Free Energy Perturbation Calculations with Combined QM/MM Potentials Complications, Simplifications, and Applications to Redox Potential Calculations

Guohui Li, Xiaodong Zhang, and Qiang Cui*

Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, 1101 University Ave, Madison, Wisconsin 53706

Received: February 3, 2003; In Final Form: May 12, 2003

The present work deals with a formal discussion on complications associated with using combined quantum mechanical and molecular mechanical (QM/MM) potentials in free energy perturbation simulations. Because quantum mechanical potentials are not trivially separable and because of the difficulty associated with QM calculations for "fractional" electronic Hamiltonians, conventional computational strategies cannot be straightforwardly applied to free energy simulations with QM/MM potentials. Here, we propose a unique coupling scheme, termed the "dual-topology-single-coordinate" scheme, in which the two chemical states are forced to adopt the same set of Cartesian coordinates during the perturbation simulation; the method is exact because the free energy difference is independent of the coupling path, although a formal proof is also given. The scheme combines the merits of the conventional dual- and single-topology approaches: both coupling potential and free energy derivative are straightforward to compute for any form of QM method (even if nonvariational) and QM/MM coupling scheme, calculations can be performed at end points without numerical/ sampling instabilities, no corrections related to artifactual reference states or Jacobian factors have to be considered, and numerical convergence is rapid. The method is illustrated with an application to the redox potential of FAD in cholesterol oxidase. It was found that the protein and solvent respond in approximately a linear fashion to the reduction of FAD with a small reorganization energy; the bulk solvent makes an essential contribution to compensate for the reorganization of the protein. The contribution of Asn 485 was found to be around 1.3 kcal/mol with several different simulations, which is close to the experimental estimate based on mutation studies.

I. Introduction

Free energy perturbation simulations play a vital role in understanding chemical processes that occur in the condensed phase.[1−4] For processes that either occur rarely (e.g., activated chemical reaction) or are intrinsically slow (e.g., ligand binding to biomolecules), which are usually difficult to simulate directly with the current computational facilities, free energy simulations are of tremendous value because they deal with thermodynamic quantities which are independent of the time scale of the process. Moreover, useful insights into the relevant kinetics can be obtained from free energy results based on the Hammond postulate; for example, an interaction that stabilizes the product state relative to the reactant state is also expected to contribute favorably to the rate of the process. Another unique feature of free energy simulations is that versatile thermodynamic cycles can be constructed to analyze experimental observables through pathways that are difficult or even impossible to follow experimentally but are convenient for simulation and analysis. This has been nicely illustrated in many "alchemy" free energy simulations that explored relative binding energies of various ligands and effect of mutation on the stability of proteins.[1−3]

Although the quantitative results depend on the quality of the potential function employed, the unique value of free energy simulations is that they allow a meaningful analysis of various contributions to the particular problem in hand.[5−7] For many chemical processes that involve complex changes in the electronic structure, however, it is crucial to employ a potential function that is sufficiently accurate and flexible. Important examples are redox processes and ligand exchange reactions concerning transition metals and light absorption/emission processes involving electronically excited states. Under those circumstances, a hybrid quantum mechanical and molecular mechanical (QM/MM) potential is an attractive compromise for reliability and computational efficiency.[8−10]

Another motivation for employing a hybrid QM/MM potential is based on the consideration that for many (though not all) problems, a direct comparison with experimental measurement is not straightforward with simulations employing pure classical potential functions. In the study of complex systems such as biomolecules, for which many approximations have to be made on the simulation details (e.g., form of potential functions, length of simulations, treatment of long-range electrostatics, etc.), it is essential to compare computed free energies with available experimental measurements such that the subsequent analyses can be taken as meaningful results. If one is interested in analyzing contributions from amino acids and solvent to the redox property of a cofactor bound to a protein, for example, the most convincing validation is to compare the computed and measured redox potentials. This is not straightforward if the standard classical force field is used, because of the lack of an explicit account of different numbers of electrons in the two oxidation states. Other examples of this sort include the study of absorption and emission spectra in the condensed phase and the prediction of the pK_a of residues in proteins and nucleic

---

* To whom correspondence should be addressed.

acids. Under these circumstances, although computing relative properties using reference systems is of interest and commonly done in the literature with classical force fields,[10,11] it is a more stringent test to employ a QM/MM treatment of the system, such that absolute properties (redox potential, excitation energy, pK_a) can be computed. In other words, accurate results in homogeneous solution do not necessarily transfer to simulations of heterogeneous systems such as biomolecules.

Although various QM/MM approaches have been widely applied to the study of chemical processes (mainly chemical reactions) in solution and biological systems,[9,12−14] considerably less work has been done to employ QM/MM potentials in free energy perturbation calculations. Exceptions for the latter include mainly solvation free energy calculations with chemically simple QM molecules such as metal ions[15] and Monte Carlo simulations in which the structure of the QM region is held fixed.[16,17] In addition to the computational cost associated with meaningful QM/MM free energy simulations, there are numerical complications (e.g., end-point problems) that cannot be easily circumvented with common strategies[18] that work for classical pairwise potentials; these complications arise mainly because the QM/MM potential is not separable. In the current work, we present a formal discussion of problems associated with using QM/MM potentials in free energy simulations. Moreover, we propose to use a specific coupling pathway that connects the reactant and product states without suffering these problems. Such a "dual-topology-single-coordinate" approach is particularly suitable to study processes that conserve the number of atoms, such as redox cycles, absorption/emission events in the condensed phase, and binding exchange of metal ions to macromolecules; the method can also be extended straightforwardly to study other processes that involve different numbers of atoms in the two end-states, such as predicting pK_a's of amino acids in proteins. This approach in the current form cannot be applied to the *same* electronic states with very different chemical structures (e.g., the reactant and product of a typical chemical reaction), which is not of serious concern because other simulation techniques, such as umbrella sampling,[19−21] would be more appropriate in those cases.

In section II, we start by formally discussing the complications that arise specifically to free energy simulations with QM/MM potentials; we then describe the "dual-topology-single-coordinate" approach. In section III, we illustrate this approach with an application to the calculation of the first reduction potential for FAD in cholesterol oxidase; both absolute reduction potential and the effect of a nearby conserved residue (Asn 485) have been studied. In section IV, we draw several conclusions. A brief discussion relevant to the current work has been reported recently.[22]

II. Free Energy Perturbation Simulations with QM/MM Potentials

In this section, we first briefly review the theory behind free energy simulations and the two general implementations: the dual-topology and single-topology approaches. We then discuss problems that arise in simulations with QM/MM potentials and describe how these problems can be circumvented with the so-called "dual-topology-single-coordinate" approach.

II.1. Free Energy Perturbation Calculations. In many areas of chemistry and biochemistry, one is interested in the free energy change associated with the transformation between two chemical states A and B (hereafter referred to as "subsystems") in the presence of the "environment" (e.g., solution or a protein), C. As discussed in details by many authors,[18,23,24] a popular

approach for obtaining the free energy difference is thermodynamic integration,[25] in which one computes the free energy derivative $(\partial F/\partial \lambda)$ with respect to the coupling parameter (λ) that converts A to B through a series of intermediate values ("windows") and then integrates these derivatives over λ from 0 to 1

$$\Delta F = \int_0^1 \frac{\partial F(\lambda)}{\partial \lambda} \, d\lambda = \int_0^1 \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{1}$$

where $U(\lambda)$ is the potential energy of the system at a particular coupling parameter, λ (see below). Although it is possible (or even preferable[26]) to include higher free energy derivatives, the common practice is to consider only the first derivatives at sufficiently large number of λ values. Thermodynamic integration is preferred in many applications over the exponential formulation[27] associated with thermodynamic perturbation (which is formally exact) because it is more straightforward to perform component analysis.[6,28]

To compute the free energy derivative, two implementations have been proposed. The "dual-topology" approach involves separate copies of the two chemical states; thus, the total potential function at a particular λ value is given by

$$U^D(\mathbf{X}_A, \mathbf{X}_B, \mathbf{X}_C; \lambda) = (1 - \lambda)[U_{AA}(\mathbf{X}_A) + U_{AC}(\mathbf{X}_A, \mathbf{X}_C)] + \lambda[U_{BB}(\mathbf{X}_B) + U_{BC}(\mathbf{X}_B, \mathbf{X}_C)] + U_{CC}(\mathbf{X}_C) \tag{2}$$

where U_{IJ} stands for the potential energy corresponds to the interaction between I and J (I, J = A, B, and C). In the alternative "single-topology" scheme, there is only one copy of a "hybrid subsystem" (hybrid of A and B) in the presence of the environment, C; that is, the coupling potential has the form

$$U^S(\mathbf{X}_H, \mathbf{X}_C; \lambda) = U_{HH}(\mathbf{X}_H; \lambda) + U_{HC}(\mathbf{X}_H, \mathbf{X}_C; \lambda) + U_{CC}(\mathbf{X}_C) \tag{3}$$

where both the hybrid-hybrid (U_{HH}) and hybrid-environment (U_{HC}) interactions are λ-dependent; this can be accomplished by switching parameters (e.g., partial charges and force constants) associated with the hybrid subsystem.

As discussed thoroughly in previous work, both approaches have their merits and caveats. The single topology approach often delivers faster convergence,[29] although the calculation can be problematic if the two states tend to adopt very different conformations (e.g., different rotamer states for protein side chains[5]); one also has to be cautious about contributions from the switched molecular parameters, e.g., Jacobian contributions due to the change in bond lengths in the two states.[30−33] The dual topology approach does not suffer from these problems, although it often encounters end-point ($\lambda = 0, 1$) problems that emerge as large structural/positional fluctuations and conformational sampling problems associated with the subsystems.[34] These problems arise from the fact that one of the states (e.g., state A in eq 2 as λ approaches one) has a vanishingly small contribution to the total energy and force. Divergence in free energy derivative can also occur because of the fast variation of specific interactions (e.g., van der Waals terms[35,36]) as a function of λ.

II.2. Using QM/MM Potentials in Free Energy Simulations−General Discussions. As mentioned in the Introduction, combined QM/MM potentials are preferred over pure classical force fields in the study of many processes that involve complex electronic structural changes.[8,9] Although the thermodynamic integration scheme (eq 1 and 2) is applicable to any form of potential energy functions, complications do arise in practical

Free Energy Perturbation Calculations

J. Phys. Chem. B, Vol. 107, No. 33, 2003 **8645**

applications if combined QM/MM potential functions are used. In the following, we present general discussions on these problems and also comment on possible solutions. For simplicity, we assume that the two "subsystems" (A and B) are treated with QM and the environment (C) is treated as MM; this is likely to be the typical choice in practical applications and extensions to other QM/MM partitions do not poise any conceptual problems.

II.2a. Dual-Topology Framework. With a combined QM/MM potential, the definition of a coupling potential that most resembles the classical scenario (eq 2) involves scaling the total QM/MM potential energies; i.e.

$$U^D(\mathbf{X}_A,\mathbf{X}_B,\mathbf{X}_C;\lambda) = (1-\lambda)[\langle\Phi_A|\hat{H}_{AA}(\mathbf{X}_A) + \hat{H}_{AC}(\mathbf{X}_A,\mathbf{X}_C)|\Phi_A\rangle] + \lambda[\langle\Phi_B|\hat{H}_{BB}(\mathbf{X}_B) + \hat{H}_{BC}(\mathbf{X}_B,\mathbf{X}_C)|\Phi_B\rangle] + U_{CC}(\mathbf{X}_C) \quad (4)$$

where $\Phi_A(\Phi_B)$ is the wave function for A (B) in the presence of the environment C from a combined QM/MM energy calculation at the geometry $(\mathbf{X}_{A(B)},\mathbf{X}_C)$. Note that, because the scaling is performed *after* the QM/MM calculations, the QM/MM interaction energies themselves do not depend on the coupling parameter λ, and it is straightforward to compute the derivative of U^D with respect to λ and consequently the free energy derivative.

The practical difficulty associated with adopting eq 4 as the coupling potential is that one of the subsystems tends to have very small contributions to the potential energy and force at the end points ($\lambda = 0, 1$), which lead to wild structural distortions and numerical instability in the free energy derivatives (vide infra); this is even more serious in QM/MM simulations than classical calculations because the QM/MM potentials are highly anharmonic and unstable (e.g., SCF convergence failure) with respect to large structural distortions. To circumvent similar end-point problems in simulations with pure classical potentials, it is often chosen to scale only the interaction between the subsystems and the environment; that is, the coupling potential takes the form

$$U'^D(\mathbf{X}_A,\mathbf{X}_B,\mathbf{X}_C;\lambda) = (1-\lambda)[U_{AC}(\mathbf{X}_A,\mathbf{X}_C)] + \lambda[U_{BC}(\mathbf{X}_B,\mathbf{X}_C)] + U_{AA}(\mathbf{X}_A) + U_{BB}(\mathbf{X}_B) + U_{CC}(\mathbf{X}_C) \quad (5)$$

In such a way, the internal structures of A and B remain close to equilibrium even at end points, which ensures the stability of the simulation. The artifact is that the end states are no longer physical; for example, $\lambda = 0$ corresponds to the desired state (A + C) plus an isolated B. This can be remedied, however, by approximate calculations of the contribution due to isolated B.[30–32,34]

In principle, the similar strategy can be adopted with a QM/MM potential, where the correspondence of eq 5 is

$$U'^D(\mathbf{X}_A,\mathbf{X}_B,\mathbf{X}_C) = \langle\Phi_A|\hat{H}_{AA}(\mathbf{X}_A) + (1-\lambda)\hat{H}_{AC}(\mathbf{X}_A,\mathbf{X}_C)|\Phi_A\rangle + \langle\Phi_B|\hat{H}_{BB}(\mathbf{X}_B) + \lambda\hat{H}_{BC}(\mathbf{X}_B,\mathbf{X}_C)|\Phi_B\rangle + U_{CC}(\mathbf{X}_C) \quad (6)$$

Although it is straightforward to implement this coupling scheme for the energy and force evaluations required for the MD simulation, there is a minor complication in the evaluation of the energy derivative with respect to λ, which is required in the calculation of free energy derivative (eq 1). Take the first term in eq 6 as an example, we have

$$\frac{\partial}{\partial\lambda}\langle\Phi_A|\hat{H}_{AA} + (1-\lambda)\hat{H}_{AC}|\Phi_A\rangle = -\langle\Phi_A|\hat{H}_{AC}|\Phi_A\rangle + \langle\partial\Phi_A/\partial\lambda|\hat{H}_{AA} + (1-\lambda)\hat{H}_{AC}|\Phi_A\rangle + \langle\Phi_A|\hat{H}_{AA} + (1-\lambda)\hat{H}_{AC}|\partial\Phi_A/\partial\lambda\rangle \quad (7)$$

Because the Hellman–Feymann theorem[37–39] does not hold for electronic structure theories in general, the derivative of the wave function (Φ_A) with respect to the coupling parameter ($\partial\Phi_A/\partial\lambda$) may be required; these quantities can be obtained by solving the coupled-perturbed equations.[40] When the QM method is stationary with respect to the variational parameters (e.g., Hartree–Fock[41] or density functional theory[42]), however, the last two terms in eq 7 can be ignored according to Wigner's theorem.[40] In any case, one has to evaluate the expectation value of the QM/MM interaction ($\langle\Phi_A|\hat{H}_{AC}|\Phi_A\rangle$) with the converged wave function as the additional effort, because QM/MM potentials are not trivially separable as classical force fields. For a relatively simple form of the QM/MM interaction ($\hat{H}_{AC}$), however, such a step is straightforward.

II.2b. Single Topology Framework. If the single topology approach is taken, there is only one potential and one set of coordinates that correspond to the hybrid subsystem (eq 2). The major issue here is the calculation of the hybrid potentials ($U_{HH}(\lambda)+U_{HC}(\lambda)$) and the free energy derivative (eq 1). With a MM potential, the λ dependence of the hybrid potential is introduced through variations in parameters such as partial charges and internal coordinate force constants; free energy derivatives can then be evaluated straightforwardly because of the simple analytical form of MM potentials. With a combined QM/MM potential, it is not always straightforward to introduce an electronic Hamiltonian for the hybrid subsystem. For processes that involve different electronic states, for example, it is nontrivial to introduce "fractionally" electronically excited states. In other cases, it is indeed possible to specify the hybrid electronic Hamiltonian, although the corresponding potential and derivative with respect to λ might be difficult to compute. To compute a one-electron reduction potential, for example, the electronic Hamiltonian of the hybrid subsystem is given by

$$\hat{H}^S(\lambda) = \hat{H}_{HH}^N + \hat{H}_{HC}^N + \lambda\left[\sum_{I\in H(N)}\frac{-Z_I}{|\mathbf{X}_I-\mathbf{r}_{N+1}|} + \sum_{i\in H(N)}\frac{1}{|\mathbf{r}_i-\mathbf{r}_{N+1}|} + \sum_{I\in C}\frac{-Q_I}{|\mathbf{X}_I-\mathbf{r}_{N+1}|}\right] \quad (8)$$

where N is the number of electrons in the oxidized form; i.e., the hybrid system has a fractional number of electrons ($N + \lambda$). Such a calculation is difficult to perform in practice with wave function based ab initio methods but is indeed possible in the framework of density functional theory.[42] Similar to scenarios in the dual-topology framework in which only the subsystem–environment interactions are scaled, derivatives of the wave function/density with respect to λ are required for calculating the free energy derivatives ($\langle\partial U^S(\lambda')/\partial\lambda'\rangle_\lambda = \langle\partial[\langle\Phi(\lambda')|\hat{H}^S(\lambda')|\Phi(\lambda')\rangle]/\partial\lambda'\rangle_\lambda$) when the QM method is nonvariational.

II.3. Efficient QM/MM Free Energy Simulations with a "Dual-Topology-Single-Coordinate" Approach. Based on the above discussions, it is clear that neither the standard dual-topology nor the single-topology approach is numerically convenient for combined QM/MM potential functions; the complication is mainly due to the fact that QM/MM interactions are not trivially separable as popular classical force fields and

that it is difficult to perform QM/MM calculations with "fractional" electronic Hamiltonians. In the current work, we propose a unique coupling scheme that is inspired by combining certain features of the conventional dual and single topology approaches, which we term the "dual-topology-single-coordinate" scheme. In this scheme, both subsystems (A and B) are present in the simulation and therefore it is a dual-topology based method; the two chemical states, however, are constrained to adopt exactly the same set of Cartesian coordinates, which calls for features of single-topology based methods. The merits of this coupling scheme are that it is straightforward to implement, free of sampling problems at the end points, does not involve artifactual reference states, and offers fast convergence.

In the dual-topology-single-coordinate scheme, there are two electronic Hamiltonians, and the potential function of the system takes the same form as in the standard dual-topology approach, i.e., U^D in eq 4. The unique feature is that the two subsystems are constrained to adopt the same Cartesian coordinates; that is, we compute a *constrained* ensemble average of the potential energy derivative with respect to λ

$$\left\langle \frac{\partial U^D(\lambda)}{\partial \lambda} \right\rangle^\delta_\lambda = \frac{1}{Q^\delta(\lambda)} \int d\mathbf{X}_A \, d\mathbf{X}_B \, d\mathbf{X}_C \, \frac{\partial U^D(\lambda)}{\partial \lambda} e^{-\beta U^D(\lambda)} \delta(\mathbf{X}_A - \mathbf{X}_B) \quad (9)$$

where the *constrained* ensemble configurational integral, $Q^\delta(\lambda)$, is given as

$$Q^\delta(\lambda) = \int d\mathbf{X}_A \, d\mathbf{X}_B \, d\mathbf{X}_C \, e^{-\beta U^D(\lambda)} \delta(\mathbf{X}_A - \mathbf{X}_B) \quad (10)$$

It can be shown straightforwardly that the left-hand-side corresponds to a *constrained* free energy derivative, i.e.

$$\left\langle \frac{\partial U^D(\lambda)}{\partial \lambda} \right\rangle^\delta_\lambda = -\beta^{-1} \frac{\partial}{\partial \lambda} \ln Q^\delta(\lambda) = \frac{\partial F^\delta(\lambda)}{\partial \lambda} \quad (11)$$

The *constrained* free energy at end points can be written out in more transparent forms. For example

$$F^\delta(1) = -\beta^{-1} \ln[\int d\mathbf{X}_A \, d\mathbf{X}_B \, d\mathbf{X}_C \, e^{-\beta U^D(\mathbf{X}_A, \mathbf{X}_B, \mathbf{X}_C; 1)} \delta(\mathbf{X}_A - \mathbf{X}_B)] = -\beta^{-1} \ln[\int d\mathbf{X}_A \, d\mathbf{X}_B \, d\mathbf{X}_C \, e^{-\beta U^D(\mathbf{X}_B, \mathbf{X}_C)} \delta(\mathbf{X}_A - \mathbf{X}_B)] = -\beta^{-1} \ln$$
$$[\int d\mathbf{X}_B \, d\mathbf{X}_C \, e^{-\beta[U_{BB}(\mathbf{X}_B) + U_{BC}(\mathbf{X}_B, \mathbf{X}_C) + U_{CC}(\mathbf{X}_C)]}] = F_{\langle B+C \rangle} \quad (12)$$

where we have used the fact that $U(\mathbf{X}_A, \mathbf{X}_B, \mathbf{X}_C; 1)$ depends only on the coordinates of B and C, and the property of the Dirac delta function to eliminate the integration over $\mathbf{X}_A$. Similarly, $F^\delta(0)$ can be shown to give nothing but $F_{\langle A+C \rangle}$. In other words, the integration of the *constrained* free energy derivative gives exactly what we are looking for, the Helmoltz free energy difference between the B + C and the A + C systems

$$\int_0^1 d\lambda \, \frac{\partial F^\delta(\lambda)}{\partial \lambda} = F^\delta(1) - F^\delta(0) = F_{\langle B+C \rangle} - F_{\langle A+C \rangle} \quad (13)$$

Physically, this result is expected because free energy change is independent of the coupling path.

Because the two subsystems adopt the same set of Cartesian coordinates during the simulation, the current scheme does not suffer from the wild structural distortions associated with $U^D(\lambda)$ in eq 4 at end points; there is always at least one chemical state that maintains the structure of the subsystem (QM region) to be close to equilibrium. Compared to the scheme of scaling only the subsystem−environment interaction (i.e., adopting $U'^D(\lambda)$ in eq 6), the free energy derivative is straightforward to compute, even if the QM method is nonvariational, and there is no need to introduce special corrections for the change of reference states at end points.[18,34] Compared to the single-topology approach, the coupling potential and its λ derivative are much easier to compute in the present dual-topology-single-coordinate scheme. Moreover, because there are two topologies, the subsystem evolves naturally from one chemical state to the other as the coupling parameter is changed from 0 to 1; thus, there is no "Jacobian" contribution that has to be considered.[30−33] Finally, the fact the two chemical states adopt the same set of coordinates points to a faster numerical convergence, as often found in the comparison between dual-topology and single-topology simulations.[29]

In the above discussion, an essential assumption is that the two chemical states have the same set of atoms; otherwise, δ-($\mathbf{X}_A - \mathbf{X}_B$) is ill-defined. This may not be as serious a limitation as it might appear to be. Although one unique feature of "alchemical" free energy perturbation simulations is that the two chemical states do not necessarily have the same chemical compositions, such applications rarely call for a QM/MM treatment of the potential functions. Most systems that require the use of QM/MM potentials would involve highly similar *atomic* arrangements (i.e., *chemical bonding patterns*) but different *electronic* configurations; examples are different oxidation states and ground vs electronically excited states. One application that does involve a different number of atoms in the two chemical states and can benefit from QM/MM treatment is pK_a prediction in macromolecules; the proton is no longer present in the deprotonated state. In this case, the proton can be treated as a dummy atom harmonically anchored to the proton donor atom in the deprotonated state, and then the dual-topology-single-coordinate approach can still be used. The contribution of the dummy atom can be removed using local configurational integrals,[43] as has been discussed extensively in the past.[30,31]

Finally, we note that the dual-topology-single-coordinate approach also simplifies the component analysis when the two chemical states have the same number of atoms. With the popular form of the hybrid QM/MM potential[8]

$$U_{Tot}^{QM/MM}(\vec{R}) = \langle \Psi(\vec{r};\vec{R})|H^{QM} + H_{el}^{QM/MM}|\Psi(\vec{r};\vec{R})\rangle + U_{van}^{QM/MM}(\vec{R}) + U_{bonded}^{QM/MM}(\vec{R}) + U^{MM}(\vec{R}) = U_{ele}^{QM/MM}(\vec{R}) + U_{van}^{QM/MM}(\vec{R}) + U_{bonded}^{QM/MM}(\vec{R}) + U^{MM}(\vec{R}) \quad (14)$$

only the electrostatic term contributes explicitly to the free energy derivative if the same set of van der Waals and bonded parameters are used for the interaction between MM atoms and the two QM states (because the latter two have the same Cartesian coordinates)

$$\frac{\partial F}{\partial \lambda}(\vec{R};\lambda) = \langle -U_{ele}^{A/MM}(\vec{R}_{QM}, \vec{R}_{MM}) + U_{ele}^{B/MM}(\vec{R}_{QM}, \vec{R}_{MM}) \rangle_\lambda \quad (15)$$

As a result, performing the free energy component analysis in a perturbative fashion is very straightforward: one turns off the partial charges on a particular MM residue, computes the change in the free energy derivatives using the original MD trajectories, and then integrates over λ. Because only the electrostatic interaction contributes explicitly, there is little worry

Free Energy Perturbation Calculations

J. Phys. Chem. B, Vol. 107, No. 33, 2003 **8647**

about the "path dependence" of the decomposition, and the components are totally symmetric with respect to direction of the perturbation (A → B vs B → A).[28] However, because of the nonadditivity of QM/MM interactions, we emphasize that results from the component analysis should only be taken as qualitative values.

III. Test Calculations

As an illustration of the dual-topology-single-coordinate approach, we present calculations for the first reduction potential of flavin adenine dinucleotide (FAD) in cholesterol oxidase (Chox).[44] Chox is a bacteria enzyme, whose major function is to oxidize cholesterol for the use as a carbon source in primary metabolism. The oxidation is made possible by FAD as the redox cofactor, which is bound noncovalently to the active site in some enzymes, while covalently attached to a His residue in others. Understanding the redox property of FAD is a key step in unraveling the catalytic mechanism of Chox, for which many questions remain to be answered despite much insights provided by previous crystallographic,[45-48] kinetic, and mutation studies.[47-50] Recent mutation and X-ray studies[47] showed that Asn 485, a conserved residue in the glucose−methanol−choline (GMC) oxidoreductase family,[51] modulates the redox property of FAD through an interaction reminiscent of the NH-π interaction found in a number of proteins.[52,53] For quantifying this contribution and systematically exploring the effect of other conserved residues or water molecules in the active site, theoretical calculations are of great value; this is true because mutations may often induce unexpected structural changes in the active site, which might complicate the interpretation of measurements. In Chox, for example,[47] the X-ray structure of the Asn485L mutant indicated a structural change of the nearby methionine (Met 122), which in turn allowed the presence of an additional water molecule in the active site. Therefore, the measured shift in the reduction potential due to the Asn485L mutation may not fully correspond to the contribution from Asn 485 in the wild-type enzyme. In the present work, we focus on the contribution from Asn 485 to the first reduction of FAD, as an illustration of QM/MM free energy simulations with the dual-topology-single-coordinate approach. Correspondingly, calculations were carried out for both the wild-type enzyme and the Asn485L mutant. More detailed and systematic analysis on contributions from other residues and solvent molecules will be reported separately.

III.1. Computational Setup. For the reduction potential calculations, the isoalloxazine ring of FAD was treated with QM (see below) and the rest of FAD and other protein atoms plus solvent molecules were described with the CHARMM22 force field.[54] The conserved amino acids in the active site,[51] such as Glu 361, Asn 485, and His 447 (Figure 1b), were also treated with MM, such that their contribution to the reduction potential can be analyzed straightforwardly through a perturbative analysis (following eq 15). Link atoms[8] were introduced between the C^{11} and C^{12} atoms in FAD to saturate the valence of the boundary QM atom (see Figure 1b). The link atoms interact with the MM atoms, except the "link host" MM atom (the C^{11} atom in this case), through electrostatic terms; no van der Waals interactions are included. This scheme has been shown to be a satisfactory way to treat the QM/MM interface, particularly when the charges of the atom in the neighborhood of the link atom are small;[55] this is true in the present case.

The QM level used here is the self-consistent-charge density-functional-tight-binding (SCC-DFTB) method,[56] which was recently introduced into CHARMM[57] in a QM/MM frame-
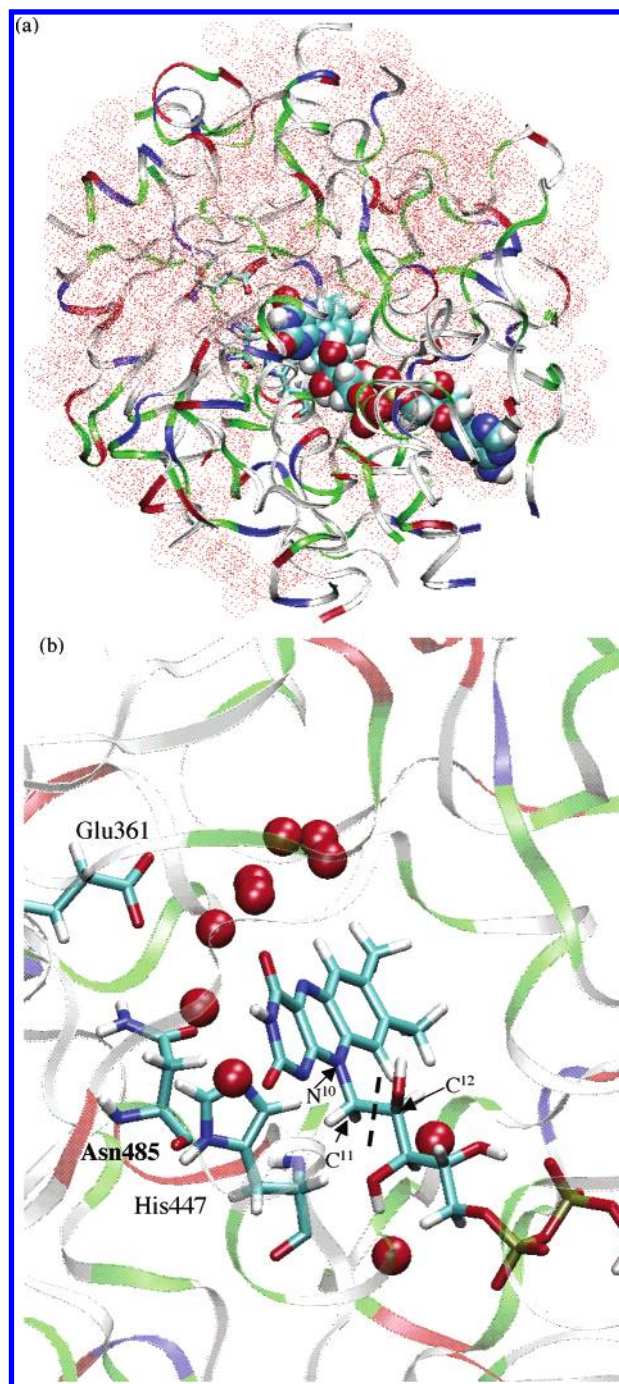

Figure 1. FAD-cholesterol oxidase system based on the X-ray structure from *Streptomyces* at 1.5 Å (PDB code 1B4V). (a) The stochastic boundary setup of 25 Å around the isoalloxazine ring in FAD. The FAD molecule is shown in space-filling models, several conserved residues (Glu 361, Asn 485, His 447) are shown in line forms, the protein backbone is shown in ribbon (color coded as the following: red, acidic; blue, basic; green, polar; white, nonpolar), and the water molecules are shown in dots. (b) A closer view of the active site of cholesterol oxidase shown in a similar format, except that FAD is also shown in line form; several active site water molecules close to FAD are shown as red van der Waals spheres. The isoalloxazine ring system of FAD is treated with SCC-DFTB, and the rest of FAD and protein atoms are described with the CHARMM 22 force field for proteins; the dash-line indicates the QM/MM boundary. Note that the side chain of Asn 485 is oriented according to the original X-ray structure (1B4V);[46] recent mutation experiments[47] and sub-Å resolution data[45] suggest that a more appropriate orientation involves NH_2 points toward the isoalloxazine ring. Reduction potentials with both orientations have been computed (see Table 1). The figure was prepared with VMD; Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33.

work.[58] Mainly because of approximations made to the two-electron integrals, it is computationally very efficient with a speed similar to that of AM1 and PM3. Benchmark calculations on the energies, geometries and vibrational frequencies of small organic molecules, peptides, base pairing and stacking interactions,[59−61] and complexes involving zinc ion(s)[62] indicated that SCC-DFTB often gives superior results compared to AM1 and PM3. Applications to proton and hydride transfer reactions in several enzymes have also been successful.[58,63] Specifically to the present calculations, as shown in the Supporting Information, SCC-DFTB gave satisfactory results for the redox properties of a model FAD (7,8-dimethyl isoalloxazine) compared to B3LYP[64−66] calculations.

As the starting structure, the X-ray result[46] for *Streptomyces* at 1.5 Å (PDB code 1B4V) with the cofactor FAD bound to the enzyme active site was used for the wild-type enzyme. The stochastic boundary setup[67−69] was used, with a radius 25 Å centered at N^5 in FAD. The final model included 6175 protein atoms and 539 water molecules (Figure 1a).

In the QM/MM free energy perturbation calculations, 11 windows were chosen with λ ranging from 0.0 to 1.0 with a 0.1 interval. As discussed in section II, the dual-topology-single-coordinate approach does not suffer from end-point problems and allows molecular dynamics calculations at $\lambda = 0.0$ and 1.0. For each window, the system was first equilibrated for 50 ps and then a production run was carried out up to 300 ps until the free energy derivative converges; that is, totally more than 4 ns trajectories were used for the free energy calculations. In the MD calculations, a time-step of 1.0 fs was used and all of the bonds involving hydrogen atoms were constrained with SHAKE.[70] The temperature was set to be 300 K with mixed Langevian/Newtonian dynamics unique to the stochastic boundary procedure.[67]

As discussed in a previous report,[22] it is crucial to carefully treat long-range electrostatics in redox potential calculations. To this end, a Poisson−Boltzmann (PB) charge-scaling scheme[71] was introduced to account for solvent shielding in addition to that from the explicit water molecules included in the model. The scaling factors are summarized in the Supporting Information. Following the QM/MM free energy calculations with the scaled partial charges, Poisson−Boltzmann (PB)[72] calculations were carried out for each λ window to obtain the charge-scaling corrections and bulk solvation free energies. One hundred configurations were taken for each λ value. Mulliken charges from SCC-DFTB/CHARMM calculations were used for the QM atoms in the PB calculations. Dielectric constants of 1.0 and 80.0 were used for the protein and bulk solvent, respectively; the grid size was set to 0.4 Å. A value of 1.0 rather than the popular value of 4.0[73] was chosen as the protein dielectric constant because conformational fluctuations (and therefore part of the protein dielectric response) were included in the MD calculations. The solvation corrected reduction free energy has the following form:

$$\Delta F_{\text{full}}^{\text{red}}(Q) = \Delta F_{\text{pert}}(q) + \Delta F_{\text{PB}}^{\text{scl}}(Q \leftarrow q) + \Delta F_{\text{PB}}^{\text{slv}}(Q) \quad (16)$$

where the three terms on the rhs indicate components from QM/MM free energy perturbation calculations, charge-scaling correction PB calculations, and solvation free energy PB calculations, respectively. The last two terms will be collectively referred to as "solvation correction" in the following discussions, and $\Delta F_{\text{full}}^{\text{red}}(Q)$ will be referred to as solvation corrected reduction free energy. The notation "q" and "Q" were used to emphasized that the scaled and full set of partial charges, respectively, were used for the charged side chains.

Concerning contributions from amino acids in the active site, we have focused on Asn 485 in the present work because it is the only quantity that has been characterized experimentally so far.[47] First, it was not possible to determine the precise orientation of the Asn 485 side chain with the X-ray data at 1.5 Å resolution. The original structure (1B4V)[46] modeled the side chain such that the NH_2 group points away from the FAD isoalloxazine ring, whereas the recent mutation results for Asn 485 suggested that a more appropriate model should arrange the $-NH_2$ to point toward the isoalloxazine ring, such that the extra electron density on FAD upon reduction can be stabilized through a dipole−π interaction.[47] To explicitly illustrate the effect of side chain orientation on the reduction potential, calculations were performed with both orientations. To further quantify the contribution of Asn 485, three types of calculations were performed. First, a perturbative analysis has been carried out by setting the partial charges on the Asn 485 side chain to zero and recomputing the free energy derivatives and solvation corrections; the results are perturbative in the sense that configurations from the MD trajectories for the wild type simulation are used. This procedure corresponds to that typically done in QM/MM analysis of reaction mechanisms.[74,75] In the second set of calculations, Asn 485 was manually mutated into a Leu residue, based on the wild-type X-ray structure (1B4V), and the full set of reduction potential calculations (QM/MM free energy perturbation and solvation corrections) were carried out. Finally, a third set of reduction potential was computed for the Asn485L mutant using the actual X-ray structure for the mutant (1IJH).[47] Formally, only the third set of simulation produces experimentally measurable redox potential for the mutant, whereas the first two sets of calculations reflect the "intrinsic" contribution from Asn 485 because no additional structural changes accompanying the mutation (rotation of Met 122 and the presence of the extra water molecule)[47] are implicated. It is very interesting to explore the numerical differences between the three sets of results, as we do below.

III.2. Free Energy Derivatives and the Absolute Reduction Potential. As shown in Figure 2, the free energy derivatives without solvation corrections (associated with $\Delta F_{\text{pert}}(q)$) show similar convergence behavior at different λ values; that is, there is no significant statistical problem associated with the end points ($\lambda = 0$ and 1), and the free energy derivatives remain finite at the extreme λ values. These are due to the fact that no atoms were created or annihilated in the dual-topology-single-coordinate approach when the numbers of atoms are the same in the two chemical states; as shown in eq 15, van der Waals terms make no explicit contributions at all.

The free energy derivatives depend on the coupling parameters in a rather linear fashion, which is qualitatively consistent with the idea that the linear dielectric response is a good approximation even for heterogeneous systems such as proteins, which was also found in previous simulations of several other redox proteins such as cytochrome c.[76] Because a linear response model, Poisson−Boltzmann, was used in obtaining the charge scaling and bulk solvation corrections, it is expected that the free energy derivative including these corrections remains to be linear as a function of the coupling parameter, as observed in Figure 2b. The highly linear behavior of the solvation corrected free energy derivatives found here justifies to a certain extent the use of continuum solvation models in estimating redox properties of cofactors in protein systems.[77,78] However, the QM/MM free energy simulation protocol is more robust in the sense that the result is less sensitive to the dielectric "constant" for

Free Energy Perturbation Calculations

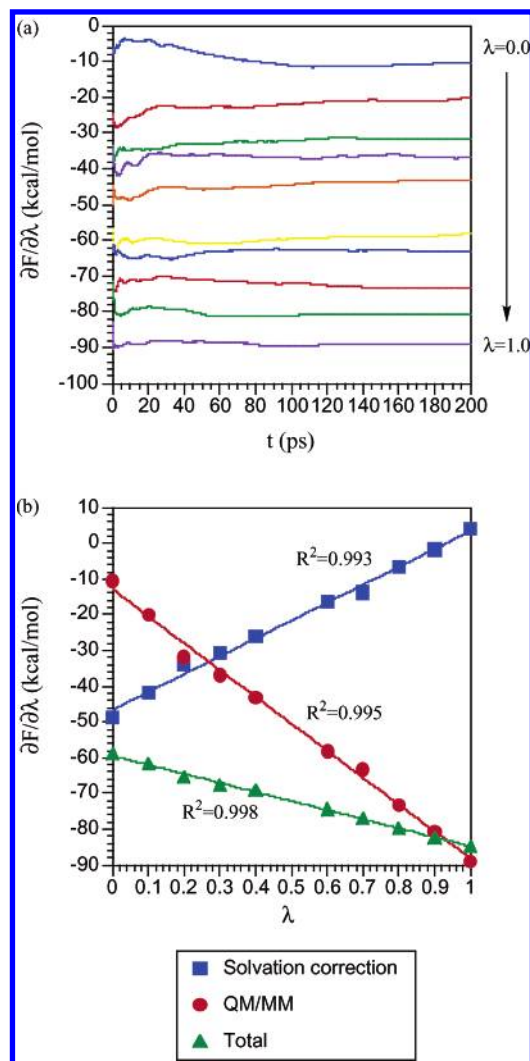*J. Phys. Chem. B, Vol. 107, No. 33, 2003* **8649**

Figure 2. Results from SCC-DFTB/CHARMM free energy perturbation calculations for the wild-type cholesterol oxidase with the side chain of Asn 485 oriented such that $-NH_2$ points toward the isoalloxazine ring of FAD. (a) The convergence behavior of free energy derivatives (without solvation corrections) as a function of production simulation length. (b) The behavior of solvation corrected free energy derivative and its two components as a function of the coupling parameter, λ: QM/MM calculations with the scaled partial charges on charged side chains, and solvation corrections that include both charge-scaling correction and bulk solvation (eq 16). Note the linear behavior of free energy derivative and its components.

the protein system, which has different values under different contexts.[79,80]

In the framework of Marcus's model for oxidation—reduction reactions in the condensed phase,[81,82] the difference between free energy derivatives at two end points ($\lambda = 0, 1$) reflects the reorganization (free) energy for the process. Taking the wild-type result in Figure 2 for example, the reorganization energy from QM/MM free energy calculations without solvation correction is -39.3 kcal/mol, which indicates that there is substantial structural reorganization in the protein (and the explicit water molecules included in the simulation) upon reduction of FAD. As shown in Figure 3, several active site residues, especially Glu 361, change positions substantially in the presence of an additional electron on FAD. The average water structure is also slightly different in the two redox states; a larger number of water molecules trapped in the active site are drawn closer to the isoalloxazine ring in the reduced state (Figure 3). Although local structural changes are expected for
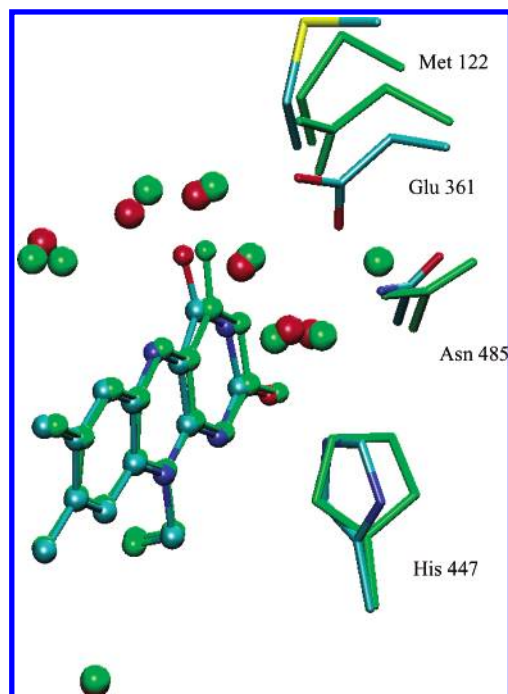

Figure 3. Superposition of the averaged structures of several residues and water molecules in the active site of wild-type cholesterol oxidase with FAD in the oxidized (colored according to atom type) and reduced (in green) states. Hydrogen atoms are neglected for clarity. The structures have been averaged over 200 ps of MD simulations using a SCC-DFTB/CHARMM potential (see text) and are collected from the free energy perturbation calculations at the two end points ($\lambda = 0$ and 1), which are made possible by the double-topology-single-coordinate scheme proposed in the present work.

the reduction process, the calculated reorganization energy is rather large compared to previous studies on redox proteins; for the reduction of heme in cytochrome c, for example, the reorganization energy was calculated to be $9-15$ kcal/mol, which is much smaller than the value in solution ($\sim$37 kcal/mol).[79] Interestingly, although the charge-scaling correction has a very small λ dependence ($\sim$0.1 kcal/mol reorganization), the bulk solvation component has a significant reorganization energy of $+26.3$ kcal/mol. As a result, the solvent corrected reorganization energy for FAD reduction in Chox is 12.9 kcal/mol, which is much closer to previous estimates for other redox proteins.[79] These results indicate that the dielectric response from the protein atoms per se (our current model includes also 539 water molecules) to the reduction of FAD is substantial, and the magnitude of the response is not sensitive to the charge scaling protocol; gas phase model calculations (see the Supporting Informatiojn) indicate that the intra-FAD reorganization energy is relatively small, on the order of 5 kcal/mol at the B3LYP/6-311+G(d,p)[83,84] level. The bulk solvent also responds significantly to the reduction process, despite the fact that the FAD isoalloxazine ring is not solvent exposed at all (Chox has a hydrophobic active site to adopt its substrate cholesterol); the sign of the solvent response, however, is opposite to that of the protein, such that the net reorganization of the solvated protein is fairly small, which has been proposed to be an important factor for efficient electron transfers in proteins relative to in solution.[85] Preliminary calculations indicate that the reorganization of solvent for FAD reduction in solution is indeed much larger than that in Chox. Thus the present calculations emphasized the importance of bulk solvent in compensating the reorganization of protein atoms in the reduction process. During the catalytic cycle, however, we expect reorganizations of the

TABLE 1: Various Components for the Calculated First Reduction Potential of FAD in Cholesterol Oxidase (Chox)[a]

system	QM/MM	solvation correction	total
WT Asn485 O1[b]	−47.3	−22.0	−69.3/−1425 (−1121)
WT Asn485 O2[b]	−50.7	−21.5	−72.2/−1299 (−995)
WT Asn485 perturbation[c]	[−5.5]	[4.2]	[−1.3/−56]
Asn485L manual[d]	−43.1 [−7.6]	−27.6 [6.1]	−70.8[−1.5/−65]
Asn485L X-ray[d]	−48.7 [−2.0]	−22.2 [0.7]	−70.9 [−1.3/−56]

[a] The QM/MM column indicates results from SCC-DFTB/CHARMM free energy perturbation calculations with scaled partial charges (see the Supporting Information) on the charged side chains; the charges on Glu 361 and Lys 225 were not scaled down due to the proximity to the isoalloxazine. The "solvation correction" component includes both charge-scaling and bulk solvation contributions (eq 16), both from Poisson−Boltzmann calculations (see text). Numbers are generally given in kcal/mol, whereas values after slashes are in mV. The mV values are reduction potentials converted relative to the hydrogen electrode (4.43 eV).[86] The experimental value for the first reduction potential of FAD in wild-type Chox in the absence of substrate is −498 mV (note that the values are given relative to Safranin T in the original reference,[47] which has a reduction potential of −249 mV relative to the hydrogen electrode). The numbers in parentheses are the contribution of Asn 485 calculated in various ways (see text and other footnotes below); a negative value indicates a favorable contribution toward reduction (i.e., stabilizes FAD⁻). The experimental estimation based on mutation studies is 78 mV (1.8 kcal/mol). [b] O1 and O2 indicate the different orientations of the Asn485 side chain, which has the −NH₂ pointing away and toward the isoalloxazine ring of FAD, respectively. [c] The contribution due to Asn 485 was estimated by removing the partial charges on the side chain of Asn 485 and computing changes in various components of the reduction potential (QM/MM free energy derivative with the scaled partial charges and solvation corrections, eq 16). For each λ window, 100 configurations were taken from the MD simulations of the wild-type Chox. [d] In the "Asn485L manual" calculations, the side chain of Asn 485 was changed to a leucine residue based on the X-ray structure of the wild-type enzyme (1B4V);[46] in the "Asn485L X-ray" calculations, the X-ray structure of the Asn485L mutant (1IJH)[47] was used, which showed that the side chain of Met212 also changed orientation upon the Asn485L mutation, and an additional water molecule appeared in the active site as a result. Note that the full set of QM/MM free energy perturbations and solvation corrections have been carried out in both simulations.

protein and solvent to be substantially smaller because the electron comes from the nearby substrate, which is absent in the present simulations. It would be of interest to perform the similar analysis for flavoproteins whose major function is electron-transfer rather than catalysis, such as flavodoxin.

As to the quantitative aspects of the calculated reduction potential, charge scaling and bulk solvation corrections make significant contributions as commented in an earlier report.[22] For the wild-type Chox with −NH₂ of Asn 485 pointing toward FAD, which is believed to be the correct orientation (also see below),[47] the reduction potential relative to the hydrogen electrode is −2231 mV based on SCC-DFTB/MM free energy calculation with the scaled partial charges on charged side chains ($\Delta F_{pert}(q)$); this is very different from the experimental value of −498 mV.[47] With the charge scaling and bulk solvation corrections taken into account, the computed reduction potential becomes −1299 mV, which is much closer to the experimental value. Finally, considering the difference between SCC-DFTB and B3LYP results for the reduction potential for a FAD analogue in the gas phase (see the Supporting Information), the best estimate from the present work is −995 mV. This value is still substantially different from the experimental measurement, and the precise origin for the discrepancy is currently under investigation. Although the MM treatment of the active site and quality of the force field certainly make contributions, the magnitude of the difference seems to indicate that certain aspects of the active site are different in the present simulation compared to the realistic protein system. In fact, the recent high-resolution structure of Chox[45] indicates that His 447 might be protonated at the Nε position rather than the originally proposed Nδ position. Because His 447 directly interacts with FAD, its protonation pattern is expected to have a substantial influence on the reduction potential. Calculations with His 447 in different protonation states are currently underway using the latest sub-Å resolution X-ray structure,[45] and the results will be reported separately because the major focus of the current work is to illustrate the convergence behavior of QM/MM free energy simulations.

III.3. Effect of Asn 485−side Chain Orientation, Dipolar-π Interaction and Unexpected Structural Change Upon Mutation. At 1.5 Å resolution, it is difficult to assign the side chain orientation of Asn 485 in Chox, and the original X-ray structure

had the −NH₂ group pointing away from the isoalloxazine ring.[46] A recent mutation study[47] found that Asn 485 favors the reduction of FAD, and therefore, it seems to be more appropriate to have the −NH₂ group pointing toward FAD such that the additional electron density on the isoalloxazine ring upon reduction is stabilized through a dipolar−π type of interaction. The present calculations support the latter assignment. As shown in Table 1, the calculated reduction potential is closer to the experimental measurement by 3 kcal/mol (126 mV) with the second set of orientation of Asn 485. More convincingly, with this side chain orientation, the contribution due to Asn 485 estimated from several different sets of calculations is in close agreement with that inferred from the mutation experiment. We note that the inherent uncertainty associated with the calculated contribution to the reduction potential is expected to be smaller than that for the absolute reduction potential due to error cancellations.

In the first set of calculations, the contribution due to Asn 485 was estimated with a perturbative approach similar to what is typically done in QM/MM reaction mechanism analysis.[74,75] As shown in Figure 4, long-range electrostatic interactions also have a substantial influence on the computed contribution from Asn 485; it is about 5.5 kcal/mol with SCC-DFTB/CHARMM calculations using the scaled partial charges, but reduced to 1.3 kcal/mol when long-range electrostatics are taken into account. Interestingly, although the free energy derivatives depend sensitively on the coupling parameter, the contribution from Asn 485 shows little λ dependence. This suggests that Asn 485 does not contribute significantly to the reorganization energy for the FAD reduction; this is consistent with the observation that the position of Asn 485 undergoes little change during FAD reduction (Figure 3).

In the other two sets of calculations, free energy perturbation and solvation correction calculations were repeated for the Asn485L mutant enzyme. The solvation corrected free energy derivatives are highly linear for both mutant simulations (Figure 4b), similar to the wild type results. Somewhat unexpectedly, the calculated reduction potential depends little on the structure used for the mutant, when the solvation corrections are taken into account (Table 1). The inferred contributions for Asn 485 from the two simulations, 1.3 and 1.5 kcal/mol, respectively, are close to the perturbative result based on the wild type MD
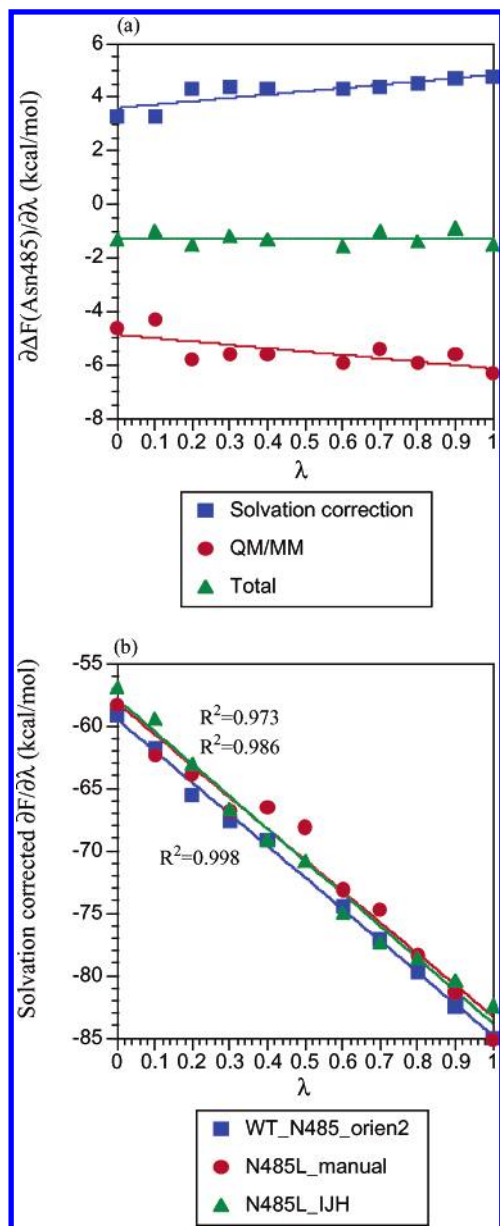
Free Energy Perturbation Calculations

J. Phys. Chem. B, Vol. 107, No. 33, 2003 **8651**

Figure 4. Results for estimating the contribution of Asn 485 to the first reduction potential of FAD. (a) Estimations based on a perturbative approach, in which the partial charges on Asn 485 are set to zero and the shift in the reduction potential was calculated using the MD trajectories for the wild-type enzyme. Note that the contribution has little *l* dependence, which suggests that Asn 485 does not contribute much to the reorganization energy of the protein (see text) during the FAD reduction. (b) Comparison of solvation corrected free energy derivatives for the wild-type Chox and two simulations for the Asn485L mutant. In the latter two simulations, the starting structures are obtained from mutating Asn485 to leucine in the wild-type enzyme (PDB code 1B4V) and from the X-ray structure of Asn485L (PDB code 1IJH), respectively.

trajectories (1.3 kcal/mol). All of these values are also close to the experimentally measured shift (78 mV $\sim$ 1.8 kcal/mol) in FAD reduction potential when Asn 485 was mutated into a leucine.[47]

The X-ray structure of Asn485L (PDB code 1IJH) showed that the side chain of Met 122 also changed its conformation upon mutation, and an extra water molecule appeared in the area originally occupied by Met 122.[47] Therefore, it was suggested that the observed shift in the reduction potential (78 mV) is a lower limit for the absolute contribution of Asn 485. During the two sets of mutant simulations, the side chain
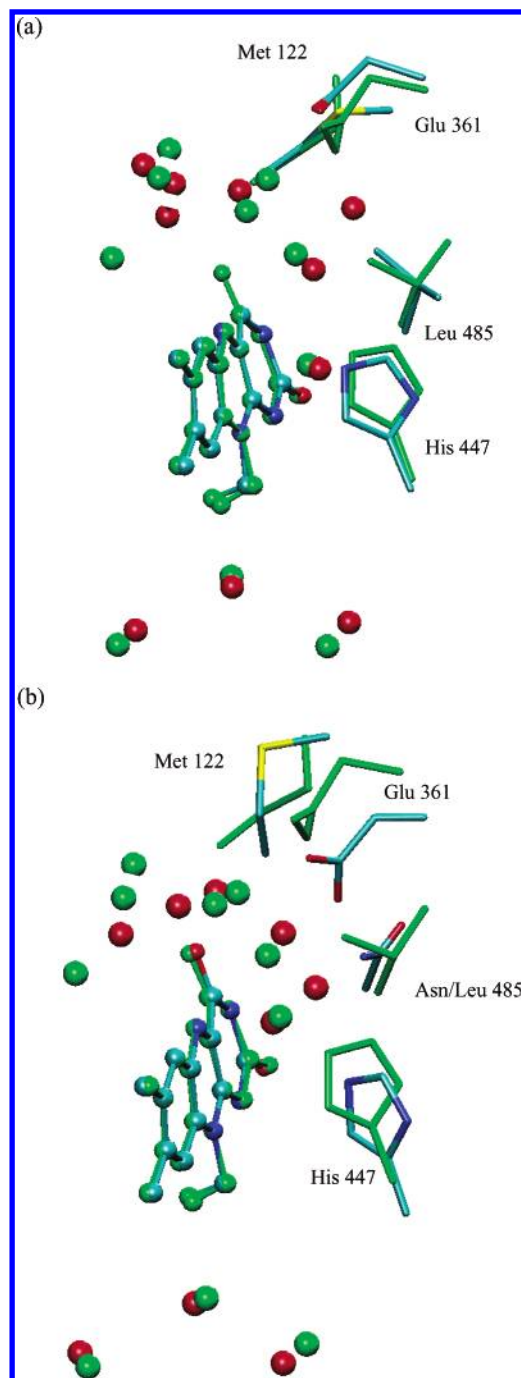

Figure 5. Superposition of averaged structure of several amino acids and water molecules in the active site of cholesterol oxidase when FAD is in the oxidized state. Hydrogen atoms are neglected for clarity. The structures have been averaged over 200 ps of MD simulations using a SCC-DFTB/CHARMM potential (see text). (a) The Asn 485 L mutant simulation with the corresponding X-ray structure (in green, PDB code 1IJH) compared to that using a manually mutated structure based on the wild-type X-ray data (color-coded based on atom type). Note the difference in the orientation of the Met 122 side chain (see text). (b) The Asn 485 L mutant result using the X-ray structure (PDB code 1IJH) compared to that of the wild-type simulation (PDB code 1B4V).

orientation of Met 122 remained different because the starting structures are based on the wild type and mutant X-ray structures, respectively. As seen in Figure 5a, however, the average positions of other active site residues and the distribution of water molecules are rather similar in the "manual mutation" simulation and the one using the X-ray structure for the Asn485L mutant. Comparing with the wild type simulation

(Figure 5b), the positions of His 447 and Glu 361 are substantially different and there is one more water molecule close (within 4 Å) to the isoalloxazine ring in the mutant simulations. Therefore, these structural features suggest that the side chain orientation of Met 122 is a secondary effect compared to the nature of residue 485 (i.e., Asn vs Leu). This is consistent with the fact that three different calculations (using either different structures or different ways to estimate contributions, Table 1) gave rather similar estimations on the contribution of Asn 485, and the results are all close to the experimentally observed shift in FAD reduction potential upon the Asn485L mutation. The minor importance of Met 122 will be further verified by calculations with the most recent sub-Å resolution X-ray structures[45] and future mutation experiments involving this residue.

IV. Conclusions

The combined quantum mechanical/molecular mechanical (QM/MM) approaches have been widely used to probe the mechanism of chemical reactions in solution and biological systems.[8,9,12,63] Much less work has been done, however, to employ QM/MM potential in free energy perturbation calculations. It is meaningful to fill such a void because the study of many chemical processes that involve complex electronic structural changes (e.g., redox reactions involving transition metal ions) requires a reliable description of the potential energy that is difficult to achieve with pure classical force fields. Many properties (such as *absolute* redox potentials) are also difficult to compute with pure classical force fields. Although we realize that it is often the *relative* quantities (e.g., the effect of mutation on the redox potential) that are of major interest, a comparison between absolute values with experimental measurements can serve as a convincing validation of the simulation protocols, which is a crucial element of computational studies of complex systems.

In the current work, we discussed the basic theoretical and computational frameworks associated with using QM/MM potential in free energy perturbation simulations. We pointed out several formal inconveniences of QM/MM free energy simulations, which arise from the fact that QM/MM interactions are not trivially separable as standard classical force fields, and that it is difficult to perform QM/MM calculations with "fractional" electronic Hamiltonians. To circumvent these problems, we propose a unique coupling scheme that is inspired by a combination of the conventional dual and single topology approaches,[18] which we term the "dual-topology-single-coordinate" scheme. In this scheme, both subsystems of interest are present in the simulation but are constrained to adopt exactly the same set of Cartesian coordinates; in so doing, simulations can be performed at the end-points ($\lambda = 0, 1$) without numerical or sampling instabilities. Moreover, the scheme combines the merits of the conventional dual- and single-topology approaches, which include that it is straightforward to compute both the coupling potential and the free energy derivative, it is not necessary to consider corrections related to artifactual reference states or Jacobian factors, and it offers rapid numerical convergence. The method is ideally suited for studying processes that conserve the number of atoms, such as redox cycle, absorption/emission events, and binding exchange of metal ions to macromolecules. It is also straightforward to extend the method to problems that involve a small change in the number of atoms, such as predicting pK_a's of amino acids in proteins.

The proposed method was applied to calculate the absolute reduction potential of FAD in cholesterol oxidase and the contribution from a nearby conserved residue, Asn 485.[47] With a hybrid SCC-DFTB/CHARMM potential[56,58] and a charge-scaling technique[71] that deals with the long-range electrostatic effects in solution, the absolute potential was found to be in decent agreement with experimental measurements. In addition to the approximations in the SCC-DFTB/MM potential functions, the remaining discrepancy between simulation and experimental data is likely to have a substantial contribution from the fact that the protonation pattern for an active site histidine in the simulation is different from that observed in the recent high-resolution structure.[45] The present simulations estimated that the "intrinsic" contribution from Asn 485 to the FAD reduction potential is on the order of 1.3 kcal/mol, which is fairly close to the experimental estimate based on mutation studies.[47] Finally, the calculations emphasized that the protein responds in an approximately linear fashion to the reduction of FAD, and has a fairly small (∼13 kcal/mol) reorganization energy; the bulk solvent was found to play an important role in compensating the reorganization of the protein.

Supporting Information Available: The scaling factors calculated for the charged residues are given. Moreover, results for benchmarking the SCC-DFTB method using a FAD analogue (7,8-dimethyl isoalloxazine) as compared to B3LYP/6-311+G(d,p) calculations are also given. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2003**, *35*, 430.

(2) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395.

(3) Straatsma, T. P.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407−435.

(4) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184−189.

(5) Archontis, G.; Simonson, T.; Moras, D.; Karplus, M. *J. Mol. Biol.* **1998**, *275*, 823−846.

(6) Boresch, S.; Karplus, M. *J. Mol. Biol.* **1995**, *254*, 801−807.

(7) Gao, J.; Kucezera, K.; Tidor, B.; Karplus, M. *Science* **1989**, *244*, 1069.

(8) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.

(9) Gao, J. In *Reviews in Computational Chemistry*; Lipkowitza, K. B., Boyd, D. B., Eds.; VCH: New York, 1996; Vol. 7, p 119.

(10) Åqvist, J.; Warshel, A. *Chem. Rev.* **1993**, *93*, 2523.

(11) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *Biochem.* **1996**, *35*, 7819−7833.

(12) Field, M. J. *J. Comput. Chem.* **2002**, *23*, 48.

(13) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.

(14) Cui, Q.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 3093−3124.

(15) Shoeib, T.; Ruggiero, G. D.; Siu, K.; Hopkinson, A. C.; Williamds, I. H. *J. Chem. Phys.* **2002**, *117*, 2762.

(16) Gao, J. *J. Phys. Chem.* **1992**, *96*, 537.

(17) Gao, J. L.; Xia, X. F. *J. Am. Chem. Soc.* **1993**, *115*, 9667−9675.

(18) Simonson, T. In *Computational biochemistry and biophysics*; Becker, O. M., Mackerell, A. D. J., Roux, B., Watanabe, M., Eds.; Marcel Dekker, Inc.: New York, 2001.

(19) Bartels, C.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 865−880.

(20) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.

(21) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(22) Formaneck, M. S.; Li, G.; Zhang, X.; Cui, Q. *J. Theor. Comput. Chem.* **2002**, *1*, 53−67.

(23) van Gunsteren, W.; Beutler, T.; Fraternali, F.; King, P.; Mark, A.; Smith, P. In *Computer simulation of biomolecules systems*; van Gunsteren, W., Weiner, P., Wilkinson, A., Eds.; ESCOM Science: Leiden, 1993.

(24) Brooks, C. L., III.; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, & Thermodynamics*; John Wiley & Sons: New York, 1988; Vol. LXXI.

(25) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1975.

(26) Hummer, G.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 2004.

(27) Zwanzig, R. *J. Chem. Phys.* **1954**, *22*, 1420.

(28) Archontis, G.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 11246−11260.

(29) Pearlman, D. A. *J. Phys. Chem.* **1994**, *98*, 1487.

(30) Boresch, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 5145.

(31) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 103−118.

(32) Boresch, S.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 119−136.

(33) Hermans, J.; Wang, L. *J. Am. Chem. Soc.* **1997**, *119*, 2702.

(34) Roux, B.; Nina, M.; Pomes, R.; Smith, J. *Biophys. J.* **1996**, *71*, 670.

(35) Simonson, T. *Mol. Phys.* **1993**, *80*, 441.

(36) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. *Chem. Phys. Lett.* **1994**, *222*, 529.

(37) Feynmann, R. P. *Phys. Rev.* **1939**, *41*, 721.

(38) Hellmann, H. *Einfurhrung in die Quantencgemie*; Leipzig, 1937.

(39) Hurley, A. C. *Proc. R. Soc. A* **1954**, *226*, 170−179.

(40) Yamaguchi, Y.; Osmura, Y.; Goddard, J. D.; Schaefer, H. F. I. *A new dimension to quantum chemistry: analytical derivative methods in ab initio molecular electronic structure theory*; Oxford University Press: Oxford, U.K., 1994.

(41) Szabo, A.; Ostlund, N. S. *Modern quantum chemistry−introduction to advanced electronic structure theory*; Dover Publications: New York, 1989.

(42) Yang, W.; Parr, R. G. *Density Function Theory of Atoms and Molecules*; Oxford University Press: Oxford, 1989.

(43) Herschbach, D. R.; Johnston, H. S.; Rapp, D. *J. Chem. Phys.* **1959**, *31*, 1652.

(44) Sampson, N. S. *Antioxidants and Redox Signaling*; 2001; Vol. 3, p 839.

(45) Lario, P. I.; Sampson, N. S.; Vrielink, A. *J. Mol. Biol.* **2003**, In press.

(46) Li, J.; Vrielink, A.; Brick, P.; Blow, D. M. *Biochemistry* **1993**, *32*, 11507.

(47) Yin, Y.; Sampson, N. S.; Vrielink, A.; Lario, P. I. *Biochemistry* **2001**, *40*, 13779.

(48) Yue, Q. K.; Kass, I. J.; Sampson, N. S.; Vrielink, A. *Biochemistry* **1999**, *38*, 4277.

(49) Kass, I. J.; Sampson, N. S. *Biochemistry* **1998**, *37*, 17990.

(50) Sampson, N. S.; Kass, I. J. *J. Am. Chem. Soc.* **1997**, *119*, 855.

(51) Cavener, D. R. *J. Mol. Biol.* **1992**, *223*, 811.

(52) Levitt, M.; Perutz, M. F. *J. Mol. Biol.* **1988**, *201*, 751.

(53) Perutz, M. F.; Fermi, G.; Abraham, D. J.; Poyart, C.; Bursaux, E. *J. Am. Chem. Soc.* **1986**, *108*, 1064.

(54) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Kraplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(55) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720−1735.

(56) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seigert, G. *Phys. Rev.* **1998**, *B58*, 7260−7268.

(57) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(58) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569−585.

(59) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. *Chem. Phys.* **2000**, *256*, 15−27.

(60) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. *Chem. Phys.* **2001**, *263*, 203-219.

(61) Bohr, H. G.; Jalkanen, K. J.; Elstner, M.; Frimand, K.; Suhai, S. *Chem. Phys.* **1999**, *246*, 13−36.

(62) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, In press.

(63) Cui, Q.; Elstner, M.; Karplus, M. *J. Phys. Chem. B* **2002**, *106*, 2721−2740.

(64) Becke, A. D. *Phys. Rev. A* **1998**, *38*, 3098−3100.

(65) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(66) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(67) Brooks, C. L., III.; Brünger, A.; Karplus, M. *Biopoly.* **1985**, *24*, 843.

(68) Brooks, C. L., III.; Karplus, M. *J. Mol. Biol.* **1989**, *208*, 159−181.

(69) Brooks, C. L., III.; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312−6325.

(70) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327−341.

(71) Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem. B* **1997**, *101*, 8349−8362.

(72) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.

(73) Simonson, T.; Perahia, D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1082.

(74) Cui, Q.; Karplus, M. *J. Am. Chem. Soc.* **2001**, *122*, 2284−2290.

(75) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826−5832.

(76) Simonson, T. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6544−6549.

(77) Konecny, R.; Li, J.; Fisher, C. L.; Dillet, V.; Bashford, D.; Noodleman, D. A. *Inorg. Chem.* **1999**, *38*, 940.

(78) Mouesca, J. M.; Chen, J. L.; Noodleman, D. A.; Bashford, D.; Case, D. A. *J. Am. Chem. Soc.* **1994**, *116*, 11898.

(79) Muegge, I.; Qi, P. X.; Wand, A. J.; Chu, Z. T.; Warshel, A. *J. Phys. Chem.* **1997**, *101*, 825−836.

(80) Schutz, C. N.; Warshel, A. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 400.

(81) Marcus, R. A. *J. Chem. Phys.* **1956**, *24*, 966−978.

(82) Marcus, R. A. *J. Phys. Chem.* **1968**, *72*, 891−899.

(83) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222.

(84) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650−654.

(85) Parson, W. W.; Chu, Z. T.; Warshel, A. *Biophys. J.* **1998**, *74*, 182−191.

(86) Reiss, H.; Heller, A. *J. Phys. Chem.* **1985**, *89*, 4207.