Reaction path potential for complex systems derived from combined ab initio quantum mechanical and molecular mechanical calculations

Zhenyu Lu and Weitao Yang

Citation: The Journal of Chemical Physics **121**, 89 (2004); doi: 10.1063/1.1757436
View online: http://dx.doi.org/10.1063/1.1757436
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/121/1?ver=pdfcov
Published by the AIP Publishing

Reaction path potential for complex systems derived from combined *ab initio* quantum mechanical and molecular mechanical calculations

Zhenyu Lu and Weitao Yang[a)]
Department of Chemistry, Duke University, Durham, North Carolina 27708

(Received 16 February 2004; accepted 9 April 2004)

Combined *ab initio* quantum mechanical and molecular mechanical calculations have been widely used for modeling chemical reactions in complex systems such as enzymes, with most applications being based on the determination of a minimum energy path connecting the reactant through the transition state to the product in the enzyme environment. However, statistical mechanics sampling and reaction dynamics calculations with a combined *ab initio* quantum mechanical (QM) and molecular mechanical (MM) potential are still not feasible because of the computational costs associated mainly with the *ab initio* quantum mechanical calculations for the QM subsystem. To address this issue, a reaction path potential energy surface is developed here for statistical mechanics and dynamics simulation of chemical reactions in enzymes and other complex systems. The reaction path potential follows the ideas from the reaction path Hamiltonian of Miller, Handy and Adams for gas phase chemical reactions but is designed specifically for large systems that are described with combined *ab initio* quantum mechanical and molecular mechanical methods. The reaction path potential is an analytical energy expression of the combined quantum mechanical and molecular mechanical potential energy along the minimum energy path. An expansion around the minimum energy path is made in both the nuclear and the electronic degrees of freedom for the QM subsystem internal energy, while the energy of the subsystem described with MM remains unchanged from that in the combined quantum mechanical and molecular mechanical expression and the electrostatic interaction between the QM and MM subsystems is described as the interaction of the MM charges with the QM charges. The QM charges are polarizable in response to the changes in both the MM and the QM degrees of freedom through a new response kernel developed in the present work. The input data for constructing the reaction path potential are energies, vibrational frequencies, and electron density response properties of the QM subsystem along the minimum energy path, all of which can be obtained from the combined quantum mechanical and molecular mechanical calculations. Once constructed, it costs much less for its evaluation. Thus, the reaction path potential provides a potential energy surface for rigorous statistical mechanics and reaction dynamics calculations of complex systems. As an example, the method is applied to the statistical mechanical calculations for the potential of mean force of the chemical reaction in triosephosphate isomerase.
 [DOI: 10.1063/1.1757436]

I. INTRODUCTION

There have been many developments in applying computational methods to the study of enzymatic catalysis (for recent reviews, see Refs. 1–8). In particular, much work has been carried out to achieve an accurate description of chemical reactions in enzymes. A widely used approach to studying enzyme reactions is to combine quantum mechanical (QM) methods with molecular mechanical (MM) force fields (QM/MM).[9–15] Warshel and Levitt developed the basic idea of QM/MM approaches in their seminal paper.[16] In the last two decades, there has been a great deal of development of such methods by a number of groups.[17–24] In the process of chemical reactions, only a small number of atoms participate in the bond forming or breaking processes, while many other atoms in the system generally serve as a steric and electrostatic environment to influence the properties and reactivities of the active site. Therefore, the small active site can be treated by quantum mechanical methods, while the rest with numerous particles can be described by molecular mechanical methods. The combined QM/MM methodology has allowed reliable electronic structure calculations for enzymatic reactions with a realistic and atomistic description of enzyme environment. Such a combined QM and MM approach can take advantage of the applicability and accuracy of the QM methods for chemical reactions in systems of several dozens of atoms and of the computational efficiency of the MM description for the rest of the enzyme and solvent, which normally consists of thousands of atoms.

The QM/MM approach can be classified into two types according to the level of theory used in the electronic structures calculations. One type uses semiempirical methods such as AM1 and PM3[18–20] while the other uses *ab initio* molecular orbital theory or density functional theory (DFT).[22–27] With an adequate basis set, the latter is more accurate in describing chemical reactions. But it is also much

a)Author to whom correspondence should be addressed. Electronic mail: weitao.yang@duke.edu

more demanding computationally, mainly in the *ab initio* quantum mechanical calculations for the QM subsystem. Because of the computational costs, rigorous statistical mechanics sampling and reaction dynamics calculations with an *ab initio* QM/MM potential are still not feasible. Therefore, *ab initio* QM/MM calculations are usually limited to the determination of a minimum energy path, which connects the reactant, transition state, and product smoothly.

We have partially addressed the issue of how to go beyond just the minimum energy path calculations in our recently developed *ab initio* QM/MM free energy approach to studying enzyme reactions.[23,24,28] This approach consists of three major components: the pseudobond *ab initio* QM/MM method which provides a smooth interface between the QM and MM subsystem and thus a well-defined potential energy surface, the efficient iterative optimization procedure which determines the reaction paths within a realistic enzyme environment, and free energy calculations which take into account the fluctuation of the enzyme system. To calculate the free energy of activation (or the potential of mean force), we made two assumptions: (1) the dynamics of the QM and MM subsystems is independent of each other and (2) the QM subsystem fluctuations are harmonic. Then we calculated the contribution to the free energy from the fluctuations in the MM subsystem with the free energy perturbation method and the contribution to the free energy from the fluctuations in the QM subsystem with harmonic frequency calculations. The method has been applied to various enzymes: enolase,[25] 4-oxalocrontonate tautomerase,[27] triosephosphate isomerase (TIM),[28] and other enzymes.[29,30] Unlike the method combining gas phase QM calculations with free energy simulations (the QM-FE approach),[31–34] in our approach the reaction path is determined in the enzyme environment rather than in the gas phase. We do not need to assume that the enzyme environment has little or no influence on the course of the reaction and that it only affects the energetic profile. Although the gas phase path assumption has been supported by successful applications for small organic molecules in solvent, it may not be the case for enzymes. The preorganization of the enzyme, which constrains the relative geometry of the fragments, is often suggested to play an important role in enzyme reactions.

The independent dynamics assumption in our *ab initio* QM/MM free energy approach would not be appropriate where there is a strong coupling of the chemical reaction dynamics with the protein dynamics. In the present paper, we address this issue. Before we present the idea of our approach, we would like to review the related reaction path Hamiltonian (RPH) of Miller, Handy and Adams for gas phase chemical reactions.[35] The idea of the RPH is to utilize a relatively modest number of electronic structure calculations along the minimum energy reaction path to construct a Hamiltonian for the system based on several assumptions, one being that the reaction valley is harmonic. Once the RPH for a system is constructed, it can be used for chemical reaction dynamics calculations. For a recent review, see Ref. 36.

Following the ideas from the RPH, we here develop a reaction path potential for statistical mechanics and dynamics simulation of chemical reactions in enzymes and other complex systems. As in RPH, our reaction path potential uses the harmonic approximation for the potential energy. But it differs from the RPH in several key aspects and is designed specifically for large systems that are described with *ab initio* QM/MM methods. The reaction path potential deals only with the potential energy and it is an analytical energy expression of the QM/MM potential energy along the minimum energy path. An expansion around the minimum energy path is made in both the nuclear and the electronic degrees of freedom for the QM subsystem, while the energy of the subsystem described with MM remains unchanged from that in the QM/MM expression and the electrostatic interaction between the QM and MM subsystems is described as the interaction of the MM charges with the QM charges that are polarizable in response to the changes in the MM degrees of freedom. The input data for constructing the reaction path potential are energies, vibrational frequencies, and electron density response properties of the QM subsystem along the minimum energy path, all of which can be obtained from QM/MM calculations. Once constructed, it costs much less for its evaluation. Thus, the reaction path potential allows for rigorous statistical mechanics and reaction dynamics calculations of complex systems.

As an example, the method is applied to the statistical mechanical calculations of the potential of mean force for the first step of the chemical reaction in triosephosphate isomerase. This reaction path potential is used to perform the constrained molecular dynamics simulations, from which the thermodynamic force is obtained for each point. During the statistical sampling, the QM subsystem undergoes dynamics coupled with the motions of the MM subsystem. In addition, the QM charges also fluctuate in response to the configuration changes of both QM and MM subsystems. The free energy profile along the minimum energy path is subsequently obtained by the thermodynamic integration (TI) method.

The outline of the paper is as follows: In Sec. II, we present the formulations of the reaction path potential. Next we describe the free energy calculation method with thermodynamic integration. In Sec. III, we present the application of methodology to the proton abstraction reaction catalyzed by TIM. Section IV summarizes the results and concludes the paper.

II. METHOD

A. Background

To set the stage for presentation, we here briefly describe our *ab initio* QM/MM methodology,[23,24,28] a practical approach for modeling the enzymatic reactions.

1. The ab initio QM/MM total energy and its components

The total energy of a QM/MM system can be written as follows:

$$E_{\text{total}} = E_{\text{qm}} + E_{\text{qm/mm}} + E_{\text{mm}}, \tag{1}$$

where E_{qm} is the quantum mechanical energy of the QM subsystem, E_{mm} is the standard molecular mechanical interaction energy involving exclusively atoms in the MM subsystem, and $E_{\text{qm/mm}}$ is the interaction energy between the two subsystems. The QM/MM interaction between the QM subsystem and the MM subsystem can be divided into three terms: the electrostatic energy, the van der Waals energy, and the MM bonded interactions, as in the following equation:

$$E_{\text{qm/mm}} = E_{\text{qm/mm}}^{\text{electrostatics}} + E_{\text{qm/mm}}^{\text{vdw}} + E_{\text{qm/mm}}^{\text{MM-bonded}}, \tag{2}$$

where the MM-bonded interaction $E_{\text{qm/mm}}^{\text{MM-bonded}}$ refers to the MM bond, angle, and dihedral angle energy terms, which involve terms with at least one atom from the MM subsystem and one atom from the QM subsystem. $E_{\text{qm/mm}}^{\text{MM-bonded}}$, $E_{\text{qm/mm}}^{\text{vdw}}$, and E_{mm} are calculated with the MM force field.

In QM calculations, the sum of E_{qm} and $E_{\text{qm/mm}}^{\text{electrostatics}}$ is calculated as the (ground-state) energy of an effective Hamiltonian H_{eff}, namely,

$$E_{\text{qm}} + E_{\text{qm/mm}}^{\text{electrostatics}} = \langle \Psi | H_{\text{eff}} | \Psi \rangle. \tag{3}$$

The effective Hamiltonian describes a total number of electrons N, which is the sum of the electrons of the active part and the valence electrons from all the boundary atoms. These electrons move in the one-electron potential generated from the nuclei of the atoms in the QM subsystem with nuclear charges $\{Z_\alpha\}$, the point charge $\{q_\beta\}$ of the MM atoms, and the effective potential (pseudopotential) of the boundary atoms.[23] The electrostatic interaction energy, $E_{\text{qm/mm}}^{\text{electrostatics}}$, is defined as the interactions between the QM electrons (described by the wave function Ψ) and nuclei with the MM charges; it is expressed as

$$E_{\text{qm/mm}}^{\text{electrostatics}} = \langle \Psi | - \sum_{i}^{N} \sum_{\beta \in \text{MM}} \frac{q_\beta}{r_{\beta i}} | \Psi \rangle + \sum_{\alpha \in \text{QM}, \beta \in \text{MM}} \frac{Z_\alpha q_\beta}{r_{\alpha\beta}}.$$

$E_{\text{qm/mm}}^{\text{electrostatics}}$ can be well approximated by $E_{\text{qm/mm}}^{\text{esp}}$,

$$E_{\text{qm/mm}}^{\text{esp}} = \sum_{\alpha \in \text{QM}, \beta \in \text{MM}} \frac{Q_\alpha q_\beta}{r_{\alpha\beta}}, \tag{4}$$

where Q_α is electrostatic potential (ESP) fitted charge assigned to atom α in the QM subsystem and N is the number of electrons. Thus during the MM subsystem optimization in our iterative optimization approach, we used the approximation[24]

$$E_{\text{qm/mm}}^{\text{esp}} \approx E_{\text{qm/mm}}^{\text{electrostatics}} \tag{5}$$

to eliminate the need of performing QM calculations.

With the *ab initio* QM/MM potential defined in Eq. (1), a minimum energy path can be determined efficiently via the iterative optimization procedure.[24]

2. Calculating the free energy with the free energy perturbation method

Next we calculate the free energy as a function of the reaction coordinate, which is also called the potential of mean force. Consider a reaction coordinate s defined by $s = f(\mathbf{r}_{\text{QM}})$, where $\mathbf{r}_{\text{QM}}$ represents the coordinates of the QM atoms in our QM/MM system. The configuration partition function of the system can be written as

$$Z = \int \exp\{-\beta[E_{\text{qm}}(\mathbf{r}_{\text{QM}}) + E_{qm/mm}(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}}) + E_{\text{mm}}(\mathbf{r}_{\text{MM}})]\} d\mathbf{r}_{\text{QM}} d\mathbf{r}_{\text{MM}} = \int Z(s) ds, \tag{6}$$

where $\beta = 1/k_B T$ is the inverse temperature and k_B is the Boltzmann constant. $\mathbf{r}_{\text{MM}}$ represents the coordinates of MM atoms, $Z(s)$ is the partial partition function associated with the reaction coordinate, and is given by

$$Z(s) = \int \delta(s - f(\mathbf{r}_{\text{QM}})) \times \exp\{-\beta[E_{\text{qm}}(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}}) + E_{qm/mm}(\mathbf{r}_{\text{QM}}, \mathbf{r}_{\text{MM}}) + E_{\text{mm}}(\mathbf{r}_{\text{MM}})]\} d\mathbf{r}_{\text{QM}} d\mathbf{r}_{\text{MM}}, \tag{7}$$

where δ is the Dirac delta function. Associated with this partial partition function is the free energy profile of the reaction

$$F(s) = -\frac{1}{\beta} \ln Z(s). \tag{8}$$

For two neighboring points A and B along the minimum energy path, the difference in free energy is

$$\Delta F^{(A \to B)} = F(s_B) - F(s_A) = -\frac{1}{\beta} \ln \langle \exp\{-\beta[E_{\text{total}}(\mathbf{r}_{\text{QM}}(s_B), \mathbf{r}_{\text{MM}}) - E_{\text{total}}(\mathbf{r}_{\text{QM}}(s_A), \mathbf{r}_{\text{MM}})]\} \rangle_A, \tag{9}$$

where $\langle \ldots \rangle_A$ represents the ensemble average at the reaction coordinate s_A. If we ignore the QM fluctuations by freezing their coordinates to those on the minimum energy path, $\mathbf{r}_{\text{QM}}^{\min}$, then we have

$$\Delta F^{(A \to B)} = \Delta E_1^{(A \to B)} + \Delta F_{\text{mm}}^{(A \to B)} = E_1(\mathbf{r}_{\text{QM}}^{\min}(s_B), \mathbf{r}_{\text{MM}}^{\min}(s_B)) - E_1(\mathbf{r}_{\text{QM}}^{\min}(s_A), \mathbf{r}_{\text{MM}}^{\min}(s_A)) - \frac{1}{\beta} \ln \langle \exp\{-\beta[E_{\text{qm/mm}}(\mathbf{r}_{\text{QM}}^{\min}(s_B), \mathbf{r}_{\text{MM}}) - E_{\text{qm/mm}}(\mathbf{r}_{\text{QM}}^{\min}(s_A), \mathbf{r}_{\text{MM}})]\} \rangle_{\text{MM},A}, \tag{10}$$

where $\langle \ldots \rangle_{\mathrm{MM},A}$ represents ensemble average over the MM subsystem and E_1 is the QM internal energy

$$E_1(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}) = \langle \Psi | H_{\mathrm{eff}} | \Psi \rangle - E^{\mathrm{esp}}_{\mathrm{qm/mm}}(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}). \quad (11)$$

Note that in Eq. (10) we assume the reaction coordinate is only defined by the QM atoms. Equation (10) only includes the effects of MM fluctuations, which is calculated by the free energy perturbation method.

Assuming that the QM atoms fluctuate harmonically and the coordinates of the MM subsystem are frozen, we can obtain the free energy contribution, $F_{\text{qm-harmonic}}$, from the fluctuations of the QM region by performing frequency calculations at points on the minimum energy path. The final free energy change is the sum of the two terms

$$\Delta F^{(A\to B)} = \Delta F^{(A\to B)}_{\mathrm{mm}} + \Delta E^{(A\to B)}_{1} + \Delta F^{(A\to B)}_{\text{qm-harmonic}}. \quad (12)$$

To calculate the free energy of activation based on Eq. (12), we use only the QM energies along the reaction path and the harmonic frequencies at the reactant and transition states from the QM calculations. The statistical averaging is performed with the MM force field. This provides an effective tool for approximating the free energy and has been used in practical applications.[27,28]

We can see that this approach decouples motions of the QM and MM subsystems. In the next subsection, we will show how to construct the reaction path potential, which allows the QM charges to be polarizable and the motion of the QM subsystem to couple with that of the MM subsystem.

B. Reaction path potential

The key idea of the reaction path potential is to approximate the QM charges to the first order, and the QM internal energy up to the second order with respect to the QM and MM fluctuations. First, we decompose the total energy in Eq. (1) into three terms:

$$E_{\mathrm{total}}(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}) = E_1(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}) + E^{\mathrm{esp}}_{\mathrm{qm/mm}}(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}) + E'_{\mathrm{mm}}(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}), \quad (13)$$

where $E_1(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}})$ is the QM internal energy defined by Eq. (11) and $E'_{\mathrm{mm}}(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}})$ is composed of three terms completely expressed with a MM force field; namely,

$$E'_{\mathrm{mm}}(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}) = E^{\mathrm{vdw}}_{\mathrm{qm/mm}}(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}) + E^{\text{MM-bonded}}_{\mathrm{qm/mm}}(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}) + E_{\mathrm{mm}}(\mathbf{r}_{\mathrm{MM}}).$$

We do not put the $E^{\mathrm{esp}}_{\mathrm{qm/mm}}$ term into E'_{mm}, because the charges for QM atoms are not taken from the MM force field. Instead, the QM charges are calculated by the electrostatic fitting method and are allowed to be polarizable. After $E^{\mathrm{esp}}_{\mathrm{qm/mm}}$ is described properly, E_1 is the minimum amount in E_{total} that cannot be described with a conventional MM force field. We then approximate E_1 up to second order in both the nuclear and the electronic degrees of freedom. In the next subsection we show how to calculate $E^{\mathrm{esp}}_{\mathrm{qm/mm}}$. It is followed by the subsection describing the E_1 term.

1. Polarizable QM charge model

To describe the response of a molecule to an external electric field, Stone and Alderton developed the distributed multipole analysis method, in which the molecule is divided into regions, and charges and higher order multipoles are assigned to each region.[37] They then calculated the distributed polarizabilities by solving the coupled-perturbed Hartree–Fock equations.[38] Morita and Kato took a simpler approach, in which they avoided dealing with the higher order multipoles and only calculated the distributed polarizability matrix for ESP point charges.[39] In terms of DFT, the distributed polarizability matrix they obtained is a coarse-grained version of the response kernel based on the atomic site representation. The response kernel matrix is defined as

$$\chi_{\alpha\beta} = \left(\frac{\partial Q_\alpha}{\partial \upsilon_\beta}\right)_N = \frac{\partial^2 E}{\partial \upsilon_\alpha \partial \upsilon_\beta}, \quad (14)$$

where N is the number of electrons, υ_β is the external electrostatic potential, and α,β denote the atomic sites.

Within the QM/MM context, it is straightforward to make the QM charges polarizable once the response kernel is obtained. Let $\mathbf{r}^{\mathrm{min}}_{\mathrm{QM}}$ and $\upsilon^{\mathrm{min}}_{\mathrm{MM}}$ be the QM coordinates and the external electrostatic potential for a reference point on the minimum energy path. Then $\mathbf{r}_{\mathrm{QM}} - \mathbf{r}^{\mathrm{min}}_{\mathrm{QM}}$ reflects the fluctuation of the QM coordinates and $\upsilon_{\mathrm{MM}} - \upsilon^{\mathrm{min}}_{\mathrm{MM}}$ is the fluctuation of the classical electrostatic potential generated by the MM atoms. Let Q^{min} be the ESP charges calculated at the reference point. Then the new charge due to the MM fluctuations at the QM atomic site α is

$$Q_\alpha = Q^{\mathrm{min}}_\alpha + \sum_{\beta\in \mathrm{QM}} \chi_{\alpha\beta}[\upsilon_{\mathrm{MM}}(\mathbf{r}_\beta) - \upsilon^{\mathrm{min}}_{\mathrm{MM}}(\mathbf{r}_\beta)]. \quad (15)$$

To accommodate the fluctuations of the QM atoms, we here define another response kernel $\kappa_{\alpha\beta}$

$$\kappa_{\alpha\beta} = \left(\frac{\partial Q_\alpha}{\partial \mathbf{r}_\beta}\right)_N, \quad (16)$$

which describes the QM charge fluctuation in response to the QM geometry displacements. Using $\kappa_{\alpha\beta}$, the QM charges are

$$Q_\alpha = Q^{\mathrm{min}}_\alpha + \sum_{\beta\in \mathrm{QM}} \kappa_{\alpha\beta}[\mathbf{r}_\beta - \mathbf{r}^{\mathrm{min}}_\beta]. \quad (17)$$

In our work, we evaluate the matrix $\kappa_{\alpha\beta}$ by numerical differentiation. Finally, combining Eq. (15) and Eq. (17), we have

$$Q_\alpha(\mathbf{r},\upsilon_{\mathrm{MM}}) = Q^{\mathrm{min}}_\alpha + \sum_{\beta\in \mathrm{QM}} \chi_{\alpha\beta}[\upsilon_{\mathrm{MM}}(\mathbf{r}^{\mathrm{min}}_\beta) - \upsilon^{\mathrm{min}}_{\mathrm{MM}}(\mathbf{r}^{\mathrm{min}}_\beta)] + \sum_{\beta\in \mathrm{QM}} \kappa_{\alpha\beta}(\mathbf{r}_\beta - \mathbf{r}^{\mathrm{min}}_\beta). \quad (18)$$

This is the first order approximation for the atomic charges which are polarizable with respect to the fluctuations in both the MM and QM subsystems.

When computing the ESP charges, some eigenvalues of the matrix used for ESP fitting might be close to zero. Thus, there will be numerical instability in the ESP charges obtained by inverting this matrix, which results in unreasonably large eigenvalues of the response kernel. This causes instabilities in the molecular dynamics simulations. We solved

this problem by using a modified ESP charge model,[40] in which all the eigenvalues of the matrix to be inverted are augmented by 0.006 a.u.

2. Second order approximation to the QM internal energy E_1

The QM internal energy, E_1 as defined by Eq. (11), is the internal energy of the QM subsystem, whose wave function is polarized by the MM point charges. It depends both on the coordinates of the QM atoms $\mathbf{r}_{\mathrm{QM}}$ and the external electrostatic potential v_{MM} which is exerted on the quantum mechanical region by the MM point charges. Now we focus on a point on the minimum energy path. We expand E_1 around this point on the minimum energy path with respect to $\mathbf{r}_{\mathrm{QM}}$ and v_{MM} up to second order,

$$E_1(\mathbf{r}_{\mathrm{QM}},v_{\mathrm{MM}})\cong E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})+(\mathbf{r}_{\mathrm{QM}}-\mathbf{r}_{\mathrm{QM}}^{\min})^T\cdot\nabla_{\mathbf{r}_{\mathrm{QM}}}E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})+\frac{1}{2}(\mathbf{r}_{\mathrm{QM}}-\mathbf{r}_{\mathrm{QM}}^{\min})^T\cdot\nabla_{\mathbf{r}_{\mathrm{QM}}}^2E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})\cdot(\mathbf{r}_{\mathrm{QM}}-\mathbf{r}_{\mathrm{QM}}^{\min})$$
$$+\int d\mathbf{r}\,\frac{\delta E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})}{\delta v_{\mathrm{MM}}(\mathbf{r})}[v_{\mathrm{MM}}(\mathbf{r})-v_{\mathrm{MM}}^{\min}(\mathbf{r})]+\frac{1}{2}\int d\mathbf{r}\,d\mathbf{r}'[v_{\mathrm{MM}}(\mathbf{r}')-v_{\mathrm{MM}}^{\min}(\mathbf{r}')]$$
$$\times\frac{\delta^2E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})}{\delta v_{\mathrm{MM}}(\mathbf{r}')\,\delta v_{\mathrm{MM}}(\mathbf{r})}[v_{\mathrm{MM}}(\mathbf{r})-v_{\mathrm{MM}}^{\min}(\mathbf{r})]+(\mathbf{r}_{\mathrm{QM}}-\mathbf{r}_{\mathrm{QM}}^{\min})^T\cdot\nabla_{\mathbf{r}_{\mathrm{QM}}}\int d\mathbf{r}\,\frac{\delta E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})}{\delta v_{\mathrm{MM}}(\mathbf{r})}[v_{\mathrm{MM}}(\mathbf{r})-v_{\mathrm{MM}}^{\min}(\mathbf{r})]. \tag{19}$$

The first three terms in Eq. (19) use the same ideas as in the reaction path Hamiltonian.[35] The last three terms in Eq. (19) are introduced in the present work to capture the polarization effect caused by the MM environment. Next, we consider an approximate and discrete representation of the external potential by $v_{\mathrm{MM}}(\mathbf{r}_\alpha)$, the value of the MM electrostatic potential at the QM atomic site $\mathbf{r}_\alpha$; namely, the energy perturbation from the external potential is only due to the changes at each atomic site, $v_{\mathrm{MM}}(\mathbf{r}_\alpha^{\min})-v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})$. Then Eq. (19) becomes

$$E_1(\mathbf{r}_{\mathrm{QM}},v_{\mathrm{MM}})\cong E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})+(\mathbf{r}_{\mathrm{QM}}-\mathbf{r}_{\mathrm{QM}}^{\min})^T\cdot\nabla_{\mathbf{r}_{\mathrm{QM}}}E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})+\frac{1}{2}(\mathbf{r}_{\mathrm{QM}}-\mathbf{r}_{\mathrm{QM}}^{\min})^T\cdot\nabla_{\mathbf{r}_{\mathrm{QM}}}^2E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})\cdot(\mathbf{r}_{\mathrm{QM}}-\mathbf{r}_{\mathrm{QM}}^{\min})$$
$$+\sum_{\alpha\in\mathrm{QM}}\frac{\partial E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})}{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)}[v_{\mathrm{MM}}(\mathbf{r}_\alpha^{\min})-v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})]+\frac{1}{2}\sum_{\alpha\in\mathrm{QM}}\sum_{\beta\in\mathrm{QM}}\frac{\partial^2E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})}{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)\partial v_{\mathrm{MM}}(\mathbf{r}_\beta)}[v_{\mathrm{MM}}(\mathbf{r}_\alpha^{\min})$$
$$-v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})][v_{\mathrm{MM}}(\mathbf{r}_\beta^{\min})-v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})]+\sum_{\beta\in\mathrm{QM}}[\mathbf{r}_{\mathrm{QM}}-\mathbf{r}_{\mathrm{QM}}^{\min}]^T\cdot\nabla_{\mathbf{r}_{\mathrm{QM}}}\frac{\partial E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})}{\partial v_{\mathrm{MM}}(\mathbf{r}_\beta)}[v_{\mathrm{MM}}(\mathbf{r}_\beta^{\min})-v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})]. \tag{20}$$

Now we derive the analytical expression for the first and second energy derivatives involved in Eq. (20). First we consider the first partial derivative of $E_{\mathrm{qm/mm}}^{\mathrm{esp}}$ with respect to v_{MM}. With $E_{\mathrm{qm/mm}}^{\mathrm{esp}}$ defined in Eq. (4), we have

$$\frac{\partial E_{\mathrm{qm/mm}}^{\mathrm{esp}}(\mathbf{r}_{\mathrm{QM}},v_{\mathrm{MM}})}{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)}=\frac{\partial\Sigma_{\beta\in\mathrm{QM}}Q_\beta v_{\mathrm{MM}}(\mathbf{r}_\beta)}{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)}$$
$$=Q_\alpha+\sum_{\beta\in\mathrm{QM}}\frac{\partial Q_\beta}{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)}v_{\mathrm{MM}}(\mathbf{r}_\beta). \tag{21}$$

Using a discrete representation for the change of the external potential, we have[39]

$$\frac{\partial\langle\Psi|H_{\mathrm{eff}}|\Psi\rangle}{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)}\cong Q_\alpha. \tag{22}$$

Combining Eqs. (11), (21), and (22), we have

$$\frac{\partial E_1(\mathbf{r}_{\mathrm{QM}},v_{\mathrm{MM}})}{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)}=-\sum_{\beta\in\mathrm{QM}}\frac{\partial Q_\beta}{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)}v_{\mathrm{MM}}(\mathbf{r}_\beta). \tag{23}$$

The second derivative is

$$\frac{\partial^2E_1(\mathbf{r}_{\mathrm{QM}},v_{\mathrm{MM}})}{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)\partial v_{\mathrm{MM}}(\mathbf{r}_\beta)}=-\frac{\partial Q_\beta}{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)} \tag{24}$$

and the QM coordinate and MM potential mixed second derivative is

$$\frac{\partial^2 E_1(\mathbf{r}_{\mathrm{QM}}, v_{\mathrm{MM}})}{\partial \mathbf{r}_\alpha \partial v_{\mathrm{MM}}(\mathbf{r}_\beta)} = -\frac{\partial Q_\alpha}{\partial v_{\mathrm{MM}}(\mathbf{r}_\beta)} \frac{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)}{\partial \mathbf{r}_\alpha}. \tag{25}$$

We note that in accordance with our charge linear response expansion in Eq. (18), the second derivatives of the ESP charges are ignored in the last two equations. The partial derivative $\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)/\partial \mathbf{r}_\alpha$ has a simple analytical expression because $v_{\mathrm{MM}}(\mathbf{r})$ is defined as

$$v_{\mathrm{MM}}(\mathbf{r}) = \sum_{\gamma \in \mathrm{MM}} \frac{q_\gamma}{|\mathbf{r} - \mathbf{r}_\gamma|}.$$

Using Eqs. (18), (23), (24), and (25) to evaluate the derivatives at the minimum energy point and substituting the results into the last three terms in Eq. (20), we have

$$\sum_{\alpha \in \mathrm{QM}} \frac{\partial E_1(\mathbf{r}_{\mathrm{QM}}^{\min}, v_{\mathrm{MM}}^{\min})}{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)} [v_{\mathrm{MM}}(\mathbf{r}_\alpha^{\min}) - v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})]$$

$$= -\sum_{\alpha \in \mathrm{QM}} \sum_{\beta \in \mathrm{QM}} \chi_{\alpha\beta} v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min}) \times [v_{\mathrm{MM}}(\mathbf{r}_\beta^{\min}) - v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})], \tag{26}$$

$$\frac{1}{2} \sum_{\alpha \in \mathrm{QM}} \sum_{\beta \in \mathrm{QM}} [v_{\mathrm{MM}}(\mathbf{r}_\alpha^{\min}) - v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})] \times \frac{\partial E_1(\mathbf{r}_{\mathrm{QM}}^{\min}, v_{\mathrm{MM}}^{\min})}{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha) \partial v_{\mathrm{MM}}(\mathbf{r}_\beta)} [v_{\mathrm{MM}}(\mathbf{r}_\beta^{\min}) - v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})]$$

$$= -\frac{1}{2} \sum_{\alpha \in \mathrm{QM}} \sum_{\beta \in \mathrm{QM}} [v_{\mathrm{MM}}(\mathbf{r}_\alpha^{\min}) - v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})] \times \chi_{\alpha\beta} [v_{\mathrm{MM}}(\mathbf{r}_\beta^{\min}) - v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})], \tag{27}$$

and

$$\sum_{\alpha \in \mathrm{QM}} \sum_{\beta \in \mathrm{QM}} (\mathbf{r}_\alpha - \mathbf{r}_\alpha^{\min}) \frac{\partial^2 E_1(\mathbf{r}_{\mathrm{QM}}^{\min}, v_{\mathrm{MM}}^{\min})}{\partial \mathbf{r}_\alpha \partial v_{\mathrm{MM}}(\mathbf{r}_\beta)} \times [v_{\mathrm{MM}}(\mathbf{r}_\beta^{\min}) - v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})]$$

$$= -\sum_{\alpha \in \mathrm{QM}} \sum_{\beta \in \mathrm{QM}} (\mathbf{r}_\alpha - \mathbf{r}_\alpha^{\min}) \times \frac{\partial v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})}{\partial \mathbf{r}_\alpha} \chi_{\alpha\beta} [v_{\mathrm{MM}}(\mathbf{r}_\beta^{\min}) - v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})]. \tag{28}$$

Now we focus on the derivatives with respect to the QM atom coordinates. The first derivative of $E_{\mathrm{qm/mm}}^{\mathrm{esp}}$ with respect to $\mathbf{r}_{\mathrm{QM}}$ is

$$\frac{\partial E_{\mathrm{qm/mm}}^{\mathrm{esp}}(\mathbf{r}_{\mathrm{QM}}, v_{\mathrm{MM}})}{\partial \mathbf{r}_\alpha} = \frac{\partial \Sigma_{\beta \in \mathrm{QM}} Q_\beta v_{\mathrm{MM}}(\mathbf{r}_\beta)}{\partial \mathbf{r}_\alpha} = \sum_{\beta \in \mathrm{QM}} \frac{\partial Q_\beta}{\partial \mathbf{r}_\alpha} v_{\mathrm{MM}}(\mathbf{r}_\beta) + Q_\alpha \frac{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)}{\partial \mathbf{r}_\alpha}.$$

The second derivative is

$$\frac{\partial^2 E_{\mathrm{qm/mm}}^{\mathrm{esp}}(\mathbf{r}_{\mathrm{QM}}, v_{\mathrm{MM}})}{\partial \mathbf{r}_\alpha \partial \mathbf{r}_\beta} = \frac{\partial Q_\alpha}{\partial \mathbf{r}_\beta} \frac{\partial v_{\mathrm{MM}}(\mathbf{r}_\alpha)}{\partial \mathbf{r}_\alpha} + \frac{\partial Q_\beta}{\partial \mathbf{r}_\alpha} \frac{\partial v_{\mathrm{MM}}(\mathbf{r}_\beta)}{\partial \mathbf{r}_\beta} + Q_\beta \frac{\partial^2 v_{\mathrm{MM}}(\mathbf{r}_\beta)}{\partial \mathbf{r}_\alpha \partial \mathbf{r}_\beta} \delta_{\alpha\beta},$$

where $\delta_{\alpha\beta}$ is Kronecker delta function. Evaluating the derivatives at the minimum energy point, we have

$$\frac{\partial E_1(\mathbf{r}_{\mathrm{QM}}^{\min}, v_{\mathrm{MM}}^{\min})}{\partial \mathbf{r}_\alpha} = \left. \frac{\partial (\langle \Psi | H_{\mathrm{eff}} | \Psi \rangle - E_{\mathrm{qm/mm}}^{\mathrm{esp}})}{\partial \mathbf{r}_\alpha} \right|_{\mathbf{r}_{\mathrm{QM}}^{\min}, v_{\mathrm{MM}}^{\min}} = \left. \frac{\partial \langle \Psi | H_{\mathrm{eff}} | \Psi \rangle}{\partial \mathbf{r}_\alpha} \right|_{\mathbf{r}_{\mathrm{QM}}^{\min}, v_{\mathrm{MM}}^{\min}} - \sum_{\beta \in \mathrm{QM}} \kappa_{\beta\alpha} v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min}) - Q_\alpha^{\min} \frac{\partial v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})}{\partial \mathbf{r}_\alpha} \tag{29}$$

and

$$\frac{\partial^2 E_1(\mathbf{r}_{\mathrm{QM}}^{\min}, v_{\mathrm{MM}}^{\min})}{\partial \mathbf{r}_\alpha \partial \mathbf{r}_\beta} = \left. \frac{\partial^2 (\langle \Psi | H_{\mathrm{eff}} | \Psi \rangle - E_{\mathrm{qm/mm}}^{\mathrm{esp}})}{\partial \mathbf{r}_\alpha \partial \mathbf{r}_\beta} \right|_{\mathbf{r}_{\mathrm{QM}}^{\min}, v_{\mathrm{MM}}^{\min}} = \left. \frac{\partial^2 \langle \Psi | H_{\mathrm{eff}} | \Psi \rangle}{\partial \mathbf{r}_\alpha \partial \mathbf{r}_\beta} \right|_{\mathbf{r}_{\mathrm{QM}}^{\min}, v_{\mathrm{MM}}^{\min}} - \kappa_{\alpha\beta} \frac{\partial v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})}{\partial \mathbf{r}_\alpha} - \kappa_{\beta\alpha} \frac{\partial v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})}{\partial \mathbf{r}_\beta} - Q_\beta^{\min} \frac{\partial^2 v_{\mathrm{MM}}^{\min}(\mathbf{r}^{\mathbf{min}}{}_\beta)}{\partial \mathbf{r}_\alpha \partial \mathbf{r}_\beta} \delta_{\alpha\beta}, \tag{30}$$

where

$$\frac{\partial \langle \Psi | H_{\mathrm{eff}} | \Psi \rangle}{\partial \mathbf{r}_{\mathrm{QM}}}$$

and

$$\frac{\partial^2 \langle \Psi | H_{\mathrm{eff}} | \Psi \rangle}{\partial \mathbf{r}_{\mathrm{QM}} \partial \mathbf{r}_{\mathrm{QM}'}}$$

are obtained analytically from *ab initio* QM/MM calculations.

3. The reaction path potential

We substitute Eqs. (26)–(30) into Eq. (20) to get the final expression for E_1:

$$E_1(\mathbf{r}_{\mathrm{QM}},v_{\mathrm{MM}})\cong E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})+\sum_{\alpha\in\mathrm{QM}}\left(\left.\frac{\partial\langle\Psi|H_{\mathrm{eff}}|\Psi\rangle}{\partial\mathbf{r}_\alpha}\right|_{\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min}}-\sum_{\beta\in\mathrm{QM}}\kappa_{\beta\alpha}v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})-Q_\alpha^{\min}\frac{\partial v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})}{\partial\mathbf{r}_\alpha}\right)(\mathbf{r}_\alpha-\mathbf{r}_\alpha^{\min})$$
$$+\frac{1}{2}\sum_{\alpha\in\mathrm{QM}}\sum_{\beta\in\mathrm{QM}}(\mathbf{r}_\alpha-\mathbf{r}_\alpha^{\min})\left(\left.\frac{\partial^2\langle\Psi|H_{\mathrm{eff}}|\Psi\rangle}{\partial\mathbf{r}_\alpha\partial\mathbf{r}_\beta}\right|_{\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min}}-\kappa_{\alpha\beta}\frac{\partial v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})}{\partial\mathbf{r}_\alpha}-\kappa_{\beta\alpha}\frac{\partial v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})}{\partial\mathbf{r}_\beta}\right.$$
$$\left.-Q_\beta^{\min}\frac{\partial^2 v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})}{\partial\mathbf{r}_\alpha\partial\mathbf{r}_\beta}\delta_{\alpha\beta}\right)(\mathbf{r}_\beta-\mathbf{r}_\beta^{\min})-\sum_{\alpha\in\mathrm{QM}}\sum_{\beta\in\mathrm{QM}}\chi_{\alpha\beta}v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})[v_{\mathrm{MM}}(\mathbf{r}_\beta^{\min})-v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})]$$
$$-\frac{1}{2}\sum_{\alpha\in\mathrm{QM}}\sum_{\beta\in\mathrm{QM}}[v_{\mathrm{MM}}(\mathbf{r}_\alpha^{\min})-v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})]\chi_{\alpha\beta}[v_{\mathrm{MM}}(\mathbf{r}_\beta^{\min})-v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})]$$
$$-\sum_{\alpha\in\mathrm{QM}}\sum_{\beta\in\mathrm{QM}}(\mathbf{r}_\alpha-\mathbf{r}_\alpha^{\min})\frac{\partial v_{\mathrm{MM}}^{\min}(\mathbf{r}_\alpha^{\min})}{\partial\mathbf{r}_\alpha}\chi_{\alpha\beta}[v_{\mathrm{MM}}(\mathbf{r}_\beta^{\min})-v_{\mathrm{MM}}^{\min}(\mathbf{r}_\beta^{\min})]. \tag{31}$$

With E_1 approximated by Eq. (31) and $E_{\mathrm{qm/mm}}^{\mathrm{esp}}$ calculated with the polarizable ESP charges given by Eq. (18), we finally obtain the reaction path potential (RPP)

$$E_{\mathrm{RPP}}=E_1(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}})+E_{\mathrm{qm/mm}}^{\mathrm{esp}}(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}})+E'_{\mathrm{mm}}(\mathbf{r}_{\mathrm{QM}},\mathbf{r}_{\mathrm{MM}}), \tag{32}$$

which is an approximation to the *ab initio* QM/MM potential. This approximation is valid as long as the deviation from the reference point is small. In the next section, the reaction path potential is used for the activation free energy calculations.

We have so far used the Cartesian coordinate in the expansion of the internal energy. An alternative is to use internal coordinates, which we present in the Appendix.

C. Application of the reaction path potential to potential of mean force calculations

The free energy change $\Delta F^{(A\to B)}$ along the reaction coordinate s, or the so-called potential of mean force (PMF), can be found from the following thermodynamic integration:

$$\Delta F^{(A\to B)}=\int_{s_A}^{s_B}ds'\left\langle\frac{\partial H}{\partial s}\right\rangle_{s'}, \tag{33}$$

where $\langle\ldots\rangle_{s'}$ denotes a canonical ensemble average in the constrained ensemble, i.e., $s=s'$. In practice the above equation is replaced with a numerical integration over discrete steps

$$\Delta F_{\mathrm{TI}}^{(A\to B)}=\sum_{i=A}^{B}\left\langle\frac{\partial H}{\partial s}\right\rangle_{s_i}\Delta s. \tag{34}$$

It has been shown that[41]

$$\left\langle\frac{\partial H}{\partial s}\right\rangle_{s_i}=\langle\lambda_s\rangle_{s_i}, \tag{35}$$

where λ_s is the Lagrange multiplier associated with the constraint $s=s_i$ during the molecular dynamics simulations.

With the reaction path potential introduced in the previous subsection, we can perform the constrained molecule dynamics simulations at each point on the minimum energy path. The constraint force, λ_s, can be solved numerically with the SHAKE algorithm.[42] To confine the sampling to the vicinity of the minimum energy path, an extra restrained potential must be added to the reaction path potential energy surface. Following the approach of Neria *et al.*,[43] we use a simple harmonic form for the restrained potential

$$V_{\mathrm{res}}=k(\mathbf{r}_{\mathrm{QM}}-\mathbf{r}_{\mathrm{QM}}^{\min})^2. \tag{36}$$

To correct the free energy contribution due to the introduction of the above term as well as the metric tensor effect, caused by the overconstraint on the momentum space,[44] we used the following formula:

$$\Delta F_{\mathrm{corr}}^{(A\to B)}=-\frac{1}{\beta}\ln\langle e^{\beta V_{\mathrm{res}}}Z^{-1/2}\rangle_{s_B}+\frac{1}{\beta}\ln\langle e^{\beta V_{\mathrm{res}}}Z^{-1/2}\rangle_{s_A}, \tag{37}$$

where Z is the Fixman determinant of the coordinate transformation,[45]

$$Z=\sum_i\frac{1}{m_i}\left(\frac{\partial s}{\partial x_i}\right)^2. \tag{38}$$

Here x is the Cartesian coordinate and m is the atomic mass. Then the total free energy change or PMF from the classical dynamic simulations is

$$\Delta F_{\mathrm{tot}}^{(A\to B)}=\Delta F_{\mathrm{TI}}^{(A\to B)}+\Delta F_{\mathrm{corr}}^{(A\to B)}. \tag{39}$$

III. CALCULATIONS AND RESULTS

The feasibility of the reaction path potential and its application to reaction free energy calculations are illustrated through a study of the proton transfer step in the reaction catalyzed by the enzyme TIM. Figure 1 shows first step of the reaction catalyzed by TIM, a proton abstraction by Glu165 of the *pro*-R hydrogen at C1 of the substrate dihydroxyacetone phosphate (DHAP). A starting model of the enzyme complex was constructed based on the crystal structure (PDB ID 7tim) and was prepared as previously described.[24] The system was partitioned into QM and MM subsystems and the pseudobond QM/MM approach[23] was

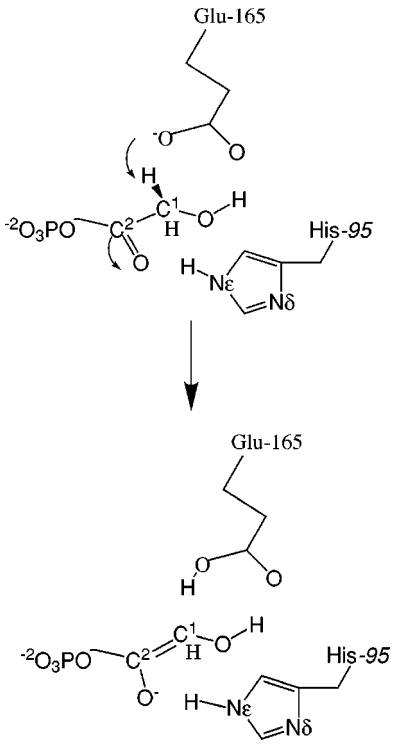

FIG. 1. Initial proton transfer step catalyzed by TIM.

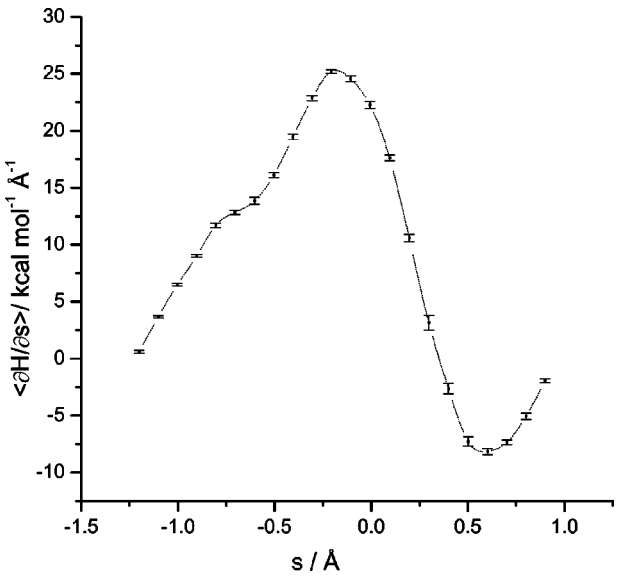

FIG. 2. Constraint force and error bar vs the reaction coordinate.

used for the QM/MM interface. The substrate DHAP and the side chains of Glu165 and His95 (totally 37 atoms) were treated at the HF/3-21G level of theory while the remaining part of the system (6051 atoms, including 752 water molecules) was modeled by the AMBER force field.[46] The calculations were performed by interfacing GAUSSIAN 98[47] with the TINKER program.[48]

We used the coordinate driving method to calculate the reaction path. The reaction coordinate was defined as the difference between the C1–H and O–H bond lengths, i.e., $s=R_{\mathrm{C-H}}-R_{\mathrm{O-H}}$. To drive the system from the reactant to the product, a harmonic potential $\frac{1}{2}k(R_{\mathrm{C-H}}-R_{\mathrm{O-H}}-s)^2$ was added to the system. By varying s from -1.2 to 0.9 Å with a step size of 0.1 Å and optimizing the system with our iterative procedure,[24] 22 points along the minimum energy path were obtained. At each point, the energy, gradient and Hessian, as well as the charge response matrices χ and κ [Eq. (18)], were calculated to construct the reaction path potential. Note that the minimum energy path can be obtained with the nudged elastic band method[49,50] or the recently developed parallel iterative reaction path optimization method[51] for those reactions without clear reaction pathways. In that case, the reaction path potential can still be constructed.

A. Potential of mean force calculations

With the reaction coordinate constrained, molecular dynamics (MD) simulations were performed for each point along the minimum energy path. All the atoms within 15 Å from the C2 atom of DHAP were allowed to move during the simulations. Each of the simulations consisted of 10 ps of equilibration and 50 ps of sampling with a time step 0.5 fs. The cutoff for the nonbonded interactions was 15 Å. The temperature was maintained at 300 K using the weak coupling method,[52] with a coupling time of 0.1 ps. For the restrained potential [Eq. (36)], the force constant k was chosen to be 1.0 kcal mol^{-1} Å^{-2}.

The constraint force (the Lagrange multiplier λ_s solved with the SHAKE algorithm) was collected at each MD step. To reduce the time-correlation effect, the recorded data were block-averaged with an interval of 2 ps. The resulting data were used to make ensemble averages and estimate statistical errors. Figure 2 shows the constraint forces and error bars along the reaction coordinate. We see that the force curve is smooth and as expected, the forces at the reactant, transition state, and product region ($s=-1.2$,0.4,0.9 Å, respectively) are close to zero. The statistical errors are generally small except for the points around the transition state.

The constraint forces were numerically integrated with the trapezoid rule

$$\Delta F_{\mathrm{TI}}^{(i\to i+1)}=\left[\left\langle\frac{\partial H}{\partial s}\right\rangle_{s_i}+\left\langle\frac{\partial H}{\partial s}\right\rangle_{s_{i+1}}\right][s_{i+1}-s_i]/2,$$

where i is the index of the points on the minimum energy path. The quantity z in Eq. (38) in this special case can be simplified as

$$z=\frac{1}{m_{\mathrm{C}}}+\frac{1}{m_{\mathrm{O}}}+\frac{2(1-\cos\phi)}{m_{\mathrm{H}}},$$

where ϕ is the C–H–O angle and m is the atomic mass. The final potential mean force (Fig. 3) was obtained after the corrections from the metric tensor effect as well as the restrained potential. The energy and free energy differences for the stationary points were shown in Table I. To make a comparison, in Fig. 3 we also present the free energy profile calculated with the free energy perturbation method,[24] which uses the frozen QM approximation and fixed QM charges. The comparison indicates the free energy barrier is increased by more than 1 kcal mol^{-1} with the current reaction path potential methodology. Such a difference is caused by including the classical motions of the QM subsystem in our

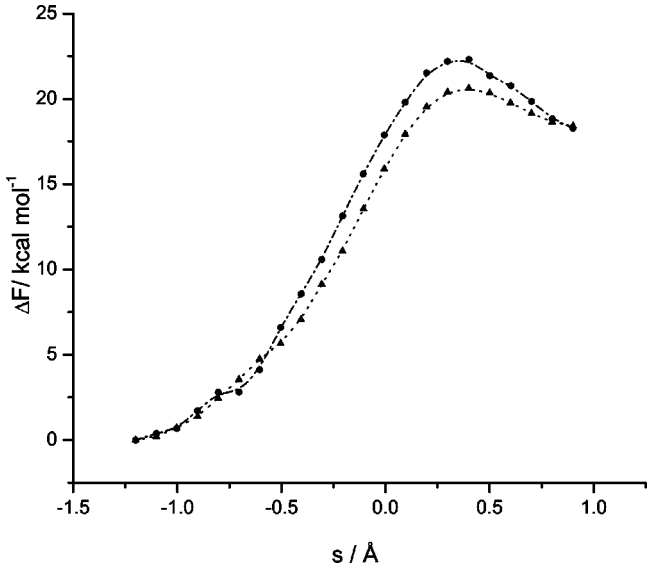

FIG. 3. Free energy changes calculated with the present reaction path potential methodology (circles) and the old methodology (Ref. 24) (triangles).

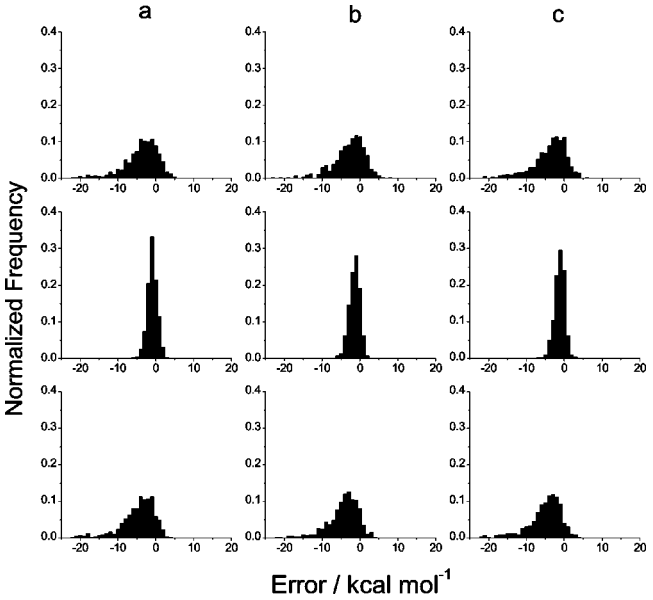

FIG. 4. Error distributions of the QM internal energy (top), QM/MM electrostatic energy (middle), and total energy (bottom). The error is defined as the quantity calculated with the reaction path potential minus the one calculated with *ab initio* QM/MM potential: (a) reactant, (b) transition state, and (c) product.

new approach, and indicates the significance of dynamics in the QM subsystem. We note that compared with experimental results, the calculated reaction barrier is too high because of the level of QM theory (HF/3-21G) used here.

B. Comparison of the reaction path potential to the *ab initio* QM/MM potential

We have shown that combined with constrained MD simulations, the reaction path potential can be used to calculate activation free energies. But how different is the derived reaction path potential from the *ab initio* QM/MM potential? Is the harmonic approximation to the QM internal energy valid? And is the polarizable charge model significantly better than the fixed charge model? Before addressing these questions, we should note that the deviation of the reaction path potential from the *ab initio* QM/MM potential results from two aspects: the harmonic approximation to the QM internal energy E_1 and the linear response approximation to the ESP fitted atomic charges. The latter will cause a difference of the QM/MM electrostatic interaction energies evaluated with the ESP fitted atomic charges and with our polarizable charge model.

We tested the performance of the reaction path potential by using the configurations saved at each 0.1 ps during the potential of mean force calculations. The energies calculated with the *ab initio* QM/MM method and the reaction path potential are compared at all these configurations. The results are displayed in Fig. 4. We see clearly that the total errors of the reaction path potential (using the results of *ab initio* QM/MM calculations as a reference) are dominated by those of the internal energy E_1. The errors in the electrostatic interaction energy are much smaller.

To make further investigation, we diagonalized the mass-weighted Hessian matrices (only for the QM subsystem) for the stationary points at the minimum energy path. Contrary to the gas phase results, the corresponding eigenvalues are all nonzero because the translational and rotational motions are mixed with the vibrational motions. All the eigenvalues of the Hessian in the reactant state are positive, while in the transition state all except one are positive, as we would expect. We then made displacements along each normal mode direction. The displacement we chose corresponds to a thermal energy of 2.38 kcal mol^{-1} ($\sim 4RT$, $T = 300$ K). A total of 222 distorted configurations for 111 normal modes were generated. We calculated the energies at these configurations and obtained the errors of the reaction path potential. The results shown in Fig. 5 indicate that while the errors are small for most of the modes, the motions along a small number of normal modes with low frequencies lead to large errors. Using internal coordinates for the expansion of the QM internal energy reduces the errors, as shown in the Appendix.

The peak error distributions (Fig. 4 middle or Fig. 6 bottom panel) demonstrate that the polarizable charge model reproduced the ESP QM/MM electrostatic energies very well and is significantly better than the fixed charge model (Fig. 6 top panel). It is also apparent from Fig. 6 that to give accurate results, the atomic charges need to be polarized with respect to changes in both the QM geometry and the external potential.

TABLE I. Energy and free energy differences (kcal mol^{-1}) between the stationary points.

	ΔE	$\Delta F_{\mathrm{TI}}^{a}$	ΔF_{corr}	ΔF_{tot}
Transition state	20.55	21.8	0.5	22.3
Product	18.84	18.8	−0.5	18.3

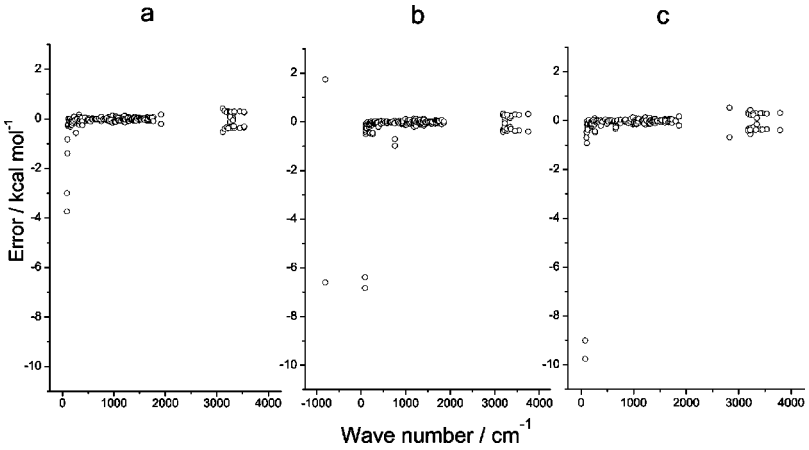

FIG. 5. Reaction path potential errors. Each point corresponds to a displacement from the minimum energy point along a normal mode direction: (a) reactant, (b) transition state, and (c) product.

IV. CONCLUSIONS

We have presented the formulation of the reaction path potential based on information from the minimum energy path derived with *ab initio* QM/MM calculations. The reaction path potential is constructed with a linear response approximation for the QM charges and a second order approximation of the QM internal energy. Combined with the thermodynamic integration method, we have also provided an efficient way to calculate the potential of mean force for enzymatic reactions using the reaction path potential. The application of the methodology was demonstrated with the study of the proton transfer reaction in TIM. In a following publication,[53] we interpolated the reaction path potential along the reaction coordinate to construct a global potential energy surface and calculate the classical reaction transmission coefficient.

Compared with the previous approach,[24] our current treatment allows the QM geometry and atomic charges to fluctuate simultaneously with the MM motions during the statistical sampling. The most important improvement is that the dynamics of the QM and MM subsystems are coupled in the reaction path potential. The results of the free energy profiles calculations indicate that the free energy contributions from the QM fluctuation are significant.

Our tests demonstrate that the accuracy of the reaction path potential depends mainly on the harmonic approximation to the QM internal energy. While most displacements of the QM subsystem can be described very well with the harmonic approximation, several low frequency motions are not harmonic, which can cause large deviations of the reaction path potential from the *ab initio* QM/MM potential. For the QM/MM electrostatic interactions, our tests indicate that our

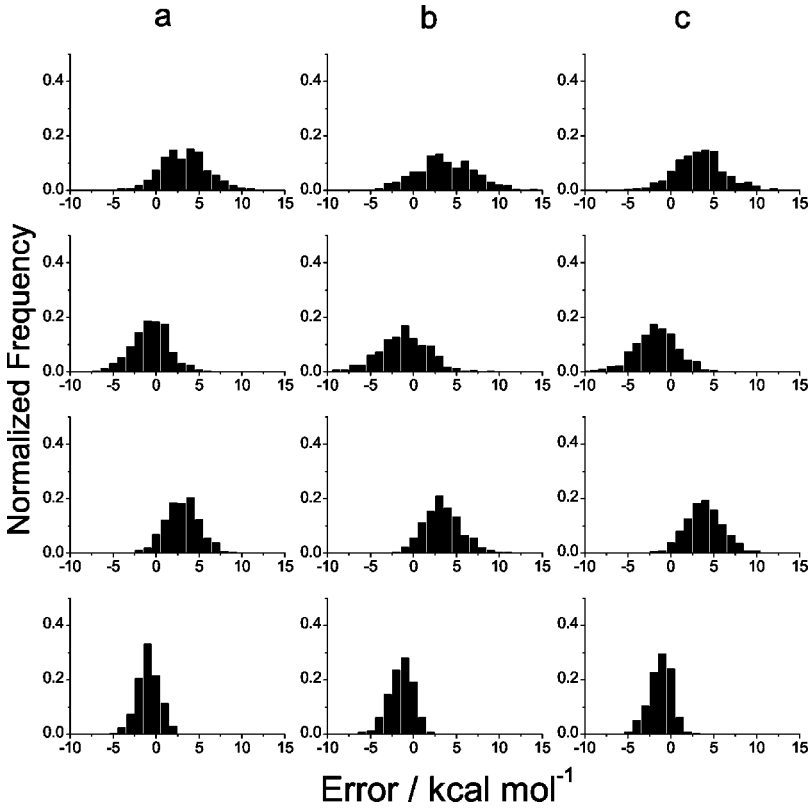

FIG. 6. Error distributions of the QM/MM electrostatic energy. Using the results evaluated with the ESP fitted atomic charges as a reference, the error distributions of different charge models were calculated: (top) fixed charge model; (second row) polarizable charge model with respect to the geometry change only; (third row) polarizable charge model with respect to the change of the external potential only; (bottom) fully polarizable charge model: (a) reactant; (b) transition state; and (c) product.

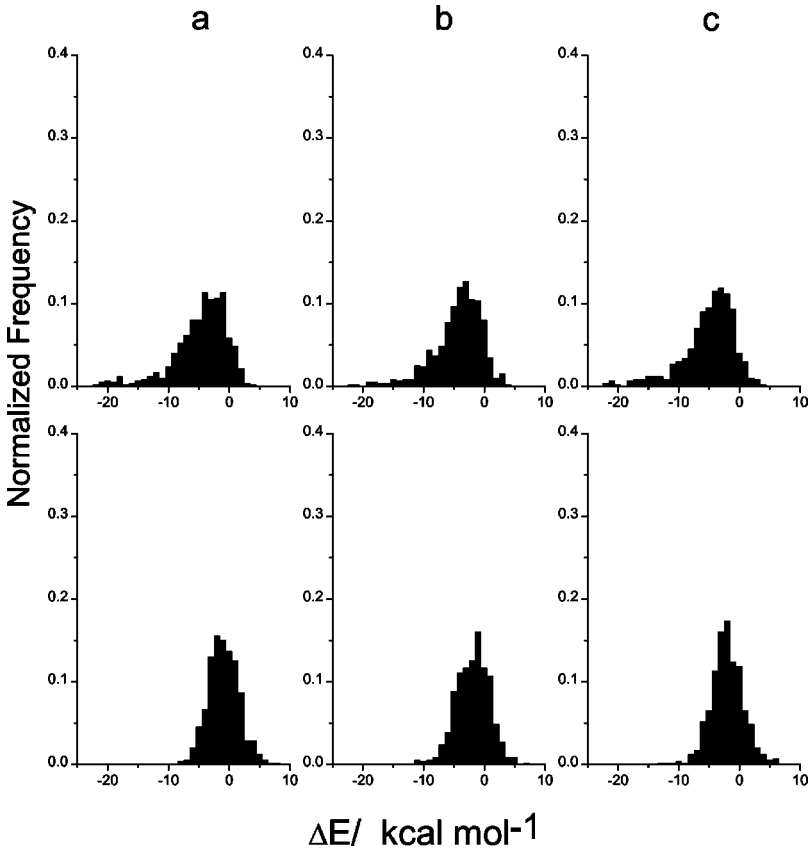

FIG. 7. Error distributions of total energy calculated with E_1 expanded with Cartesian (up) and internal coordinates(bottom): (a) reactant, (b) transition state, and (c) product.

polarizable charge model has made significant improvement over the fixed charge model and inclusion of the charge polarizability with respect to the geometry change is important.

The reaction path potential provides a potential energy surface allowing for rigorous statistical mechanics and reaction dynamics calculations of complex systems. We expect it will be very useful for exploring the coupling of chemical reaction dynamics with protein dynamics.

ACKNOWLEDGMENT

This work has been supported by the National Institute of Health.

APPENDIX: THE EXPANSION OF QM INTERNAL ENERGY AND INTERNAL COORDINATES

In Sec. II B 2, we expanded QM internal energy E_1 with respect to the QM coordinates and MM external potential. Cartesian coordinates were used for the QM subsystem. Here we develop the expansion in a set of internal coordinates. The motivation behind is that the harmonic approximation may be more valid for internal coordinates.

Because all the gradient and Hessian information at the minimum energy point are initially calculated in Cartesian coordinates, we need to transform them to internal coordinates first. Such transformations have been discussed in details.[54,55] Here we present the results relevant for our reaction path potential.

Let us consider a set of redundant internal coordinates $q=\{q,q_2,\ldots,q_n\}$ constructed automatically based on the bonding information[55] of the QM subsystem structure. To describe the translation and rotation of the QM subsystem, we augmented the internal coordinate set by adding the Cartesian coordinates of three nonlinear QM atoms. The infinitely small Cartesian displacements dX are related to dq through the Wilson B matrix ($B_{ij}=\partial q_i/\partial x_j$), i.e.,

$$q=BX. \tag{A1}$$

If we define the G matrix as

$$G=BB^T$$

and G^- as its generalized inverse, then it can be proved[54,55] that

$$g_q=G^-Bg_x \tag{A2}$$

and

$$H_q=G^-B(H_x-B'^Tg_q)B^TG^-, \tag{A3}$$

where g_x and H_x denote the Cartesian gradient and Hessian, g_q and H_q denote the same quantities in the redundant internal coordinates, and $B'_{ijk}=\partial^2 q_i/\partial x_j\partial x_k$.

After performing the above transformations, we can simply express E_1 in the redundant internal coordinates as

$$E_1(\mathbf{r}_{\mathrm{QM}},v_{\mathrm{MM}})=E_1(\mathbf{r}_{\mathrm{QM}}^{\min},v_{\mathrm{MM}}^{\min})+g_q^T\Delta q+\tfrac{1}{2}\Delta q^TH_q\Delta q+E_1^{\mathrm{rest}}, \tag{A4}$$

where Δq is the displacement in internal coordinate and E_1^{rest} is the sum of all the terms involving the partial derivative of v_{mm} in Eq. (31).

To compare with the results of the expansion in Cartesian space, we calculated the total energies for the configurations saved during the potential of mean force calculations

using the formulation of E_1 given by Eq. (A4). From Fig. 7 we see that the expansion of E_1 in internal coordinates does reduce the errors considerably.

[1] T. C. Bruice and K. Kahn, Curr. Opin. Chem. Biol. **4**, 540 (2000).
[2] V. Gogonea, D. Suárez, A. van der Vaart, and J. K. M. Merz, Curr. Opin. Struct. Biol. **11**, 217 (2001).
[3] M. J. Field, J. Comput. Chem. **23**, 48 (2002).
[4] P. A. Kollman, B. Kuhn, and M. Peräkylä, J. Phys. Chem. B **106**, 1537 (2002).
[5] J. Villà and A. Warshel, J. Phys. Chem. B **105**, 7887 (2001).
[6] J. Gao and D. G. Trular, Annu. Rev. Phys. Chem. **53**, 467 (2002).
[7] S. J. Benkovic and S. Hammes-Schiffer, Science **301**, 1196 (2003).
[8] M. Garcia-Viloca, J. Gao, M. Karplus, and D. Truhlar, Science **303**, 186 (2004).
[9] A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes* (Wiley, New York, 1991).
[10] J. Gao, *Review in Computational Chemistry* (VCH, New York, 1995), Vol. 7, pp. 119–185.
[11] R. A. Friesner and M. D. Beach, Curr. Opin. Struct. Biol. **8**, 257 (1998).
[12] K. M. Merz, Jr. and R. V. Stanton, *Encyclopedia of Computational Chemistry* (Wiley New York, 1998), pp. 2330–2343.
[13] Q. Cui, M. Elstner, and M. Karplus, J. Phys. Chem. B **106**, 2721 (2002).
[14] Q. Cui and M. Karplus, J. Am. Chem. Soc. **124**, 3093 (2002).
[15] Q. Cui and M. Karplus, Adv. Protein Chem. **66**, 315 (2003).
[16] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[17] U. C. Singh and P. Kollman, J. Comput. Chem. **7**, 718 (1986).
[18] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).
[19] J. Gao and X. Xia, Science **258**, 631 (1992).
[20] R. V. Stanton, D. S. Hartsough, and K. M. Merz, Jr., J. Phys. Chem. **97**, 11868 (1993).
[21] V. Thery, D. Rinaldi, and J.-L. Rivail, J. Comput. Chem. **15**, 269 (1994).
[22] F. Maseras and K. Morokuma, J. Comput. Chem. **16**, 1170 (1995).
[23] Y. Zhang, T. Lee, and W. Yang, J. Chem. Phys. **110**, 46 (1999).
[24] Y. Zhang, H. Liu, and W. Yang, J. Chem. Phys. **112**, 3483 (2000).
[25] H. Liu, Y. Zhang, and W. Yang, J. Am. Chem. Soc. **122**, 6560 (2000).
[26] Q. Cui and M. Karplus, J. Am. Chem. Soc. **123**, 2284 (2001).
[27] G. A. Cisneros, H. Liu, Y. Zhang, and W. Yang, J. Am. Chem. Soc. **125**, 10384 (2003).
[28] Y. Zhang, H. Liu, and W. Yang, in *Computational Methods for Macromolecules-Challenges and Applications*, Springer Verlag's Lecture Notes Series in Computational Science and Engineering, edited by T. Schlick and H. H. Gan (Springer, New York, 2002), pp. 332–354.
[29] Y. Zhang, K. Jeremy, and J. A. McCammon, J. Phys. Chem. B **107**, 4459 (2003).
[30] Y. Zhang, J. Kua, and J. A. McCammon, J. Am. Chem. Soc. **124**, 10572 (2002).
[31] R. V. Stanton, M. Perakyla, D. Bakowies, and P. A. Kollman, J. Am. Chem. Soc. **120**, 3448 (1998).
[32] J. Chandrasekhar, S. F. Smith, and W. L. Jorgensen, J. Am. Chem. Soc. **107**, 154 (1985).
[33] J. Chandrasekhar and W. L. Jorgensen, J. Am. Chem. Soc. **107**, 2974 (1985).
[34] W. L. Jorgensen, Acc. Chem. Res. **22**, 184 (1989).
[35] W. H. Miller, N. C. Handy, and J. E. Adams, J. Chem. Phys. **72**, 99 (1980).
[36] E. Kraka, in *Encyclopedia of Computational Chemistry*, edited by P. Schleyer (Wiley, New York, 1998), pp. 2437–2463.
[37] A. J. Stone and M. Alderton, Mol. Phys. **56**, 1047 (1985).
[38] A. J. Stone, Mol. Phys. **56**, 1065 (1985).
[39] A. Morita and S. Kato, J. Am. Chem. Soc. **119**, 4021 (1997).
[40] A. Morita and S. Kato, J. Phys. Chem. A **106**, 3909 (2002).
[41] T. Mülders, P. Krüger, W. Swegat, and J. Schlitter, J. Chem. Phys. **104**, 4869 (1996).
[42] J. P. Ryckaert, G. Ciccoti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[43] E. Neria, S. Fischer, and M. Karplus, J. Chem. Phys. **105**, 1902 (1996).
[44] J. Schlitter and M. Klähn, J. Chem. Phys. **118**, 2057 (2003).
[45] M. Fixman, Proc. Natl. Acad. Sci. U.S.A. **71**, 3050 (1974).
[46] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[47] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 98, Revision A.5, Gaussian, Inc., Pittsburgh, PA, 1998.
[48] J. W. Ponder, *TINKER, Software Tools for Molecular Design, Version 3.6*, (February 1998), The most updated version for the TINKER program can be obtained from J. W. Ponder's World Wide Web site ⟨http://dasher.wustl.edu/tinker⟩.
[49] G. Mills and H. Jonsson, Phys. Rev. Lett. **72**, 1124 (1994).
[50] X. Li, H. Liu, and W. Yang, J. Chem. Phys. **120**, 8039 (2004).
[51] H. Liu, Z. Lu, G. A. Cisneros, and W. Yang, J. Chem. Phys. (in press).
[52] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunstereno, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[53] M. Wang, Z. Lu, and W. Yang, J. Chem. Phys. **121**, 101 (2004), following paper.
[54] P. Pulay and G. Fogarasi, J. Chem. Phys. **96**, 2856 (1992).
[55] C. Peng, P. Y. Ayala, H. B. Schlegel, and M. J. Frisch, J. Comput. Chem. **17**, 49 (1996).