MIXED QUANTUM MECHANICAL/MOLECULAR MECHANICAL SIMULATIONS OF CHEMICAL REACTIONS IN SOLUTION AND IN ENZYMES BY THE CLASSICAL TRAJECTORY MAPPING APPROACH

JIANN-JONG PAN & JENN-KANG HWANG*

Department of Life Sciences
National Tsing Hua University
Hsin Chu 300, Taiwan

We present a practical hybrid quantum mechanical/molecular mechanical approach to study chemical reactions in solution and in enzymes. In this method, referred to as the "Classical Trajectory Mapping" method, trajectories are calculated on the classical potential surfaces and, by using the classical surfaces as a reference state for the actual quantum mechanical ground state potential, the free energy profile of the chemical reaction is obtained by the free energy perturbation technique. This method was applied to proton-transfer reactions both in aqueous solution and in papain. The encouraging results indicate the applicability of our method to chemical reactions in the condensed phase and the biological systems.

Introduction

To understand enzyme catalysis is one of the most important problems in biochemistry. The capability to compute the free energy profiles of enzymatic reactions will offer a useful insight into how enzymes work. Despite great advances in both computer technology and computational methodology, it is still impossible to study chemical reactions in enzymes by a fully quantum mechanical(QM) representation. Approximations[1-6], therefore, have to be introduced in one way or another in order to obtain feasible results in a reasonable amount of computational time. One usual approximation is to divide the system in two parts: the reaction region consisting of the atoms participating in the bond-breaking or bond-forming reactions is treated quantum mechanically, while the surrounding environment such as water or enzymes is treated by molecular mechanics(MM). This approach alleviates the computational loads by avoiding quantum mechanical calculations on the atoms in the classical region, but in the case of large dynamics simulations, because the

quantum mechanical forces and energies have to be evaluated at each iteration, the computational time required for the quantum mechanical calculations of even a relatively small number of atoms is still quite long. Here we propose an efficient approach which avoids quantum mechanical calculations at each iteration step. In this approach, referred to as the Classical Trajectory Mapping (CTM) method, the trajectories propagates only on a classical reference potential surface and the free energy profile is calculated by the free energy perturbation (FEP) method which maps the reference state to the actual ground state described by the semiempirical AM1 method. However, it should be noted that our strategy is quite general and will not be limited to semiempirical potential surfaces. In fact, more sophisticated *ab initio* potential surfaces can also be incorporated into our approach.

Methods

Following the approach by Warshel and coworkers[1,4,6], we consider formally the solute-solvent system as a supermolecule and write down the molecular orbital (MO) functions of the system

$$\phi_i = \sum_\mu c_{\mu i}^S \chi_\mu^S + \sum_\mu c_{\mu i}^s \chi_\mu^s \tag{1}$$

where S denotes the solute, s the solvent, χ the atomic orbital, ϕ the molecular orbitals (MO), and c is the molecular orbital coefficient, which is obtained by diagonalizing the SCF equation for the solvent-solute system,

$$\begin{vmatrix} \mathbf{F}^S & \mathbf{F}^{Ss} \\ \mathbf{F}^{Ss} & \mathbf{F}^s \end{vmatrix} \mathbf{c}_i = E_i \mathbf{c}_i \tag{2}$$

where $\mathbf{F}^S$, $\mathbf{F}^s$ and $\mathbf{F}^{Ss}$ are the Fock matrixes[7], designating the solute-solute, solvent-solvent and solute-solvent interactions, respectively. Assuming that $\mathbf{F}^{Ss} = 0$, which implies that there is no charge transfer interaction between the solute and the solvent, we can separate the molecular orbitals of the solute from those of the solvent. The solute part of the $\mathbf{F}$ matrix is now written as

$$\begin{aligned} F_{\mu\mu}^S &\approx \overline{F}_{\mu\mu}^S - V_{coul}^A \\ F_{\mu\nu}^S &= \overline{F}_{\mu\nu}^S \end{aligned} \tag{3}$$

where $\overline{F}^S_{\mu\mu}$ and $\overline{F}^S_{\mu\nu}$ are the SCF matrix elements of the isolated solute molecule (subscripts μ and ν denote the atomic orbitals on the solute atoms), and V^A_{coul}, the Coulombic potential due to the solvent atom B at the solute A, is given by

$$V^A_{coul} = \sum_B \frac{e^2 q_B}{r_{AB}}$$

where q_B is the charge distribution on the solvent atom. The total effective potential of the system is then given by[1,3-6]

$$\varepsilon_{tot} = \varepsilon_{qm} + \sum_A \sum_B \frac{a_A a_B}{r^{12}_{AB}} - \frac{b_A b_B}{r^6_{AB}} + V_s \qquad (4)$$

where ε_{qm} is the quantum mechanical energy calculated from Eq. (3), the second term is the van der Waals interactions between the quantum mechanical solute and the classical solvent atoms, a and b are the van der Waals parameters and V_s is the solvent-solvent interaction. With the potential functions on hand, the free energy surface is calculated by the FEP method. The basic idea is to use a mapping potential as a reference state for the quantum mechanical ground state surface. The choice of the reference mapping potentials is arbitrary, as long as they are easy to evaluate and can be conveniently calibrated to either experimental data or *ab initio* calculations. Warshel's empirical valence bond (EVB) method[8] offers a convenient prescription for the classical mapping potential. In the EVB method, the system is treated as a solute-solvent supermolecule: the reaction region is defined as the "solute" while the surrounding environment the "solvent". The chemical reaction is represented by several valence bond (VB) structures. The main feature of the EVB method is its calibration possibilities that allow for the incorporation of experimental data[9] or *ab initio* calculations[10] into the Hamiltonian. With proper calibrations, the EVB method can usually reproduce the correct energetics of both the reactant and product states. another advantages of the EVB method is that the diabatic VB surfaces can be evaluated quite efficiently because they are represented by simple analytical functions. Using the EVB diabatic surfaces, we can define the mapping potential as,

$$\varepsilon_m = (1 - \lambda_m)\varepsilon^1_{cl} + \lambda_m \varepsilon^2_{cl} \qquad (5)$$

where λ_m is the mapping parameter, and ε_{cl}^1 and ε_{cl}^2 are the EVB diabatic potentials refined by EVB calibration procedure[11]. In the CTM method, the trajectories propagate on these classical reference potential surfaces, and the free energy surface is calculated by the relation,

$$\beta \Delta g(X) = -\sum_{m'=0}^{m} \ln \left\langle \exp\{-\beta[\varepsilon_{m'+1} - \varepsilon_{m'}]\} \right\rangle_{m'}$$
$$- \ln \left\langle \exp\{-\beta[\varepsilon_{tot}(X) - \varepsilon_m(X)]\} \right\rangle_m \tag{6}$$

where β is $1/k_B T$, k_B the Boltzmann constant and T temperature, X is the reaction coordinate and $<...>_m$ designates an average over the mth mapping potential. This seemingly complicated expression can be easily understood if one notices that the first term on the right hand side of Eq. (6) is the usual FEP expression, which gives the free energy difference between the neighboring mapping states, and the second term is actually the expression for umbrella sampling, which gives the relative free energy profile over the specific region of the reaction coordinate covered by the mapping potential. These two different mapping approaches are shown schematically in Fig. 1. In the CTM method, the quantum mechanical calculations need to be done only once in every 10~20 iterations steps, depending on how often energy information is collected during the simulation. The reduction of the number of the quantum mechanical calculations will effectively speed up the dynamical calculations, as will be shown later.
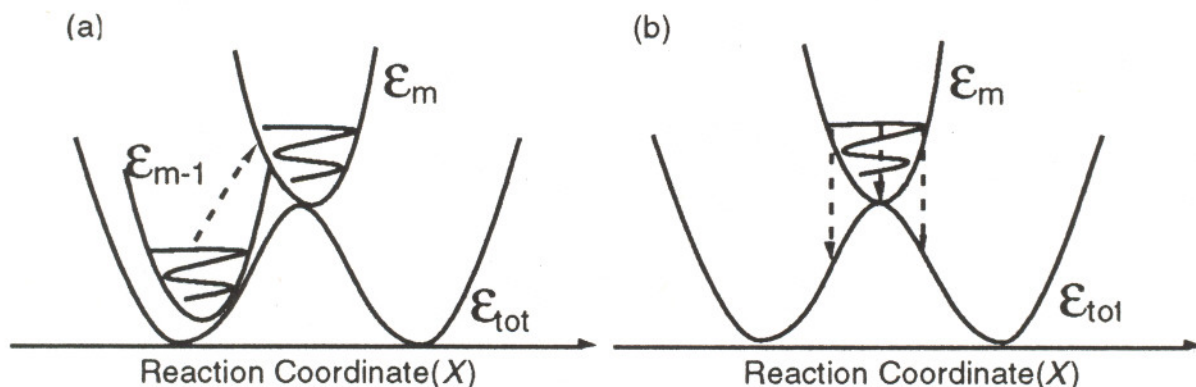

(a)   (b)

ε_m ε_m

ε_{m-1}

ε_{tot} ε_{tot}

Reaction Coordinate(X)   Reaction Coordinate(X)

Figure. 1 Two kinds of mappings used in the CTM method. Fig. 1a shows the

usual FEP mapping from one reference state to another (the mapping direction is indicated by the dotted line with an arrow). The wavy line is the trajectory that propagates on a classical reference potential. The first mapping gives us the free energy difference between mapping states. Fig.1b shows the umbrella sampling which calculates the relative free energy profile in the region covered by the mapping potential. The notations used in the figure is the same as those in Eq. 6.

Applications

We will use the CTM method developed above to study the proton transfer reactions in solution and in papain.

The proton-transfer reaction in solution

As a first example, we will study proton transfer between two water molecules in aqueous solution:

$$2H_2O \rightarrow OH^- + H_3O^+$$

The ΔH_{PT} for this reaction in gas phase is around 221 kcal/mol[12], but in water the proton transfer free energy difference reduces to around 21 kcal/mol[13]. This reaction offers a good example of the importance of the solvation effects on the proton transfer reaction. We will use the CTM method to calculate the free energy profile for this reaction in aqueous solution. The proton transfer reaction between two water molecules could be treated as an effective two-state problem corresponding to the two VB structures[11]

$$\psi_1 = \begin{bmatrix} H_2O & H_2O \end{bmatrix}$$
$$\psi_2 = \begin{bmatrix} OH^- & H_3O^+ \end{bmatrix}$$

It has been shown that the VB structures corresponding to the higher energy configuration could be effectively incorporated into these structures[14]. The diabatic potential function of the ith structure was expressed in the following form[9]

$$\varepsilon_{cl}^{i} = \sum_{j} \Delta M_{j}^{(i)}\left(b_{j}^{(i)}\right) + \sum_{j} \xi_{j}^{(i)} K_{j}^{(i)} (\theta_{j} - \theta_{0,j}^{(i)})^{2} + V_{ss}^{(i)} + V_{Ss}^{(i)}$$

$$+ V_{s} + \alpha^{(i)} \tag{8}$$

Here, the first term $\Delta M_{j}^{(i)}$ denotes the Morse potential corresponding to the jth bond in the ith VB structure, the second term describes the bond angle bending interactions. The factor $\xi_{j}^{(i)}$ in the second term is a coupling between bonds that are being broken or formed and those angles depending on those bonds. The third term and the fourth term represent the solute-solute and the solute-solvent nonbonded interactions. The last term $\alpha^{(i)}$ is the parameter used to adjust the vertical height of the ith VB structure to be close to the quantum mechanical energy. These parameters were adjusted so that they could reproduce the solvation free energies of the reacting species, i.e., H_2O, H_3O^+ and OH^-. All the calibration of the parameters was done by the FEP method.

The proton-transfer in papain

As a second example, we studied the proton-transfer reaction in papain. This enzyme, which catalyzes the hydrolysis of polypeptide substrates, is a cysteine protease composed of 212 amino acid residues that include a cysteine (Cys-25) and a histidine (His-159) in the active site. Although in solution the normal pK_a's for histidine and cysteine are around 6 and 10, respectively, experiment evidence[15] indicates that in protein these two residues form an ion pair, i.e., a proton is transferred from the SH group of Cys-25 to $N\delta_1$ of His-159. The proton transfer reaction is described by the equation

$$SH + Im \rightarrow S^- + HIm^+ \tag{9}$$

where the notations, SH and Im, denote cysteine and histidine, respectively. The VB structures and the diabatic potential functions were similar to Eq. (7) and (8). Extra terms such as the torsional interactions have to be included in the classical force field to account for the imidazole ring, and the s's appearing in Eq. (8) are understood to denotes both water and the enzyme, which solvate the quantum mechanical atoms. In simulations, the imidazole ring of histidine and the side-chain group, i.e., CH_3SH, of cysteine are treated quantum mechanically. while the other atoms of the systems are described by classical force fields.

In the simulations of papain, unlike in the first example, there are now hybrid bonds that connect the quantum mechanical and classical atoms. To handle this problem, we introduce a dummy atom along the hybrid bond and treat the dummy atom as a quantum mechanical hydrogen atom which does not interact with the rest of the classical atoms. This approach was used by several groups[3,16-18] and was shown to be quite useful.

Since we are interested not in the absolute magnitude but the relative magnitude of solvation effects on the thiolate-imidazolium ion pair between the surrounding enzyme environment and the aqueous solution, we calibrated the van der Waals parameters of cysteine and histidine such that they reproduced the corresponding pK_a values in solution, and then used the refined parameter set in the simulation of the protein papain.

Calculational Details

The molecular dynamics (MD) calculations were performed by ENZYMIX[19] which is implemented with the MNDO/AM1 option. The reaction regions, i.e., two water molecules, or the imidazolyl ring of histidine and the methyl thiol group of cysteine, were treated quantum mechanically by the semiempirical AM1 method, and the surrounding environments such as water molecules and the amino acid residues were treated molecular mechanically. The classical force fields are those of ENZYMIX. The refinement of the potential parameters of the reacting species was carried out by the FEP approach. The X-ray crystal structure of papain used in simulations was that of Kamphuis *et al*[20]. In simulations, the quantum mechanical reaction regions were surrounded by a 16 Å sphere of SCAAS water molecules[21]. The region outside the explicit water molecules was done by a continuum model[14]. There is no nonbonded cutoff between the solvent water molecules, or between the quantum mechanical atoms and the solvent molecules. The long range electrostatic interactions beyond the cut-off radius 8 Å were treated by the local reaction field method[22]. All calculations were done at a temperature of 300K and a step-size of 1 fs. The calculated energies were written to an external file every 50 steps. A typical trajectory time for the simulation of each mapping state was around 10 ps and the total trajectory time for one complete run was about 200 ps. The final results were the average of those of forward and backward mappings. All calculations were performed on IBM RISC/6000 375 and 3BT.

Results and Discussion

The calculated free energy profile for the proton transfer reaction of two water

molecules in solution is shown in Fig.2. The calculated activation barrier and ΔG^0 is 27.7 kcal/mol and 20.6 kcal/mol, respectively. Compared with the experimental values of 24 kcal/mol and 21 kcal/mol[13], the calculated ΔG^0 is quite close to the experimental value while the activation free energy is overestimated by 3.7 kcal/mol. However, considering the large value of the enthalpy in gas phase, $\Delta H_{PT}^{gas} = 221$ kcal/mol, our results clearly indicate that the CTM method can properly describe the solute-solvent interactions. Since the polarization of the solute atoms were naturally incorporated into the Hamiltonian, there is no need for any ad hoc scaling of the solute charges along the reaction path[23].
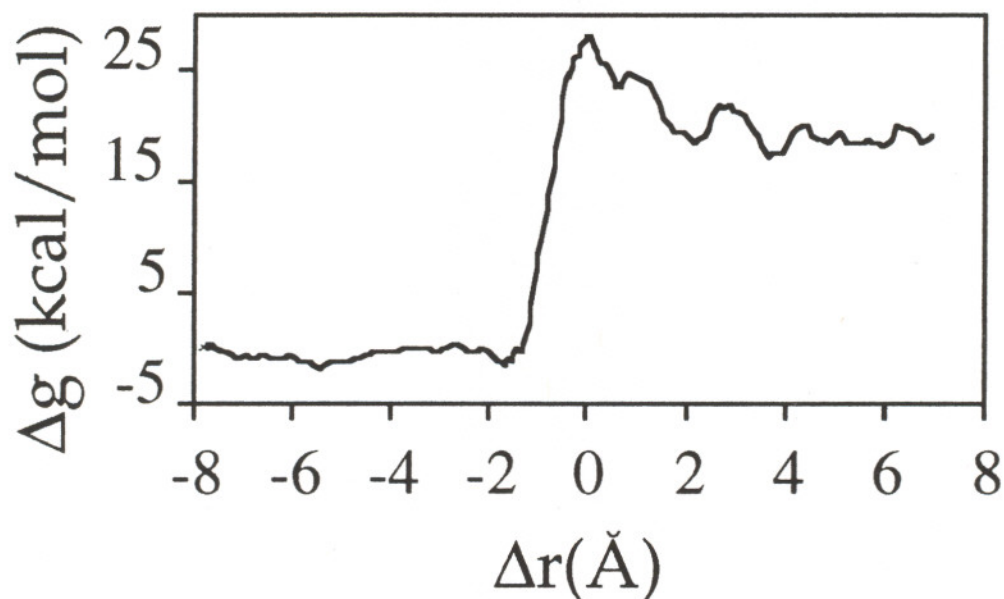

Figure 2: The calculated free energy profile for the proton transfer reaction in water. The reaction coordinate is defined by $\Delta r = r_{OH} - r_{O'H} - \delta r^{\neq}$, where r_{OH} and $r_{O'H}$ are the bond lengths being broken and formed, respectively, and $\delta r^{\neq}$ is the difference between them at the transition state.

In the study of chemical reactions ocurring in complicated environments such as enzymes, it is usually more convenient to use the energy difference of the diabatic energies as the reaction coordinate[24]. Fig.3 shows the results of our second test case - the free energy profiles vs. the energy gap for the proton transfer reactions both in solution and in papain. The calculated free energy for the proton transfer reaction in

papain has a negative value, i.e., ΔG_{PT}=-5 kcal/mol, which is in reasonable agreement with the experimental data of around -1.5 kcal/mol[25]. The calculated $\Delta\Delta G_{PT}$, i.e., the difference between ΔG_{PT}s in papain and in water, is around -10 kcal/mol, showing that the formation of the thiolate-imidazolium ion-pair is more favorable in the active site of papain than in solution. The calculated $\Delta\Delta G_{PT}$ is quite close to the experimental value of -7 kcal/mol[25]. These results indicate that the CTM method is able to account for the relative stability of the thiolate-imidazoilim ion pair between in water and in papain.
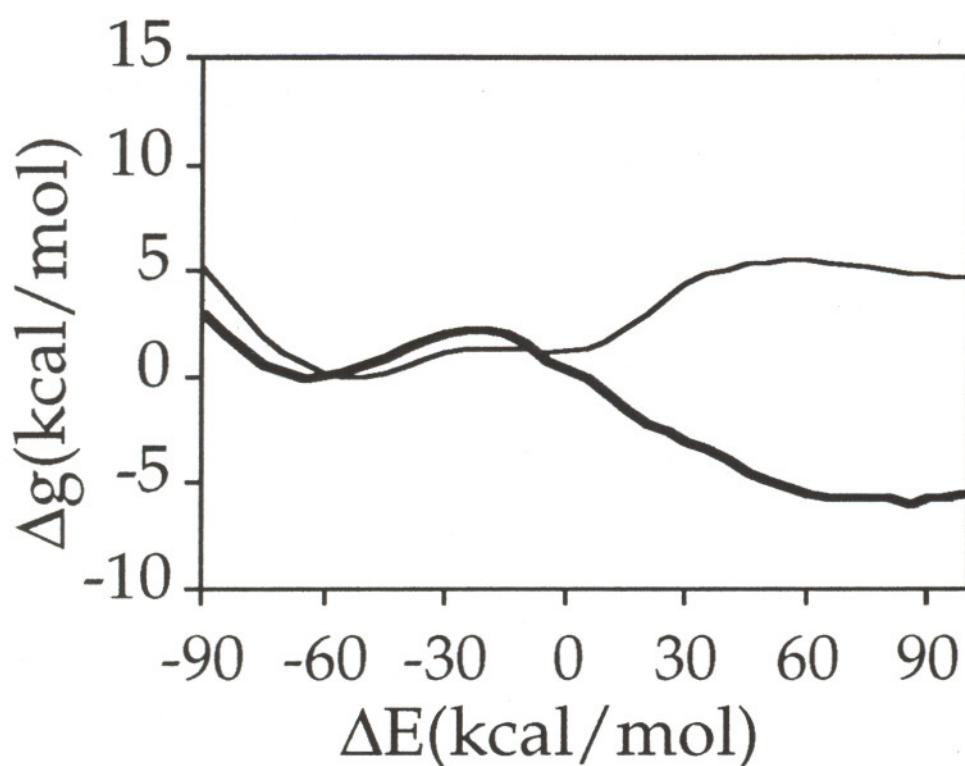

Figure 3: The free energy profile vs. the energy gap for the proton transfer between His and Cys in solution (the thin line) and in papain (the thick line).

To examine the performance of the CTM method in simulations of large systems, we compared the CPU times between the CTM method and the usual hybrid approach that performs quantum mechanical calculation at each iteration. The systems for the benchmark tests consist of various numbers of quantum mechanical water molecules surrounded by a 12 Å sphere of SCASS water molecules. As shown in Table 1, when the number of the quantum mechanical water molecules increases, the speed of the CTM method could be one order of magnitude faster than that of the conventional method. This saving of CPU time should be quite significant for the dynamics

simulations of enzymatic reactions. In conclusion, we have developed an efficient approach which allows us to run trajectories on a classical potential surface and, by mapping the classical state to the quantum mechanical states, to obtain the free energy profile for the chemical reactions in solution and in enzymes. We applied this method to the proton transfer reactions both in solution and in papain, and the results are quite encouraging, indicating that this method is applicable to the chemical reactions both in solution and in enzymes.

Table 1. Comparison of the CTM method and the conventional QM/MM method for the calculations of a 12 Å sphere of classical water molecules surrounding various numbers of quantum mechanical water molecules.

Number of QM water molecules	10	20	30	35	40
CPU time ratio[1]	2	5	9	10	12

[1] The ratio between the CPU time of the conventional approach and that of the CTM method

Acknowledgments

References

(1)   Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227-249.

(2)   Kollman, P. A.; Hayes, D. M. *J. Am. Chem. Soc.* **1981**, *103*, 2955.

(3)   Field, M. J.; Bash, P.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700-733.

(4)   Luzhkov, V.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4491-4499.

(5)   Gao, J.; Xia, X. *Science* **1992**, *258*, 631-635.

(6)   Luzhkov, V.; Warshel, A. *J. comput. Chem.* **1992**, *13*, 199-213.

(7)   Warshel, A.; Lappicirella, A. *J. Am. Chem. Soc.* **1981**, *103*, 4664.

(8)   Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218-6226.

(9)   Åqvist, A.; Warshel, A. *J. Mol. Biol.* **1992**, *224*, 7.

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

(11)   Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523-2544.

(12)   Stewart, J. J. P. *J. comp. chem.* **1989**, *10*, 221-264.

(13)   Eigen, M.; de Maeyer, L. *Z. Elektrochem.* **1955**, *59*, 986.

(14)   Warshel, A.; Russell, S. T. *Q. Rev. of Biophys.* **1984**, *17*, 283-422.

(15)   Polgar, L. *FEBS Lett* **1974**, *47*, 15-18.

(16)   Davies, T. D.; Maggiora, G. M.; Christoffersen, R. E. *J. Am. Chem. Soc.* **1974**, *96*, 7878.

(17)   Christoffersen, R. E.; Maggiora, G. M. *Chem. Phys. Lett.* **1969**, *3*, 419.

(18)   Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1986**, *7*, 718.

(19)   Warshel, A.; Creighton, S. *Computer Simulation of Biomolecules Systems*; ESCOM: Leiden, 1989, pp 120.

(20)   Kamphuis, I. G.; Kalk, K. H.; Swarte, M. B. A.; Drenth, J. *J. Mol. Biol.* **1984**, *179*, 233-256.

(21)   Warshel, A.; King, G. *Chem. Phys. Lett.* **1985**, *121*, 124-129.

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

(23)   Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. *J. Am. Chem. Soc.* **1985**, *107*, 154-163.

(24)   Warshel, A.; Hwang, J.-K.; Åqvist, J. *Faraday Discuss.* **1992**, *93*, 225-238.

(25)   Ménard, R.; Khouri, H. E.; Plouffe, C.; Laflamme, P.; Dupras, R.; Vernet, T.; Tessier, D. C.; Thomas, D. Y.; Storer, A. *Biochemistry* **1991**, *30*, 5531-5538.