1,2,26] The solvent contributions to the Hamiltonian are calculated by evaluating the interactions between the solvent and the solute atoms in different resonance structures. Thus we describe the diagonal elements of the EVB Hamiltonian by analytical function of the form

$$H_{ii} = \epsilon_{ii} = V^{(i)}_{QM} + V^{(i)}_{QM/MM} + V_{MM} + \Delta^{(i)} \tag{8}$$

where V_{QM} is the EVB potential of the quantum region, $V_{QM/MM}$ is the coupling between the quantum and classical region, and V_{MM} is the force field of the classical region. V_{QM} is described by the analytical expression

$$V^{(i)}_{QM} = \sum_m D[1 - \exp\{-\beta(b^{(i)}_m - b^{(i)}_{0,m})\}] + \sum_m (K_\theta/2)(\theta^{(i)}_m - \theta^{(i)}_{0,m})^2 + \sum_{k,k'}(Ar^{-12}_{k,k'} - Br^{-6}_{k,k'}) + \sum_{k,k'} 332Q^{(i)}_k\, Q^{(i)}_{k'}(1 - \exp^{-\mu_s r_{k,k'^2}})/r_{k,k'} - \sum_{k,k'} 166\alpha/r^4_{k,k'} \tag{9}$$

Here the nonbonded and electrostatic interactions include only atoms that are not bonded to each other in the given resonance structure. The nonbond interactions between hydrogens that are bonded to the same oxygen are set to zero. The r^{-4} term is activated only between oxygens and it represents a very rough approximation to a true self-consistent inductive treatment that exists in some EVB implementations.[15] The coupling between the classical and quantum regions involves a 6-12 nonbond term and electrostatic interactions between the QM charges (the $Q^{(i)}$) and the classical water charges. The classical term V_{MM} is described by the standard SCAAS treatment (see above).

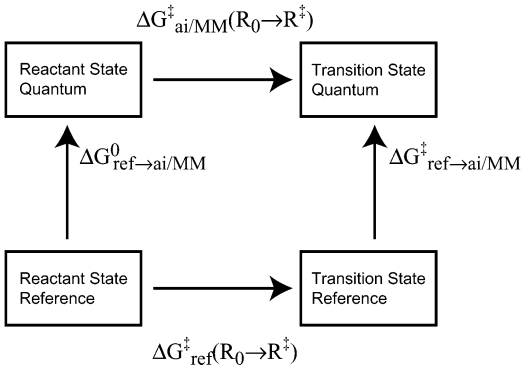

Figure 2. Thermodynamic cycle used to calculate the QM(ai)/MM activation free energy $\Delta G^{\ddagger}_{ai/MM}(R_0 \rightarrow R^{\ddagger})$.

The so-called "gas phase shift" $\Delta^{(i)}$ is a parameter that determines the energy of the fragments of the *i*th resonance form at infinite separation in solution.[25] This parameter is usually taken as a constant (obtained by forcing the corresponding asymptotic energies to reproduce the observed energy). However, in the present case, when we try to obtain a simultaneous fitting to ab initio and experimental information at both the infinite separation and intermediate distances, it was found convenient to use a distance dependent function

$$\Delta^{(i)} = \Delta^{(i)}_0 + \frac{\delta^{(i)}}{2}[1 + \tanh\{\xi^{(i)}(-R + \omega^{(i)})\}] \tag{10}$$

where R is the distance between the oxygens of the proton acceptor and the proton donor. The actual value of $\Delta^{(i)}_0$ was taken as the B3LYP/6-31+G* ab initio gas phase energy for the autodissociation reaction, when the reactant and product are held at infinite distance.

The off-diagonal elements between two states (i and j) coupled by a transfer of proton (m) from an oxygen (l) to another oxygen (n) is given by

$$H_{ij} = A_{ij}\exp\{-\alpha'_{ij}\mathbf{r}_{l,n} - \beta'_{ij}\mathbf{r}_{m,ln}{}^2 - \gamma'_{ij}(\theta - \theta'_0)^2\} \tag{11}$$

where $\mathbf{r}_{l,n}$ is the distance between two oxygens, $\mathbf{r}_{m,ln}$ is the distance of the proton from the midpoint of the two oxygens and θ, and θ'_0 are angles formed by the $l-m-n$ atoms. θ'_0 was taken from the ab initio optimized structure of the corresponding reference system. The parameters A, α', β', and γ' can be adjusted to reproduce the observed barrier for the reaction in solution or to reproduce gas phase ab initio calculations. In the present work we used gas phase DFT (B3LYP/6-31+G*) energies to refine the EVB force field.

With the EVB or any other reference potential we can obtain the QM(ai)/MM PMF free energy using the cycle of Figure 2. In this cycle we first evaluate the change of the EVB-PMF free energy $\Delta G^{\ddagger}_{ref}(R_0 \rightarrow R^{\ddagger})$ and then evaluate the free energies $\Delta G^{\circ}_{ref \rightarrow ai/MM}$ and $\Delta G^{\ddagger}_{ref \rightarrow ai/MM}$ of moving from the EVB surface to the QM/MM surface at the reactant and transition states.

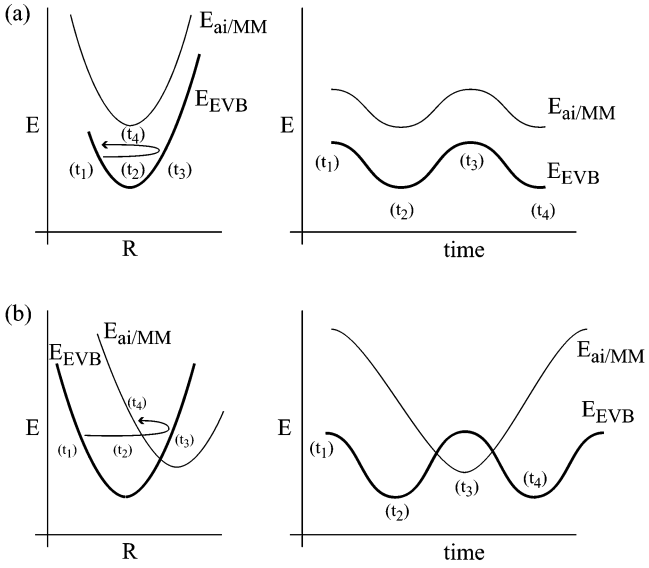

Figure 3. Illustrating the factors that determine the energy gap ($E_{EVB} - E_{ai/MM}$) for trajectories over the EVB surface. The upper figure shows the situation when the two surfaces are similar. The lower figure shows the situation when the two surfaces are significantly different. Obviously, it is much harder to obtain converging results for $\Delta G_{ref \rightarrow ai/MM}$ in the latter case.

These free energies are evaluated by a FEP treatment using

$$\exp\{-\beta \Delta G_{EVB \rightarrow ai/MM}(R_i)\} = \langle \exp\{-\beta[E_{ai/MM}(R_i) - E_{EVB}(R_i)]\}\rangle_{E_{EVB}(R_i)} \quad (12)$$

Note, however, that we are using a single step FEP procedure and collect configurations only from trajectories on the EVB surface. Now we obtain the PMF free energy of the QM(ai)/MM surface by

$$\Delta G(R \rightarrow TS)_{ai/MM} = \Delta G(R \rightarrow TS)_{EVB} + \Delta G_{EVB \rightarrow ai/MM}(TS) - \Delta G_{EVB \rightarrow ai/MM}(R) \quad (13)$$

Although this approach worked quite effectively in modeling solvation effects (i.e., mainly electrostatic interactions), it encountered difficulties in describing the free energy contributions from the solute fluctuations.[1] This problem reflected the fact that the ab initio and the EVB surface were not sufficiently similar (see Figure 3). There are several options that can help to solve this problem. One obvious option is to refine the EVB surface and thus to reduce the difference between the ab initio and EVB surfaces.

Another possibility is to add to the diagonal EVB energies constraint potentials

$$\epsilon_i' = \epsilon_i + V^i_{constr} \quad (14)$$

where V^i_{constr} is the potential added to ϵ_i. V^i_{constr} can be chosen in a way, for example, to minimize the difference between the EVB and ab initio cluster geometries. It is frequently more convenient to use simple constraint potentials than to refine the terms in eq 8.

In the present case we refined the EVB surface by considering several gas phase model systems and by fitting the EVB surfaces to the corresponding ab initio surfaces calculated by the DFT (B3LYP/6-31+G*) method. The systems considered in this fitting procedure included a gas phase proton transfer between H_3O^+ and H_2O, between OH^- and H_2O and between two water

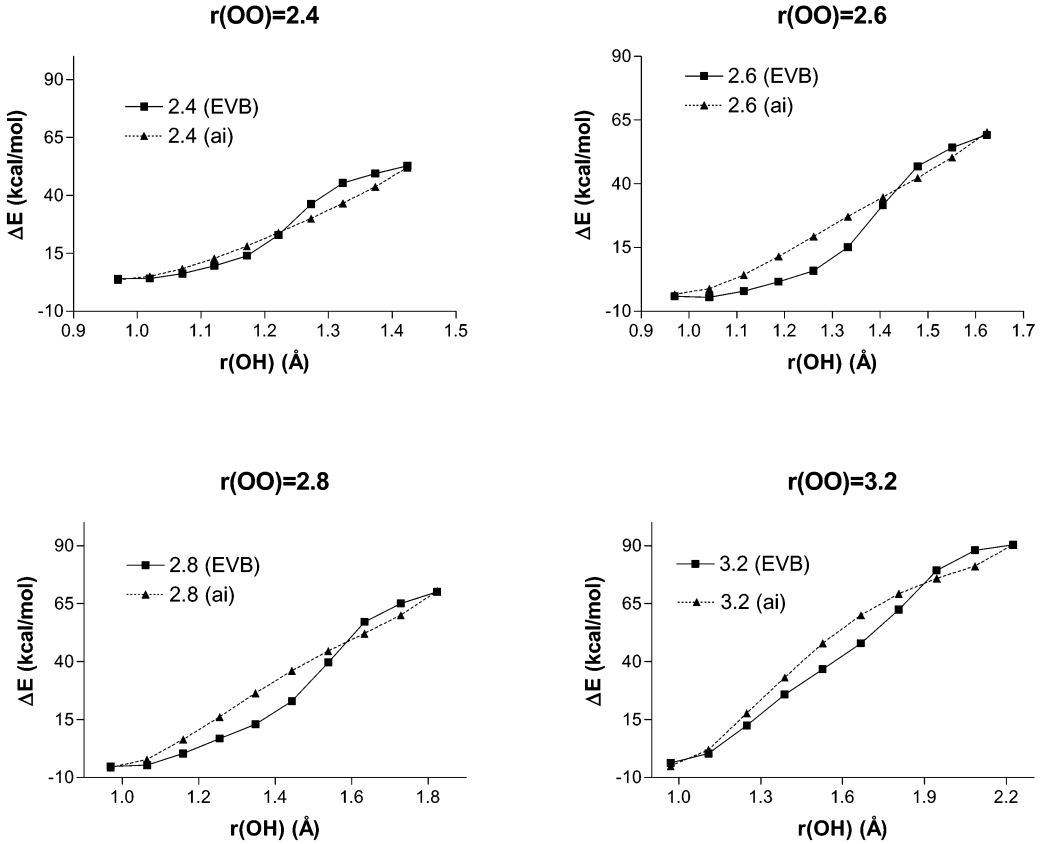

Figure 4. Showing the EVB reaction profile (solid line) and the corresponding ab initio profile (broken line) for the $H_2O + H_2O \rightarrow OH^- + H_3O^+$ gas phase system. The EVB parameters were refined by fitting the EVB surface to the corresponding ab initio B3LYP/6-31+G* surface for different O···H and O···O distances along the proton transfer coordinate.

TABLE 1: EVB Parameters[a]

bond parameters	D	b_0	β
O_W-H_W	120.0	0.988	2.0
O_H-H_H	80.0	0.988	2.2
O_O-H_O	102.0	0.975	2.0
angle parameters	$K/2$	θ_0	
O_W	70.0	106.5	
O_H	70.0	112.0	
O_O	70.0	106.5	
H_W	0.0	0.0	
H_H	0.0	0.0	
H_O	0.0	0.0	
nonbond (bonded in one or more RS)	A	B	
$O_W\cdots H_W$	250.0	0.0	
$O_W\cdots H_H$	30.0	0.0	
$O_O\cdots H_W$	250.0	0.0	
$O_O\cdots H_H$	250.0	0.0	
nonbond (never bonded)	A	B	
$O_W\cdots O_W$	500000.0	0.0	
$O_H\cdots O_W$	650000.0	0.0	
$O_H\cdots O_O$	300000.0	0.0	
$O_O\cdots O_W$	400000.0	0.0	
$H_H\cdots O_W$	30.0	0.0	
$H_O\cdots O_W$	230.0	0.0	
$H_W\cdots O_O$	200.0	0.0	
$H_H\cdots O_O$	70.0	0.0	
$H_W\cdots O_W$	100.0	0.0	
$H_W\cdots H_W$	300.0	0.0	
nonbond (with environment)	A	B	
O_W	774.0	24.0	
O_H	220.0	24.0	
O_O	700.0	24.0	
H_W	0.12	0.0	
H_H	4.0	0.0	
H_O	50.0	0.0	
inductive interaction	α		
$O_W\cdots O_W$	0.7		
$O_W\cdots O_O$	2.0		
$O_W\cdots O_H$	1.4		
elstat screening	μ_s		
$O_H\cdots O_W$	2.5		
$O_H\cdots H_W$	2.2		
$O_W\cdots H_H$	3.5		
$H_H\cdots H_W$	2.5		
$O_O\cdots H_W$	3.0		
$H_O\cdots H_W$	1.5		
$O_W\cdots O_O$	4.0		
$O_W\cdots O_W$	2.0		
$O_W\cdots H_W$	1.6		
$H_W\cdots H_W$	5.0		
$O_O\cdots H_H$	1.5		
$O_O\cdots O_H$	1.5		
$O_H\cdots H_O$	2.2		
EVB atom	**charges**		
O_W	−0.80		
O_H	−0.65		
O_O	−1.10		
H_W	0.40		
H_H	0.55		
H_O	0.10		

gas phase shifts	$\Delta_0^{(i)}$	$\delta^{(i)}$	$\xi^{(i)}$	$\omega^{(i)}$
2w ($i=2$)	210.0	6.0	0.6	7.0
3w ($i=2,3$)	210.0	13.0	0.6	7.0
4w ($i=2-4$)	210.0	20.0	0.6	7.0
7w ($i=2-7$)	210.0	45.0	0.6	7.0

off-diagonal H_{ij}'s	A	α'	β'	γ'	θ_0'
$OH^-\cdots H_3O^+$	150.0	0.55	2.0	0.0	175.0
$H_2O\cdots OH^-$	280.0	1.0	0.8	20.0	180.0
$H_2O\cdots H_3O^+$	160.0	0.9	0.0	0.0	0.0

[a] Notation of atoms according to Figure 1. The symbols used correspond to the symbols in eqs 9−11.

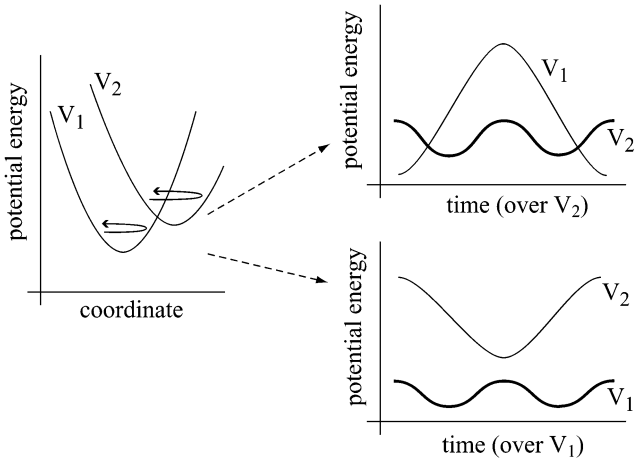

Figure 5. Schematic illustration of the LRA approach. This method evaluates the free energy associated with changing V_1 to V_2. This is done by running trajectories over V_1 (upper figure) and V_2 (lower figure), calculating $\langle\Delta V\rangle$ in each case and then using eq 15 to evaluate the relevant free energy.

molecules. The results of the fitting for the most important system (i.e., for proton transfer between two water molecules) are presented in Figure 4, and the complete set of EVB parameters is given in Table 1. As seen from the figure the fitting is not perfect and a more systematic refinement would have helped. Note, however, that we are dealing with a charge separation reaction that involves a major compensation by solvation energy. Modeling such a reaction in solution is much more challenging than modeling charge translocation reactions such as proton transfer between H_3O^+ and a water molecule (see refs 12 and 13). In the case of the autodissocation reaction it is probably essential to adjust the gas phase model upon moving to solution. Such parametrization is possible but it is somewhat out of the scope of the present paper, because we deal intentionally with cases where the agreement between the EVB and ab initio surfaces is not optimal (see below). In the present work we will only perform minor adjustments by modifying the gas phase shift parameter (the Δ of eq 8).

2.3. Improving the Convergence by Using the LRA Treatment. In cases where the difference between the EVB and ab initio surfaces is significant (see Figures 3 and 5) it might be impractical to use the expression of eq 13. In principle one can use a full FEP approach and move gradually from the EVB to the ab initio surface, but such an approach is too expensive. Thus we use here the LRA method[20] and evaluate the free energy $\Delta G^{\mathrm{LRA}}_{\mathrm{EVB}\to\mathrm{ai/MM}}(R)$ of transfer between the two potential surfaces E_{EVB} and $E_{\mathrm{ai/MM}}$ by

$$\Delta G^{\mathrm{LRA}}_{\mathrm{EVB}\to\mathrm{ai/MM}}(R) = {}^1/_2(\langle E_{\mathrm{ai/MM}} - E_{\mathrm{EVB}}\rangle_{\mathrm{EVB}} + \langle E_{\mathrm{ai/MM}} - E_{\mathrm{EVB}}\rangle_{\mathrm{ai/MM}})_R \quad (15)$$

where R designates a particular point along the reaction coordinate, e.g., the reactant state (RS) or the transition state (TS), and $\langle\,\rangle$ designates an average over the designated surface. Now, performing the average $\langle\,\rangle_{\mathrm{EVB}}$ is rather simple and the problem is to obtain a reasonable average over the QM(ai)/MM surface. Here we exploit the fact that the interaction between the reacting region and its surrounding is similar in the EVB and ab initio treatments. Thus, we use the following approximation

$$\langle \Delta E \rangle_{\mathrm{EVB}} \equiv \sum_{i}^{N} \langle \Delta E \rangle_{\mathrm{EVB}(\tau_0 + i\Delta\tau)}/N$$

$$\langle \Delta E \rangle_{\mathrm{ai/MM}} \equiv \sum_{j=m}^{M} \langle \Delta E \rangle_{\mathrm{ai/MM}(\tau_i + j\Delta\tau')}/(M-m) \qquad (16)$$

where $\Delta E = E_{\mathrm{ai/MM}} - E_{\mathrm{EVB}}$. The average $\langle \Delta E \rangle_{\mathrm{EVB}}$ is evaluated by running over the EVB surface a long relaxation time (τ_0) and then collecting ΔE from N configurations separated by an equal time interval ($\Delta\tau$). The average $\langle \Delta E \rangle_{\mathrm{ai/MM}}$ is evaluated by picking ΔE over a trajectory that was propagated on the EVB surface up to $\tau_i = \tau_0 + i\Delta\tau$ and then switched to the ai/MM surface. The trajectory is allowed to run on $E_{\mathrm{ai/MM}}$ for M steps ($M < N$). Only the last $M - m$ steps from the $E_{\mathrm{ai/MM}}$ trajectory are included in our average, which allows the trajectories to undergo a partial equilibration after switching from E_{EVB} to $E_{\mathrm{ai/MM}}$. It is assumed that only minor reequilibration is needed with regards to the "solvent" coordinates and that the equilibration in the solute coordinates is sufficient for our purpose. The use of this procedure is much less expensive than a full LRA treatment because the long run needed for convergence with regards to the solvent coordinates is done mainly on the EVB surface.

The values of N and M depend of course on the requirements for convergence of the calculated free energy and on the available computer time. Typically, our calculations involve 5000 points on the EVB surface (N in the above discussion) and from 300 to 1000 points on the ai/MM surface (M in the above discussion). The step size $\Delta\tau$ is typically 1 fs. The computer times needed for systems of different sizes will be discussed in section 3.5.

2.4. CDFT Treatments. Although the QM/MM approach is quite reasonable, there may be cases that require very large quantum regions to obtain reliable results. Here it might be too expensive to use standard ab initio approaches even with the EVB as a reference potential. A promising way of moving to large QM regions is offered by the constraint density functional theory (CDFT)[27,28] and its frozen density functional theory (FDFT) version.[27,30] Although the corresponding formulation is described elsewhere (e.g., see ref 27) we repeat here the main points to clarify the discussion of the current features.

The FDFT and CDFT approaches considers the reacting region (region I) and a part of its surrounding (region I′) by DFT formulation but freezes (or constrains) the densities of region I′. Thus we have

$$\rho(\mathbf{r}) = \rho_{\mathrm{I}}(\mathbf{r}) + \rho_{\mathrm{I}'}(\mathbf{r}) \qquad (17)$$

where we have N' and $N - N'$ electrons in region I and I′, respectively. Now, requiring that $\rho_{\mathrm{I}'}(\mathbf{r})$ obeys the constraint

$$\int \rho_{\mathrm{I}'}(\mathbf{r})\, \mathrm{d}\mathbf{r} = N - N' \qquad (18)$$

we can obtain $\rho_{\mathrm{I}}(\mathbf{r})$ by minimizing the energies of the total system using the Kohn−Sham formulation.[29] In this case we have

$$\rho_{\mathrm{I}}(\mathbf{r}) = \sum_{i=1}^{N} \varphi_i^*(\mathbf{r})\, \varphi_i(\mathbf{r}) \qquad (19)$$

where the $\varphi(\mathbf{r})$'s satisfy the modified Kohn−Sham equation[29]

$$[-{}^1/_2\nabla^2 + V_{\mathrm{eff}}^{\ddagger}(\mathbf{r})]\varphi_i(\mathbf{r}) = \epsilon_i \varphi_i(\mathbf{r}) \qquad (20)$$

and the effective potential $V_{\mathrm{eff}}^{\ddagger}(\mathbf{r})$ is given as

$$V_{\mathrm{eff}}^{\ddagger}(\mathbf{r}) = V_{\mathrm{eff}}^{\mathrm{KS}}(\mathbf{r}) + \frac{\delta T_s^{\mathrm{nadd}}[\rho_{\mathrm{I}}, \rho_{\mathrm{I}'}]}{\delta \rho_{\mathrm{I}}} \qquad (21)$$

Here, the potential $V_{\mathrm{eff}}^{\mathrm{KS}}(\mathbf{r})$ is given by

$$V_{\mathrm{eff}}^{\mathrm{KS}}(\mathbf{r}) = V_{\mathrm{ext}}(\mathbf{r}) + \int \frac{\rho_{\mathrm{I}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, \mathrm{d}\mathbf{r}' + \int \frac{\rho_{\mathrm{I}'}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, \mathrm{d}\mathbf{r}' + V_{\mathrm{xc}}[\rho(\mathbf{r})] \qquad (22)$$

where $V_{\mathrm{eff}}^{\mathrm{KS}}(\mathbf{r})$ is the exchange-correlation potential and $V_{\mathrm{ext}}(\mathbf{r})$ is the external potential. $T_s^{\mathrm{nadd}}[\rho_{\mathrm{I}}, \rho_{\mathrm{I}'}]$ is the nonadditive kinetic energy term that couples $\rho_{\mathrm{I}}(\mathbf{r})$ and $\rho_{\mathrm{I}'}(\mathbf{r})$, and is given by[30,31]

$$T_s^{\mathrm{nadd}}[\rho_{\mathrm{I}}, \rho_{\mathrm{I}'}] = T_s[\rho_{\mathrm{I}} + \rho_{\mathrm{I}'}] - T_s[\rho_{\mathrm{I}}] - T_s[\rho_{\mathrm{I}'}] \qquad (23)$$

Using the above partitioning of the kinetic energy, the Kohn−Sham total energy functional can be written as

$$E[\rho] = E[\rho_{\mathrm{I}} + \rho_{\mathrm{I}'}] = T_s[\rho_{\mathrm{I}}] + T_s[\rho_{\mathrm{I}'}] + T_s^{\mathrm{nadd}}[\rho_{\mathrm{I}}, \rho_{\mathrm{I}'}] + {}^1/_2 \int\int \frac{(\rho_{\mathrm{I}}(\mathbf{r}') + \rho_{\mathrm{I}'}(\mathbf{r}'))(\rho_{\mathrm{I}}(\mathbf{r}) + \rho_{\mathrm{I}'}(\mathbf{r}))}{|\mathbf{r}' - \mathbf{r}|}\, \mathrm{d}\mathbf{r}'\, \mathrm{d}\mathbf{r} + \sum_A \int \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|}(\rho_{\mathrm{I}}(\mathbf{r}) + \rho_{\mathrm{I}'}(\mathbf{r}))\, \mathrm{d}\mathbf{r} + E_{\mathrm{xc}}[\rho_{\mathrm{I}} + \rho_{\mathrm{I}'}] \qquad (24)$$

where the Kohn−Sham definition of $E_{\mathrm{xc}}(\rho)$ is retained and the explicit analytical form of the external potential is given. The above formulation for partitioning the electron density makes it possible to apply a hierarchy of practical implementations. In the CDFT class of approaches the electron density of all subsystems (region I and all the fragments in region I′) is obtained by minimizing the total energy. In this freeze-and-thaw procedure the electron density of each subsystem enters into eqs 21−23 in an alternating way as $\rho_{\mathrm{I}}(\mathbf{r})$ and $\rho_{\mathrm{I}'}(\mathbf{r})$ until convergence is reached.[32] In the FDFT approach only the electron density of the system of interest (region I) is obtained from eq 19. This approach is obviously less computationally expensive than the CDFT approach, because it restricts the minimization problem to a selected region. Although the FDFT is less rigorous than the CDFT method it still provides a first principle embedding scheme, where the effective potential of an embedded electron density is given by eq 21. Because in FDFT approaches the embedding electron density is given as an input, various implementations of FDFT are possible differing in the way the frozen electron density is obtained. Both the CDFT and FDFT approaches were implemented in the deMon program.[33] The Becke88 exchange potential, Perdew86 correlation potential, and the nonadditive kinetic energy functional of eq 23 were used in all the calculations. Standard 6-31+G* and 6-31G* basis sets were used for regions I and I′, respectively.

Our general treatment involves the incorporation of the classical force field (region II) of the MOLARIS program in the CDFT Hamiltonian in the same way considered in section 2.1 for the regular ab initio Hamiltonian. The total potential surface is obtained as in any QM/MM approach from the sum of the CDFT energy (region I and I′) and the "self-energy" of region II, making sure that the interaction with the CDFT region is not counted twice. Thus, the van der Waals interactions between region II and regions I′ and I are calculated by the ENZYMIX force field of MOLARIS, whereas the electrostatic

interaction between region II (point charges) and region I and I′ (electron densities) are calculated by the CDFT program. Another important point is the treatment of the interaction between the groups in region I′ (the water molecules). These interactions are treated classically by the ENZYMIX force field (point charge − point charge electrostatic interactions and van der Waals interactions) whereas the interaction between region I′ and I is treated quantum mechanically. As explained previously (e.g., see ref 28) the classical treatment of the interactions between the groups in region I′ is the main saving factor of the CDFT approach. It is also important to emphasize that this is a well-defined consistent embedding model where the interaction of the full quantum mechanical region (region I) with its immediate surrounding is evaluated by what is probably the most consistent current coupling scheme, whereas the interactions between the surrounding groups are evaluated classically. At any rate, the implementation of the QM/MM-MD simulation treatment of MOLARIS requires us to evaluate the forces of the CDFT regions. The corresponding procedure is outlined below.

The energy and forces are calculated in a three step procedure. In the first step, the zero order density of region I is obtained by a QM/MM method, in which regions I′ and II are treated classically. In the second step a CDFT calculation is carried out for each group in region I′ where the electron densities are obtained by SCF iteration (typically 15 steps) whereas the electron densities of region I are frozen and the interaction with other groups in region I′ are excluded (these interactions are included, however, in the calculations of the forces). In the third step the calculations are repeated for region I whose electron density is obtained by SCF iterations (typically 15 steps) and the electron density of region I′ is frozen. The forces in region I are calculated analytically in the third step by differentiating the total energy (eq 24) with respect to the coordinates of the atoms in region I. The forces on each group in region I′, due to the interaction between the given group and region I as well as the interaction between the atoms inside this group, are calculated analytically in the second step, by differentiating the total energy (eq 24) with respect to corresponding coordinates. In doing so we evaluate the forces due to the interaction between region I and I′ by differentiating the nonadditive kinetic energy functional term and the terms due to electrostatic interaction between region I′ and I. The differentiation of the self-energy of each group in region I′, with regard to the coordinates of its atoms is done analytically when the corresponding group is treated as region I in the freeze and thaw procedure of the second step. Note in this respect, that it is essential to evaluate all the relevant integrals within each group in order to obtain the correct quantum forces. This calculation is easily implemented if we use a standard SCF procedure for the given group. This consideration is, in fact, the main reason we use a CDFT rather than FDFT procedure. Fortunately, these calculations are not so expensive because typically the groups in region I′ are small (here they are simply taken as individual water molecules). The calculations of the integrals between different groups in region I′ are not needed, however, because we treat these interactions classically. As stated above, this classical treatment is the main time saving element in the CDFT approach.

The above treatment presents a consistent CDFT treatment because in each MD step we continue from the density of region I, obtained in the previous step in the freeze and thaw procedure. The validity of this approach was established by verifying that the calculated numerical forces provide a very good approximation for the corresponding analytical forces

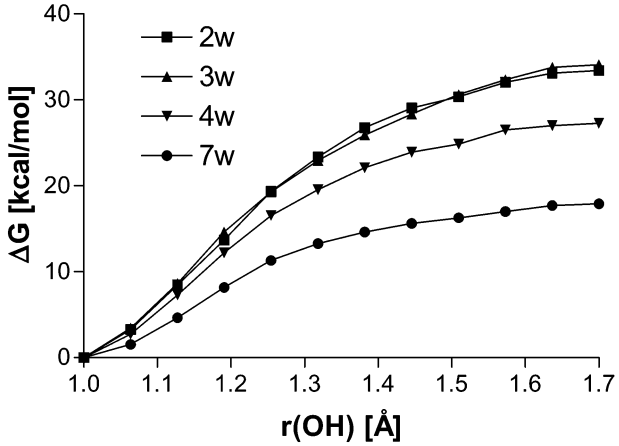

Figure 6. Direct QM/MM PMF profiles for different sizes of the quantum region.

3. Results

3.1. Direct QM(ai)/MM Calculations. As stated above, we took the autodissociation of water in water as a test case for our approaches. Now, in order to establish a benchmark for our calculations with the EVB as a reference potential, we found it useful to evaluate PMF profiles by direct QM(ai)/MM calculations. Because this approach involves evaluation of QM(ai) energies for each MD step along the whole reaction coordinate, it is likely to be prohibitive for systems with large quantum regions. Fortunately, for our small QM regions ranging from 2 to 7 water molecules, we could evaluate the actual QM(ai)/MM PMF surface in a direct way (although this required significant computer time). In order to facilitate the direct calculations we started by running MD trajectories on the EVB surface and generating configurations where the solvent is already relaxed in response to the solute charges. These configurations were then used as starting points for the QM(ai)/MM PMF calculations. After examination of the preliminary calculations of the 7 water molecules cluster, we decided to explore the PMF surface along a reaction coordinate defined by the O···H distance. The PMF was evaluated while changing the reaction coordinate from 1 Å to 1.7 Å. Increasing the O···H distance beyond 1.7 Å lead to a proton transfer from O_2 to O_3 (in the notation of Figure 1) and eventually back to O_1 and thus back to the reactant state. Thus, we assigned the transition state as the point where the reaction coordinate is equal to 1.7 Å (see also ref 18). The calculated PMF profiles for different QM/MM models (differing by the size of the QM regions) are summarized in Figure 6. The computer time needed for these calculations will be discussed in section 3.4.

The calculated results should be compared to the experimental estimates (see ref 16 for the analysis of the relevant experiments) of $\Delta G \sim 21$ kcal/mol and $\Delta G^{\ddagger} \sim 24$ kcal/mol. As seen from Figure 6, the barrier becomes more reasonable when the number of water molecules is increased. Note, however, that the PMF barrier for the 7 water molecules quantum system is lower than the observed barrier. But this might be a correct result for an "incorrect" treatment. That is, as will be discussed in section 4, the PMF treatment neglects the nonequilibrium solvation barrier that corresponds to the work that has to be invested into the reorganization of the solvent (see refs 2 and 34). Interestingly, a FPMD treatment[18] obtained a similar underestimate. At any rate, although in this work we focus on the ability to evaluate

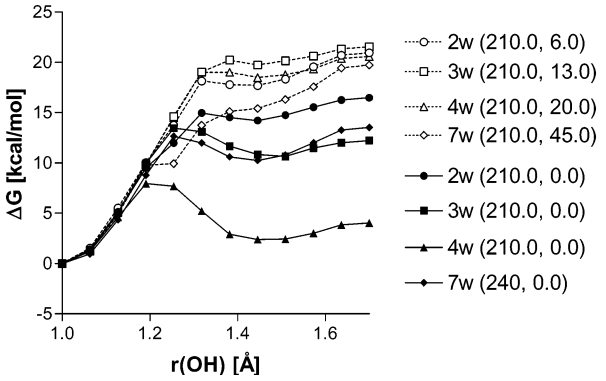

Figure 7. EVB PMF profiles for different sizes of the quantum region. The gas phase shifts parameters ($\Delta^{(i)}$ and $\delta^{(i)}$ of eq 10) used for each model are given in brackets. The EVB surfaces denoted by filled symbols were used as reference potentials for the QM(ai)/MM calculations. The profiles denoted with empty symbols are the potential surfaces obtained with gas phase shifts tuned to reproduce the experimentally observed $\Delta G°$ of the autodissociation reaction.

QM(ai)/MM free energies of different models rather than the validity of these models, we will discuss in section 4 the significance of the calculated $\Delta G^{\ddagger}$.

3.2. EVB Calculations. Several models were explored with different numbers of water molecules in the quantum region. In the present case we performed for simplicity PMF calculations rather than the special umbrella sampling procedure, which is usually implemented in EVB studies (e.g., see refs 1 and 2). The umbrella sampling procedure is more rigorous because it captures nonequilibrium solvation effects[1,26] (see also section 4), but it is hard to implement this treatment in QM(ai)/MM studies and thus it was not used in the present work. The PMF calculations used the O···H distance of the transferred proton as the mapping parameter. The corresponding EVB PMF profiles are given in Figure 7. These profiles were evaluated with the specified EVB model and with the SCAAS/LRF boundary condition. The calculations involved 11 windows of 5000 time steps each (total of 55 ps simulation time) in a 18 Å water sphere. Obviously, we can easily force the EVB surfaces to reproduce the proper "observed" PMF activation barrier by tuning the value of the gas phase shift. Nevertheless, it was found that at present we need different gas phase shifts for different sizes of the quantum region. Obtaining a more consistent EVB surface is clearly possible with more careful parametrization and better analysis of the coupling between the different resonance structures of the larger clusters. However, this is not the aim of the present work. In fact, at least formally, the result of our LRA approach should not depend on the EVB surface (of course, the closer the EVB and QM(ai) surfaces are the better will be the convergence). To demonstrate this point, we intentionally used the EVB surfaces obtained with less reliable gas phase shifts to examine the performance of our LRA approach.

3.3. Using the EVB Surface as a Reference Potential. With the EVB potential surface we are ready to evaluate the free energies of the QM(ai)/MM surfaces (see section 2.1).

As a first step in our attempt to evaluate the QM(ai)/MM free energy surface, with the EVB as a reference potential, we examined the difference between E_{EVB} and $E_{ai/MM}$ along trajectories on the EVB surfaces. Typical behavior is shown in Figures 8 and 9, which were obtained for a cluster of 2 water molecules in region I.

Although the ab initio and EVB fluctuations are similar, they may not be sufficiently close to allow a reliable estimate of

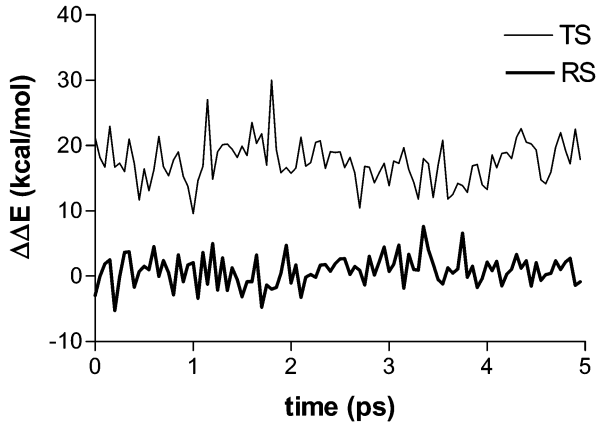

Figure 8. Fluctuations of the ($E_{ai/MM} - E_{EVB}$) energy gap for a system with 2 water molecules in region I. The fluctuations were obtained for trajectories on the EVB surfaces of the RS and TS regions.

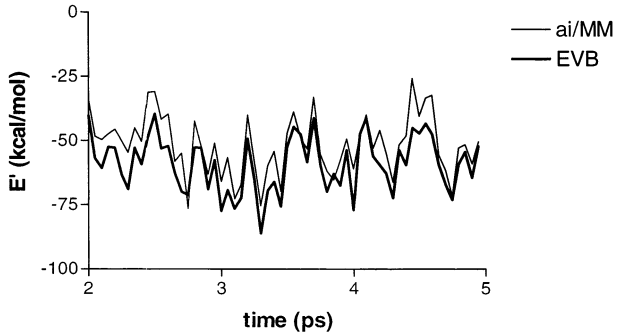

Figure 9. Fluctuations of $E'_{ai/MM}$ (broken line) and E'_{EVB} (solid line) of a system with 2 water molecules in region I. The fluctuations were obtained for trajectories over the EVB surfaces at the TS region. E' designates the energy of the quantum region plus the interaction between the quantum and classical regions.

$\Delta G_{EVB\rightarrow ai/MM}$ by eq 12. This will clearly be the case in situations that start to resemble Figure 3b. In such cases we will need extremely long trajectories to properly capture the chance of being at the minimum of the ab initio surface because our system spent most of the time near the minimum of the EVB surface. For larger QM regions it is likely that the difference between the ab initio and EVB surface will be larger because the present fitting procedure focused only on the interaction between pairs of molecules. Indeed, this point is illustrated by the results presented in Table 2. The free energies calculated by eq 13 are quite different than the benchmark energies $\Delta G^{direct}_{PMF(ai/MM)}$ in case of systems with 4 and 7 water molecules.

Of course, one can improve the results obtained by eq 12 by refining the EVB surfaces to minimize the difference between the EVB and ab initio surfaces. The refinement can be done by adjusting of the EVB parameters or by adding correction potentials to the diagonal EVB elements. However, as stated repeatedly above, in this work we like to explore a more general strategy that would not require special effort in optimization of the EVB parameters. Thus, we examined the performance of the LRA approach of eq 15.

The results of the LRA approach are summarized in Table 3. As seen from the table, this approach appears to be more robust and provides for better convergence for the bigger clusters. Unlike the approach based on eq 12, the LRA method is capable of handling the large fluctuations in the energy gap between the EVB and ai/MM surfaces for the systems of 4 and 7 water molecules. The necessity to run additional calculations to obtain the energy gap for trajectories over the ai/MM surface

TABLE 2: QM(ai)/MM PMF Activation Barrier, $\Delta G^{eqs12,13}_{PMF(ai/MM)}$, for the Autodissociation Reaction Obtained by the Single Step FEP Method of Eqs 12 and 13[a]

system	$\Delta G_{PMF(EVB)}$	$\Delta G^{eq12}_{R(EVB\rightarrow ai/MM)}$	$\Delta G^{eq12}_{TS(EVB\rightarrow ai/MM)}$	$\Delta G^{eqs12,13}_{PMF(ai/MM)}$	$\Delta G^{direct}_{PMF(ai/MM)}$
2w	16.5	−0.5	12.6	29.6	33.4
3w	12.0	−1.0	22.7	35.7	34.1
4w	4.1	−15.2	18.4	37.7	27.3
7w	13.5	−6.7	8.2	28.4	17.9

[a] Energies in kcal/mol. $\Delta G^{direct}_{PMF(ai/MM)}$ is the barrier obtained by direct QM(ai)/MM calculations (see section 3.1).

TABLE 3: QM(ai)/MM PMF Barrier, $\Delta G^{eqs13,15}_{PMF(ai/MM)}$, for the Autodissociation Reaction Obtained by the LRA Approach of Eq 15[a]

		reactant		transition state			
system	$\Delta G_{PMF(EVB)}$	$\langle E_{ai/MM} - E_{EVB}\rangle_{EVB}$	$\langle E_{ai/MM} - E_{EVB}\rangle_{ai/MM}$	$\langle E_{ai/MM} - E_{EVB}\rangle_{EVB}$	$\langle E_{ai/MM} - E_{EVB}\rangle_{ai/MM}$	$\Delta G^{eqs13,15}_{PMF(ai/MM)}$	$\Delta G^{direct}_{PMF(ai/MM)}$
2w	16.5	0.9	−5.0	17.0	3.0	28.6	33.4
3w	12.2	1.0	−5.0	30.0	12.0	35.2	34.1
4w	4.1	17.9	−17.0	51.6	4.0	31.5	27.3
7w	13.5	18.0	−14.4	39.9	−26.0	18.6	17.9

[a] Energies in kcal/mol.

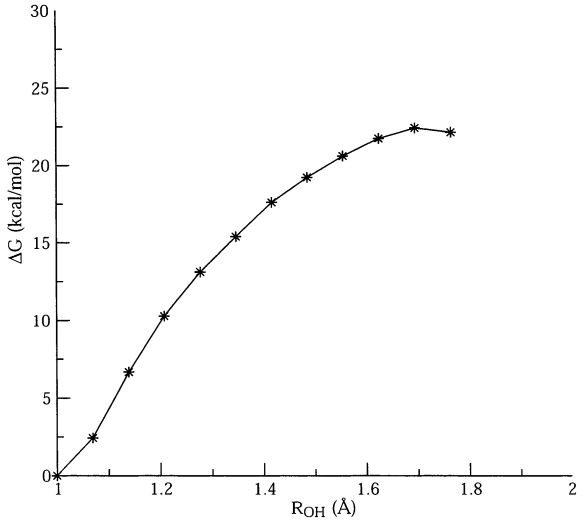

Figure 10. A QM(CDFT)/MM for a system with 10 water molecules in the quantum region. The profile was obtained by direct PMF calculations.

(see eq 15) is a relatively small price for achieving convergence in the calculated free energy.

3.4. CDFT Prescription for a Large QM Region. As is clear from the examples above, the size of the QM region may have a significant effect on the calculated free energy surfaces. One of the promising ways of increasing the size of the QM region is to use the CDFT embedding method (see section 2.4). In the present case we used the CDFT to study the system of 7 water molecules in the QM region. This was done with a QM-(CDFT)/MM model where the QM model included 2 water molecules in region I (the region with a full DFT treatment) and 8 water molecules in region I′ (the region with constrained densities). Regions I and I′ were embedded in a classical MM region modeled by the regular SCAAS water model. The QM-(CDFT)/MM model was used in direct PMF calculations and the corresponding results are described in Figure 10. The activation barrier obtained here is about 23 kcal/mol, which is a few kcal/mol larger than the result obtained by the PMF of the full DFT QM(ai)/MM treatment (∼18 kcal/mol). The overestimate might reflect the absence of charge transfer between region I and region I′. However, this deficiency can be reduced by the approach used in our previous work,[28] which allows charge equilibration between regions I and I′ at the beginning of each PMF mapping step. At any rate, the main point of the present treatment is a demonstration that the CDFT method can be used in direct PMF calculations with modest computational resources. The CDFT treatment was about 8 times faster than the corresponding full DFT treatment when using the same integration grid.

TABLE 4: CPU Times for the Different Types of PMF Calculations Reported in This Work[a]

QM system	method[b]	windows	steps	time/step (min)	tot. time (min)
7w	EVB	20	5000	0.004	400
2w	QM(ai)/MM	11	600	0.6	3960
3w	QM(ai)/MM	11	500	1.7	9350
4w	QM(ai)/MM	11	1000	2.8	30800
7w	QM(ai)/MM	11	300	13.3	43890
7w	QM(ai)/MM	20	5000	13.3	13300000[c]
10w	CDFT	11	700	3.0	23100
7w	eq 12	$20_{EVB} + 2_{QM}$	300[d]	13.3	8380
7w	eq 15	$20_{EVB} + 4_{QM}$	300[d]	13.3	16360

[a] The reported values were obtained on IBM Netfinity Cluster equipped with Pentium 3 1.2 GHz processor. [b] The QM(ai)/MM method treated region I by the B3LYP/6-31+G* method implemented in Gaussian 98. The CDFT calculation treated region I and I′ by the approach described in section 2.4. A modified version of the deMon program was used for CDFT calculations. [c] A hypothetical calculation where preequilibration using EVB method was not used. [d] The number of steps per window corresponds to the evaluation of the QM energy (the number M in eq 16).

3.5. Assessing the Time Savings of the LRA Approach. The main purpose of the present work is to develop practical ways for QM(ai)/MM free energy calculations. The time saving of our approaches can be assessed by considering the summary given in Table 4 and the discussion below.

To have a clear perspective, we should start with the time needed for a proper convergence of the PMF calculations of the test system with 7 water molecules in region I (which is the smallest system that gives a reasonable barrier). For a typical PMF we need 20 windows (each step in eq 6) of 5000 steps for each window. Using Table 4, we can see that this would 1.3×10^6 min. Because this is an enormous amount of time, we obtained the results reported in Figure 7 by starting the PMF from relaxed solvent configurations obtained from the EVB calculations. We also used a minimal number of steps (300 instead of 5000) and 11 windows instead of 20. This is the lowest limit for reasonable convergence and it can only be applied for cases with relatively small geometrical changes in the reaction coordinate. Now, if we use the method of eq 12, we have to evaluate the mapping from the EVB to the QM/MM surfaces in the reactant and transition states. Thus, the total

time for the system with 7 water molecules in region I is 400 min for obtaining the full EVB PMF reference surface and $2 \times M \times 13.3$ min (8380 min when $M = 300$) for the QM calculation. This approach can be used as long as the EVB and QM(ai) surface are similar; otherwise, we have to use the LRA treatment. The use of the LRA approach requires to perform calculation of $4 \times M \times 13.3$ min. Thus with $M = 300$ we obtain $15960 + 400 = 16360$ min. This is a factor of 80 less than the time needed for a full convergence of the QM(ai)/MM PMF calculations. Now, relative to the lower limit used in the present direct QM(ai)/MM calculations (which started from the relaxed EVB configurations) the saving is lowered to a factor of 3. However, for the PMF calculation that requires more windows, the time saving by the LRA treatment would be much more significant. Furthermore, when the EVB and QM(ai)/MM surface are more similar, we need fewer steps in the LRA calculations. Finally, as seen from Table 4, we can achieve additional saving by using the CDFT treatment. Of course, the time saving will increase when we consider more water molecules in the QM regions (region I and I′).

4. Concluding Remarks

This work addresses the challenge of obtaining QM(ai)/MM free energies by developing and examining several alternative approaches. Two approaches involve the use of the EVB surface as a reference potential for the evaluation of the ab initio free energy. The first approach follows our previous strategy[2,27] and evaluates the free energy of moving from the EVB to the ai/MM surface by running trajectories on the EVB surface and then using a single step FEP method to obtain the relevant free energy. This approach was found to give reasonable results for cases where the reference potential and the ab initio surfaces are similar. The second approach involves the use of the LRA formulation to overcome the convergence problems that occur in cases of large difference between reference potential and the ab initio surface. The use of this approach appears to provide a promising general stragety for ab initio free energy calculations. Finally, the use of the CDFT in direct QM(ai)/MM PMF calculations also provided encouraging results, demonstrating our ability to treat a large quantum region by a fast yet reasonably accurate approach. Furthermore, it should be possible to implement the CDFT method as a reference potential for the full QM(ai)/MM treatment.

Whereas most of this work involved methodological aspects, the present findings provide instructive hints about the autodissocation reaction. Although this is one of the most fundamental reactions in nature,[17] its rigorous modeling is still extremely challenging. The EVB method has allowed us to model this and related reactions for a long time (e.g., see refs 15 and 16). This was done, however, on the semiempirical level and the real challenge is to provide a quantitative picture on the ab initio level. Small cluster calculations[35] gave reasonable energetics, but such studies do not represent the behavior of the system in the bulk (see below). FPMD studies have provided an instructive insight[18,19] but could not reproduce the correct barrier. That is, the calculated barrier is about 16 kcal/mol and the observed barrier is ~24 kcal/mol. This problem might reflect several factors. First, and most importantly, the FPMD calculations evaluate the PMF free energy for the reaction and this does not include the nonequilibrium solvation effect (see ref 2). The reason is that in the PMF calculations the reaction coordinate involves only solute coordinates (the solvent is allowed to equilibrate at each point of the reaction coordinate). Consequently, the contribution corresponding to its reorganization is not included in the energy of the activation barrier. Second, the periodic model used in the FPMD calculations involve relatively few water molecules, which might not reflect the proper bulk solvation effect (see ref 36 for possible problems with the FPMD energetics).

An attempt to explore the nonequilibrium transition state by propagating trajectories toward the transition state was reported by Geisller et al.[37] Unfortunately, this approach did not evaluate the important contribution of the nonequilibrium solvation barrier to the overall activation free energy. We believe that the EVB approach provides a much more effective way of exploring transition states (TS) in solution than the method used in ref 37. That is, the main advantage of the EVB approach is in having a clear physical concept that guides the solvent from its reactant to product configuration. In this way one can easily drive the system to any point in the combined solute−solvent hypersurface. This provides a general way for exploring nonequilibrium solvation effects (see ref 34). In this approach we fix the solute coordinates at the maximum of the PMF and perform the EVB umbrella sampling procedure[25] to find the TS in the complete solute−solvent space. For example, for the EVB surface of the system with 7 water molecules in the quantum region we obtained a nonequilibrium solvation barrier of 4 kcal/mol. The challenging extension of this method to the QM(ai)/MM surfaces is now in progress in our laboratory.

Acknowledgment. This work was supported by NSF grant MCB-0003872 and NIH grant GM 24492. We are grateful to Dr. Wesolowski for useful discussions. We thank the USC's High Performance Computing and Communication Center (HPCC) for computer time.

References and Notes

(1) Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 2293−2301.
(2) Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1995**, *99*, 17516−17524.
(3) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(4) Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269−282.
(5) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483−3492.
(6) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298−305.
(7) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580−10594.
(8) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
(9) Friesner, R.; Beachy, M. D. *Curr. Op. Struct. Biol.* **1998**, *8*, 257−262.
(10) Monard, G.; Merz, K. M. *Acc. Chem. Res.* **1999**, *32*, 904−911.
(11) Field, M. *J. Comput. Chem.* **2002**, *23*, 48−58.
(12) Vuilleumier, R.; Borgis, D. *Chem. Phys. Lett.* **1998**, *284*, 71−77.
(13) Schmitt, U. W.; Voth, G. A. *J. Phys. Chem. B* **1998**, *102*, 5547−5551.
(14) Warshel, A. *J. Phys. Chem.* **1979**, *83*, 1640−1650.
(15) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218−6226.
(16) Åqvist, J.; Warshel, A. *J. Mol. Biol.* **1992**, *224*, 7.
(17) Eigen, M.; Maeyer, L. d. *Z. Elektrochem.* **1955**, *59*, 986−993.
(18) Trout, B.; Parrinello, M. *J. Phys. Chem. B* **1999**, *103*, 7340−7345.
(19) Sprik, M. *Chem. Phys.* **2000**, *258*, 139−150.
(20) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161−185.
(21) Chu, Z. T.; Villa, J.; Strajbl, M.; Schutz, C. N.; Shurki, A.; Warshel, A. "MOLARIS version beta9.05," University of Southern California, in preparation.
(22) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen,

W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.11.1; Gaussian, Inc.: Pittsburgh, PA, 2001.

(23) King, G.; Warshel, A. *J. Chem. Phys.* **1990**, *93*, 8682−8692.

(24) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100−3107.

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

(26) Hwang, J.-K.; King, G.; Creighton, S.; Warshel, A. *J. Am. Chem. Soc.* **1988**, *110*, 5297−5311.

(27) Wesolowski, T.; Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1996**, *100*, 15444−15449.

(28) Hong, G.; Strajbl, M.; Wesolowski, T. A.; Warshel, A. *J. Comput. Chem.* **2000**, *21*, 1554−1561.

(29) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.

(30) Wesolowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993**, *97*, 8050−8053.

(31) Cortona, P. *Phys. Rev. B* **1991**, *44*, 8454.

(32) Wesolowski, T. A.; Weber, J. *Chem. Phys. Lett.* **1996**, *248*, 71.

(33) St-Amant, A.; Salahub, D. R. *Chem Phys. Lett.* **1990**, *169*, 387.

(34) Villa, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 7887−7907.

(35) Siegbahn, P. *J. Comput. Chem.* **1995**, *17*, 1099−1107.

(36) The problem with the FPMD solvation contribution is expected to increase with a larger charge separation. This might be the case with the energetics of the solvent separated ion pair that was calculated to be around 17.5 kcal/mol and assumed to be in a good agreement with the observed pK_a of liquid water.[19] Actually, the observed pK_a corresponds to reaction free energy 21.4 kcal/mol[16] and this energy should be evaluated with the relevant ion pair at an infinite separation. The analysis of ref 19 probably ignored the effect of the concentration of liquid water (55 M) on the equilibrium constant of the autodissociation reaction.

(37) Geisller, P. L.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. *Science* **2001**, *291*, 2121.