Variational calculation of quantum mechanical/molecular mechanical free energy with electronic polarization of solvent

Hiroshi Nakano and Takeshi Yamamoto

Articles you may be interested in

Simple and exact approach to the electronic polarization effect on the solvation free energy: Formulation for quantum-mechanical/ molecular-mechanical system and its applications to aqueous solutions
J. Chem. Phys. **136**, 214503 (2012); 10.1063/1.4722347

Comparison of polarizable continuum model and quantum mechanics/molecular mechanics solute electronic polarization: Study of the optical and magnetic properties of diazines in water
J. Chem. Phys. **135**, 144103 (2011); 10.1063/1.3644894

Combining ab initio quantum mechanics with a dipole-field model to describe acid dissociation reactions in water: First-principles free energy and entropy calculations
J. Chem. Phys. **132**, 074112 (2010); 10.1063/1.3317398

Variational and perturbative formulations of quantum mechanical/molecular mechanical free energy with mean-field embedding and its analytical gradients
J. Chem. Phys. **129**, 244104 (2008); 10.1063/1.3041381

An application of the novel quantum mechanical/molecular mechanical method combined with the theory of energy representation: An ionic dissociation of a water molecule in the supercritical water
J. Chem. Phys. **122**, 044504 (2005); 10.1063/1.1839858

Re-register for Table of Content Alerts

Create a profile.   Sign up today!

Variational calculation of quantum mechanical/molecular mechanical free energy with electronic polarization of solvent

Hiroshi Nakano and Takeshi Yamamoto[a)]
Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan

Quantum mechanical/molecular mechanical (QM/MM) free energy calculation presents a significant challenge due to an excessive number of QM calculations. A useful approach for reducing the computational cost is that based on the mean field approximation to the QM subsystem. Here, we describe such a mean-field QM/MM theory for electronically polarizable systems by starting from the Hartree product ansatz for the total system and invoking a variational principle of free energy. The MM part is then recast to a classical polarizable model by introducing the charge response kernel. Numerical test shows that the potential of mean force (PMF) thus obtained agrees quantitatively with that obtained from a direct QM/MM calculation, indicating the utility of self-consistent mean-field approximation. Next, we apply the obtained method to prototypical reactions in several qualitatively different solvents and make a systematic comparison of polarization effects. The results show that in aqueous solution the PMF does not depend very much on the water models employed, while in nonaqueous solutions the PMF is significantly affected by explicit polarization. For example, the free energy barrier for a phosphoryl dissociation reaction in acetone and cyclohexane is found to increase by more than 10 kcal/mol when switching the solvent model from an empirical to explicitly polarizable one. The reason for this is discussed based on the parametrization of empirical nonpolarizable models. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.3699234]

I. INTRODUCTION

Molecular dynamics (MD) simulation provides a powerful tool for studying complex molecular systems in solution and biological systems. The method of MD simulation ranges widely from purely classical ones based on empirical force fields to purely quantum ones involving all electrons (i.e., *ab initio* MD). The combined quantum mechanical/molecular mechanical (QM/MM) methods[1–3] bridge the gap between those extremes by describing only the chemically relevant part quantum mechanically while the remaining part with classical force fields. Such a QM/MM simulation is typically performed with empirical nonpolarizable MM models, e.g., the TIP3P model[4] for water and OPLS-AA (Ref. 5) for organic solvents. Those models are usually parametrized so as to reproduce bulk properties of solvent and have been employed successfully in numerous studies. However, since electronic polarization is implicit in those models, there is no guarantee that they can also describe electrostatic interactions equally well, e.g., in the vicinity of solvated ions. Our goal in this paper is thus to explore to what extent explicit MM polarization is needed for accurately modeling a system of interest, particularly focusing on chemical reactions in solution (see Ref. 6 for recent reviews on polarizable simulations).

The kinetics of solution-phase reactions can be characterized most fundamentally by free energy of reaction and activation. Several QM/MM calculations have been performed for free energy with explicit polarization included.[7–14] Acevedo and Jorgensen studied the Menshutkin reaction by performing semiempirical QM/MM calculation with a polarizable OPLS model.[11] They observed that for nonpolar aprotic solvents (e.g., cyclohexane), the free energy barrier obtained with nonpolarizable models deviates significantly from experimental results, while inclusion of explicit polarization improves the agreement by more than 10 kcal/mol. Similar improvement has been observed for the binding of charged ligands to solvated proteins[7,15] and pKa shifts of ionizable groups between proteins and aqueous solution.[16,17] In contrast, there are also studies that report essentially no or only minor effects of explicit polarization.[9,12] For example, Lu and Zhang[9] performed *ab initio* QM/MM free energy calculation for the Finkelstein reaction in aqueous solution and observed only a relatively minor change in the free energy barrier (∼1.3 kcal/mol) compared to the nonpolarizable result. The above observations suggest that the effects of explicit polarization are rather dependent on the solute and solvent under study.

On the computational side, the QM/MM free energy calculation presents a significant challenge in itself. This is because free energy calculation involves extensive statistical sampling of the total system, typically requiring a large number of QM calculations greater than 10^5. As such, it is common to employ a fast QM method such as semiempirical Hamiltonian. To employ more accurate *ab initio* methods, it is necessary to introduce some tricks or approximations for reducing the number of QM calculations. One such approach is a variety of mean field approximation for the electrostatic coupling between the QM and MM subsystems. Warshel and co-workers[18] demonstrated that QM/MM free energy calculation can be accelerated significantly by performing a

a)Electronic mail: yamamoto@kuchem.kyoto-u.ac.jp.

136, 134107-1

partial statistical average of the MM subsystem to calculate the QM energy. The average solvent electrostatic potential (ASEP) method developed by Aguilar and co-workers[19–22] takes a complete average of the MM subsystem and performs geometry optimization in solution[19] based on the free energy gradient method.[23–25] The ASEP method is thus analogous to the continuum solvation models[26, 27] and the reference interaction site model (RISM) self-consistent field (SCF) method[28–30] in which the average solvent potential is included into the QM Hamiltonian. It is also noteworthy that the mean field of the environment is used as the zeroth-order part of the QM/MM minimum free energy path method developed by Yang and co-workers.[31–33] In a previous paper,[34] we thus explored a theoretical basis for the mean-field QM/MM method by starting from a variational principle of free energy and presented a simple analytical expression for the free energy gradient.[35] The main advantage here is that it can reduce the number of QM calculations significantly (typically on the order of 100), while allowing an extensive sampling of the MM subsystem. In this sense, the mean-field QM/MM approach can be regarded as intermediate between a direct QM/MM calculation and more traditional approaches (e.g., the continuum and RISM-SCF methods).

The purpose of this paper is two-fold: First, in Sec. II we present a mean-field QM/MM theory for electronically polarizable systems by extending the previous formalism for non-polarizable MM systems.[34] This is performed by first making the Hartree product ansatz for the total system (as often employed in fragment based QM approaches)[36–43] and invoking a variational principle of free energy. We then recast the MM wave function to a classical polarizable model by introducing the charge response kernel.[44, 45] The accuracy of the obtained method is examined by comparison with a direct QM/MM calculation. In Sec. IV, we apply the method to three prototypical reactions in solution (S_N2 reactions of types I and II and a phosphoryl dissociation reaction). We consider six different solvents for this purpose, namely, polar protic solvents (water and methanol), polar aprotic solvents (acetonitrile, acetone, and N, N-dimethylformamide) and a nonpolar aprotic solvent (cyclohexane). Through extensive comparison among the calculated PMF, we demonstrate that the use of explicitly polarizable models is crucial for accurately evaluating the PMF in organic solvents, while the empirical nonpolarizable models (such as TIP3P) (Ref. 4) are rather sufficient for reactions in aqueous solution. We also calculate the PMF with continuum models for comparison and show that the obtained PMF is rather insensitive to the dielectric constant of solvents. The main conclusions obtained are summarized in Sec. V.

II. THEORY

A. QM/MM free energy for electronically polarizable systems

In this section, we define the QM/MM free energy for electronically polarizable systems (without mean field approximation). Our goal here is to calculate the Helmholtz free energy of a system consisting of one solute molecule (with

fixed geometry) immersed in N solvent molecules, that is,

$$A(\mathbf{R}) = -\frac{1}{\beta} \ln \int d\mathbf{r}^N \exp[-\beta E_{\text{tot}}(\mathbf{R}, \mathbf{r})], \quad (1)$$

where $\mathbf{R}$ and $\mathbf{r}^N = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ denote the Cartesian coordinates of solute and solvent molecules, respectively, $E_{\text{tot}}(\mathbf{R}, \mathbf{r}^N)$ is the total energy of the system, and $\beta = 1/(k_B T)$ is reciprocal temperature. The above $A(\mathbf{R})$ provides a free energy surface (or PMF) as a function of solute geometry $\mathbf{R}$. We start from a fragment based description of the total system, in which the solute and solvent molecules are given individual molecular wave functions, Ψ and $\{\psi_i\}$ ($i = 1, \ldots, N$). The molecules are assumed to interact through electrostatic (ES) and van der Waals (vdW) interactions. The total energy in Eq. (1) is defined here as

$$E_{\text{tot}}(\mathbf{R}, \mathbf{r}) = \min_{\Psi} \min_{\{\psi_i\}} E[\Psi, \{\psi_i\}], \quad (2)$$

where the following energy function:

$$E[\Psi, \{\psi_i\}] = \langle \Psi | \hat{H}^0 | \Psi \rangle + \sum_i \langle \psi_i | \hat{h}_i^0 | \psi_i \rangle + \sum_i \mathbf{Q} \cdot \mathbf{D}_i \cdot \mathbf{q}_i$$

$$+ \frac{1}{2} \sum_{i \neq j} \mathbf{q}_i \cdot \mathbf{D}_{ij} \cdot \mathbf{q}_j + U_{\text{vdW}} \quad (3)$$

is minimized with respect to Ψ and $\{\psi_i\}$ for each configuration ($\mathbf{R}, \mathbf{r}^N$). Here, $\hat{H}^0$ and $\hat{h}_i^0$ represent the gas-phase electronic Hamiltonian of the solute and solvent molecules, respectively, and U_{vdW} denotes the vdW interactions in the total system (typically modeled with the Lennard-Jones (LJ) potential). The third and fourth terms in Eq. (3) are the solute-solvent and solvent-solvent ES interactions, where $\mathbf{Q} = \{Q_\gamma\}$ and $\mathbf{q}_i = \{q_{ia}\}$ denote collectively partial charges of the solute and ith solvent molecule, respectively. The $\mathbf{D}_i$ and $\mathbf{D}_{ij}$ matrices in Eq. (3) provide a short-hand notation for the reciprocal distances between solute and solvent atoms given by

$$[\mathbf{D}_i]_{\gamma,a} = \frac{1}{|\mathbf{R}_\gamma - \mathbf{r}_{ia}|}, \quad [\mathbf{D}_{ij}]_{a,b} = \frac{1}{|\mathbf{r}_{ia} - \mathbf{r}_{jb}|}, \quad (4)$$

where index γ runs over the solute atoms and indices a and b run over the ith and jth solvent molecules, respectively. With this notation the third and fourth terms in Eq. (3) read as

$$\mathbf{Q} \cdot \mathbf{D}_i \cdot \mathbf{q}_i = \sum_\gamma \sum_a^{(i)} \frac{Q_\gamma q_{ia}}{|\mathbf{R}_\gamma - \mathbf{r}_{ia}|},$$

$$\mathbf{q}_i \cdot \mathbf{D}_{ij} \cdot \mathbf{q}_j = \sum_a^{(i)} \sum_b^{(j)} \frac{q_{ia} q_{jb}}{|\mathbf{r}_{ia} - \mathbf{r}_{jb}|}. \quad (5)$$

Hereafter, we assume that the partial charges are obtained with the ESP fitting protocol,[46] which allows us to write $\mathbf{Q}$ and $\mathbf{q}_i$ in the form of expectation values[28, 29]

$$\mathbf{Q} = \langle \Psi | \hat{\mathbf{Q}} | \Psi \rangle, \quad \mathbf{q}_i = \langle \psi_i | \hat{\mathbf{q}}_i | \psi_i \rangle, \quad (6)$$

where $\hat{\mathbf{Q}}$ and $\hat{\mathbf{q}}_i$ are one-electron operators that generate ESP derived partial charges. Now inserting the above expressions for $\mathbf{Q}$ and $\mathbf{q}_i$ into Eq. (3) and performing energy minimization

with respect to Ψ and $\{\psi_i\}$, we obtain

$$[\hat{H}^0 + \hat{\mathbf{Q}} \cdot \mathbf{V}]|\Psi\rangle = \mathcal{E}_{\text{QM}}|\Psi\rangle, \qquad (7a)$$

$$\left[\hat{h}_i^0 + \hat{\mathbf{q}}_i \cdot \mathbf{v}_i\right]|\psi_i\rangle = \epsilon_i|\psi_i\rangle, \qquad (7b)$$

where $\mathbf{V}$ and $\mathbf{v}_i$ are ESP values acting on the solute and ith solvent molecule, namely,

$$\mathbf{V} = \sum_i \mathbf{D}_i \mathbf{q}_i, \qquad (8a)$$

$$\mathbf{v}_i = \sum_{j(\neq i)} \mathbf{D}_{ij} \mathbf{q}_j + \mathbf{D}_i^T \mathbf{Q}. \qquad (8b)$$

Note that Eq. (7) represents coupled equations for Ψ and $\{\psi_i\}$, because $\mathbf{V}$ and $\{\mathbf{v}_i\}$ depend on $\mathbf{Q} = \langle\Psi|\hat{\mathbf{Q}}|\Psi\rangle$ and $\mathbf{q}_i = \langle\psi_i|\hat{\mathbf{q}}_i|\psi_i\rangle$ via Eq. (6). By solving the above coupled equations, we obtain the optimal molecular wave functions for each solute and solvent configuration, i.e., $|\Psi(\mathbf{R}, \mathbf{r})\rangle$ and $|\psi_i(\mathbf{R}, \mathbf{r})\rangle$.

Before proceeding, it is interesting to note that the total energy in Eq. (2) can be written as

$$E_{\text{tot}}(\mathbf{R}, \mathbf{r}) = \langle\Psi_{\text{tot}}|\hat{\mathcal{H}}|\Psi_{\text{tot}}\rangle, \qquad (9)$$

where Ψ_{tot} is the Hartree product wave function for the total system

$$|\Psi_{\text{tot}}\rangle = |\Psi(\mathbf{R}, \mathbf{r})\rangle \prod_i |\psi_i(\mathbf{R}, \mathbf{r})\rangle, \qquad (10)$$

and $\hat{\mathcal{H}}$ is the total Hamiltonian given by

$$\hat{\mathcal{H}} = \hat{H}^0 + \sum_i \hat{h}_i^0 + \sum_i \hat{\mathbf{Q}} \cdot \mathbf{D}_i \cdot \hat{\mathbf{q}}_i$$
$$+ \frac{1}{2} \sum_{i \neq j} \hat{\mathbf{q}}_i \cdot \mathbf{D}_{ij} \cdot \hat{\mathbf{q}}_j + U_{\text{vdW}}. \qquad (11)$$

The above Hartree product approximation is commonly employed in a variety of fragment based approaches[43] (e.g., the X-Pol method)[38] for describing electronic polarization in condensed phases.

B. Mean field approximation to QM/MM free energy

The evaluation of $A(\mathbf{R})$ in Eq. (1) is expensive because one needs to calculate the solute wave function $\Psi(\mathbf{R}, \mathbf{r})$ for a large number of solvent configurations $\mathbf{r}$. To avoid this, we introduce mean field approximation for the QM wave function. Specifically, we consider an approximate solute wave function $\tilde{\Psi}$ that depends only on the solute coordinates $\mathbf{R}$, and consider a total wave function given by

$$|\tilde{\Psi}_{\text{tot}}\rangle = |\tilde{\Psi}(\mathbf{R})\rangle \prod_i |\psi_i\rangle. \qquad (12)$$

Here, we suppose that $\tilde{\Psi}(\mathbf{R})$ represents some average or coarse-grained approximation to the original $\Psi(\mathbf{R}, \mathbf{r})$. The corresponding total energy may be defined as

$$\tilde{E}_{\text{tot}}(\mathbf{R}, \mathbf{r}^N; \tilde{\Psi}) \equiv \min_{\{\psi_i\}} \langle\tilde{\Psi}_{\text{tot}}|\hat{\mathcal{H}}|\tilde{\Psi}_{\text{tot}}\rangle = \min_{\{\psi_i\}} E[\tilde{\Psi}, \{\psi_i\}], \qquad (13)$$

by minimizing with respect to $\{\psi_i\}$. One can then define an approximate QM/MM free energy as

$$\tilde{A}(\mathbf{R}; \tilde{\Psi}) = -\frac{1}{\beta} \ln \int d\mathbf{r}^N \exp\{-\beta \tilde{E}_{\text{tot}}(\mathbf{R}, \mathbf{r}; \tilde{\Psi})\}. \qquad (14)$$

To proceed, we utilize the inequality

$$E_{\text{tot}}(\mathbf{R}, \mathbf{r}) \leq \tilde{E}_{\text{tot}}(\mathbf{R}, \mathbf{r}; \tilde{\Psi}), \qquad (15)$$

which holds true for any $\tilde{\Psi}$ because $\tilde{E}_{\text{tot}}(\mathbf{R}, \mathbf{r}; \tilde{\Psi})$ is minimized only for the solvent wave functions, while $E_{\text{tot}}(\mathbf{R}, \mathbf{r})$ is minimized both for the solute and solvent wave functions. The above relation suggests

$$A(\mathbf{R}) \leq \tilde{A}(\mathbf{R}; \tilde{\Psi}), \qquad (16)$$

which states that $\tilde{A}(\mathbf{R}; \tilde{\Psi})$ gives an upper bound on the original $A(\mathbf{R})$ irrespective of the choice of $\tilde{\Psi}$. This means that the best $\tilde{\Psi}$ is obtained by minimizing $\tilde{A}(\mathbf{R}; \tilde{\Psi})$ with respect to $\tilde{\Psi}$. We write the minimal value of $\tilde{A}(\mathbf{R}; \tilde{\Psi})$ thus obtained as $A_{\text{MF}}(\mathbf{R})$, i.e.,

$$A_{\text{MF}}(\mathbf{R}) = \min_{\tilde{\Psi}} \tilde{A}(\mathbf{R}; \tilde{\Psi}). \qquad (17)$$

By performing the variational procedure,[47] we obtain the equation for $\tilde{\Psi}$ as

$$[\hat{H}^0 + \hat{\mathbf{Q}} \cdot \bar{\mathbf{V}}]|\tilde{\Psi}\rangle = \mathcal{E}_{\text{QM}}|\tilde{\Psi}\rangle, \qquad (18)$$

where $\bar{\mathbf{V}} = \{\bar{V}_\alpha\}$ is the statistical average of the solvent ESP,

$$\bar{\mathbf{V}} = \langle\mathbf{V}\rangle = \left\langle \sum_i \mathbf{D}_i \mathbf{q}_i \right\rangle, \qquad (19)$$

with the ensemble average defined by

$$\langle\cdots\rangle = \frac{\int d\mathbf{r}^N \exp[-\beta\mathcal{E}_{\text{MM}}](\cdots)}{\int d\mathbf{r}^N \exp[-\beta\mathcal{E}_{\text{MM}}]}. \qquad (20)$$

The $\mathcal{E}_{\text{MM}}$ in Eq. (20) is the sum of solute-solvent and solvent-solvent interactions in the presence of solute charge $\tilde{\mathbf{Q}}$,

$$\mathcal{E}_{\text{MM}}(\mathbf{R}, \mathbf{r}; \tilde{\mathbf{Q}}) = \sum_i \langle\psi_i|\hat{h}_i^0|\psi_i\rangle + \sum_i \tilde{\mathbf{Q}} \cdot \mathbf{D}_i \cdot \mathbf{q}_i$$
$$+ \frac{1}{2} \sum_{i \neq j} \mathbf{q}_i \cdot \mathbf{D}_{ij} \cdot \mathbf{q}_j + U_{\text{vdW}}, \qquad (21)$$

where $\tilde{\mathbf{Q}}$ and $\mathbf{q}_i$ are, respectively, partial charges obtained from $\tilde{\Psi}$ and ψ_i,

$$\tilde{\mathbf{Q}} = \langle\tilde{\Psi}|\hat{\mathbf{Q}}|\tilde{\Psi}\rangle, \quad \mathbf{q}_i = \langle\psi_i|\hat{\mathbf{q}}_i|\psi_i\rangle. \qquad (22)$$

The solvent wave functions $\{\psi_i\}$ are defined such that they minimize $E[\tilde{\Psi}, \{\psi_i\}]$ in Eq. (13) or equivalently the right-hand side of Eq. (21). The resulting equation for ψ_i is

$$\left[\hat{h}_i^0 + \hat{\mathbf{q}}_i \cdot \mathbf{v}_i\right]|\psi_i\rangle = \epsilon_i|\psi_i\rangle, \qquad (23)$$

where $\mathbf{v}_i$ is the ESP acting on the ith solvent molecule

$$\mathbf{v}_i = \sum_{j(\neq i)} \mathbf{D}_{ij}\mathbf{q}_j + \mathbf{D}_i^T \tilde{\mathbf{Q}}. \qquad (24)$$

Since repeated evaluation of $\{\psi_i\}$ is expensive, we further introduce the charge response kernel (CRK) model for the

solvent.[44,45,48] This model performs second-order expansion of ϵ_i in Eq. (23) with respect to $\mathbf{v}_i$, which gives

$$\mathcal{E}_{\mathrm{MM}}(\mathbf{R}, \mathbf{r}; \tilde{\mathbf{Q}}) \simeq \sum_i \epsilon_i^0 - \sum_i \frac{1}{2} \mathbf{v}_i \cdot \mathbf{K}_i^0 \cdot \mathbf{v}_i + \sum_i \tilde{\mathbf{Q}} \cdot \mathbf{D}_i \cdot \mathbf{q}_i$$

$$+ \frac{1}{2} \sum_{i \neq j} \mathbf{q}_i \cdot \mathbf{D}_{ij} \cdot \mathbf{q}_j + U_{\mathrm{vdW}} \tag{25}$$

(see Appendix A for the derivation). In this model the solvent charge $\mathbf{q}_i$ is obtained by the linear response approximation

$$\mathbf{q}_i \simeq \mathbf{q}_i^0 + \mathbf{K}_i^0 \mathbf{v}_i, \tag{26}$$

where $\mathbf{q}_i^0$ is the solvent charge in the gas phase and $\mathbf{K}_i^0$ is the charge response matrix.[44,45,48] The $\mathbf{q}_i$ and $\mathbf{v}_i$ are obtained by solving Eqs. (24) and (26) for each given $(\mathbf{R}, \mathbf{r}, \tilde{\mathbf{Q}})$. The corresponding total energy is

$$\tilde{E}_{\mathrm{tot}}(\mathbf{R}, \mathbf{r}) = \langle \tilde{\Psi} | \hat{H}^0 | \tilde{\Psi} \rangle + \mathcal{E}_{\mathrm{MM}}(\mathbf{R}, \mathbf{r}; \tilde{\mathbf{Q}}), \tag{27}$$

where $\mathcal{E}_{\mathrm{MM}}$ is given by Eq. (25). Substituting the above total energy into Eq. (14), we obtain the mean field approximation to QM/MM free energy

$$A_{\mathrm{MF}}(\mathbf{R}) = \langle \tilde{\Psi} | \hat{H}^0 | \tilde{\Psi} \rangle + \Delta\mu(\mathbf{R}, \tilde{\mathbf{Q}}), \tag{28}$$

where $\Delta\mu$ is the solvation free energy of the solute with fixed $\mathbf{R}$ and $\tilde{\mathbf{Q}}$, namely,

$$\Delta\mu(\mathbf{R}, \tilde{\mathbf{Q}}) = -\frac{1}{\beta} \ln \int d\mathbf{r}^N \exp\{-\beta \mathcal{E}_{\mathrm{MM}}(\mathbf{R}, \mathbf{r}; \tilde{\mathbf{Q}})\}. \tag{29}$$

Obviously, the above expression for $A_{\mathrm{MF}}(\mathbf{R})$ is quite analogous to those employed in traditional solvation models.[26,27,30] The essential difference here is that we perform explicit sampling of solvent in order to obtain the mean field of solvent, $\langle \mathbf{V} \rangle = \{\langle V_\alpha \rangle\}$. Since the solvation free energy $\Delta\mu$ is calculated explicitly with an atomistic model, the above approach may be regarded as intermediate between a direct QM/MM calculation and more traditional solvation models.[26,27,30]

III. COMPUTATIONAL DETAILS

A. Calculation of the PMF profile

In this section, we describe the details for calculating the $A_{\mathrm{MF}}(\mathbf{R})$ for solution-phase reactions. Specifically, we calculate the profile of $A_{\mathrm{MF}}(\mathbf{R})$ (or PMF) by integrating the free energy gradient $\nabla A_{\mathrm{MF}}(\mathbf{R})$ along a given reaction coordinate, $\xi(\mathbf{R})$. The free energy gradient in the mean field approximation is given by

$$\frac{\partial}{\partial R} A_{\mathrm{MF}}(\mathbf{R}) = \left. \frac{\partial \mathcal{E}_{\mathrm{QM}}(\mathbf{R}, \bar{\mathbf{V}})}{\partial R} \right|_{\bar{\mathbf{V}}}$$

$$+ \left\langle \sum_i \tilde{\mathbf{Q}} \cdot \frac{\partial \mathbf{D}_i}{\partial R} \cdot \mathbf{q}_i + \frac{\partial U_{\mathrm{vdW}}}{\partial R} \right\rangle_{\tilde{\mathbf{Q}}}, \tag{30}$$

where R is an arbitrary element of $\mathbf{R}$ (see Appendix B for the derivation). Using the gradient in Eq. (30), we perform geometry optimization on the free energy surface $A_{\mathrm{MF}}(\mathbf{R})$ with geometric constraint $\xi(\mathbf{R}) = \xi'$. We then obtain the PMF by

integrating the free energy gradient as

$$A_{\mathrm{MF}}(\mathbf{R}(\xi)) = \int_{\xi_0}^{\xi} d\xi' \left. \frac{\partial A_{\mathrm{MF}}(\mathbf{R})}{\partial \mathbf{R}} \right|_{\mathbf{R}=\mathbf{R}(\xi')} \cdot \frac{d\mathbf{R}(\xi')}{d\xi'} + \mathrm{const}, \tag{31}$$

where $\mathbf{R}(\xi')$ denotes the optimized geometry at ξ'. In practice, we discretize the reaction coordinate into a set of grid points $\{\xi_k\}$ and evaluate the integral as follows:

$$A_{\mathrm{MF}}(\mathbf{R}(\xi_K)) - A_{\mathrm{MF}}(\mathbf{R}(\xi_0))$$

$$\simeq \sum_{k=1,K} \frac{1}{2} [\nabla A_{\mathrm{MF}}(\mathbf{R}(\xi_k)) + \nabla A_{\mathrm{MF}}(\mathbf{R}(\xi_{k-1}))]$$

$$\cdot \left[\mathbf{R}(\xi_k) - \mathbf{R}(\xi_{k-1}) \right]. \tag{32}$$

In this paper the grid spacing is set to 0.2 Å, which is found to be sufficiently small for the present application (see Sec. IV).[49]

For the free energy gradient in Eq. (30) to be valid, one needs to ensure that the solute wave function $\tilde{\Psi}$ is determined self-consistently via Eq. (18). This requirement is common to essentially all the solvation models.[26,27,30] In principle, the self-consistency can be achieved by repeating the QM calculation of the solute and statistical sampling of solvent as follows:

1. Prepare an initial guess of the solute wave function $\tilde{\Psi}$.
2. Calculate the solute partial charge $\tilde{Q}_\alpha = \langle \tilde{\Psi} | \hat{Q}_\alpha | \tilde{\Psi} \rangle$ and perform statistical sampling of the solvent in the presence of $\tilde{\mathbf{Q}} = \{\tilde{Q}_\alpha\}$.
3. Calculate the solute wave function $\tilde{\Psi}$ in the presence of the mean solvent ESP, $\langle \mathbf{V} \rangle = \{\langle V_\alpha \rangle\}$.
4. Repeat steps 2 and 3 until $\tilde{\mathbf{Q}}$ and $\langle \mathbf{V} \rangle$ converge.

Once the self-consistency is achieved for given $\mathbf{R}$, one evaluates the free energy gradient by Eq. (30) and advance geometry optimization one step further on the free energy surface $A_{\mathrm{MF}}(\mathbf{R})$. However, this approach is very expensive because of the repeated statistical sampling of solvent for each $\mathbf{R}$. As such, a more efficient approach has been proposed (called the sequential sampling and optimization method).[32] The latter approach is obtained by modifying steps 3 and 4 as follows:

3. Perform full geometry optimization of $\mathbf{R}$ in the presence of an *approximate* average solvent potential obtained from the previous MD run, subject to the geometric constraint $\xi(\mathbf{R}) = \xi'$.
4. Repeat steps 2 and 3 until $\mathbf{R}$, $\tilde{\mathbf{Q}}$, and $\langle \mathbf{V} \rangle$ converge.

This approach has been utilized previously to accelerate the geometry optimization on free energy surface[19,32,34] and it is also utilized in the present paper.

In the application presented in Sec. IV, we choose the reaction coordinate as

$$\xi = r(\mathrm{Cl} - \mathrm{C}) - r(\mathrm{C} - \mathrm{Cl}') \tag{33}$$

for the Finkelstein reaction (Sec. IV A) and

$$\xi = r(\mathrm{Cl} - \mathrm{C}) - r(\mathrm{N} - \mathrm{C}) \tag{34}$$

for the Menshutkin reaction (Sec. IV B). For a phosphoryl dissociation reaction discussed in Sec. IV C, we do not

utilize a distinguished reaction coordinate but rather calculate the reaction path approximately using the string method[50] (see Sec. IV C for details).

The QM calculation was performed using a modified version of the GAMESS program[51] with the ESP charge operator $\hat{Q}_\alpha$ and its derivative implemented.[28,29] The ESP grid points were generated using the Spackman scheme[52] with the number of grid points being ∼3000. We performed coupled-cluster with single, double, and perturbative triple excitations (CCSD(T)) and MP2 calculations[53,54] with the 6-311++G(3*df*,3*p*) basis and density functional theory (DFT) and Hartree-Fock (HF) calculations with the 6-31+G(*d*,*p*) basis, unless otherwise noted. When MP2 method is used for the QM part, we calculated the solute partial charge as

$$Q_\gamma^{MP2} = Z_\gamma - tr[\mathbf{P}^{MP2}\mathbf{b}_\gamma], \quad (35)$$

where Z_γ is nuclear charge, $\mathbf{P}^{MP2}$ is the sum of SCF density and its MP2 correction (known as MP2 response density), and $\mathbf{b}_\gamma$ is the electron part of the ESP charge operator in the atomic orbital basis. We calculated QM/MM free energy at the MP2 level by generalizing Eq. (28) as

$$A_{MF}(\mathbf{R}) = \mathcal{E}_{QM}^{MP2}(\mathbf{R}, \bar{\mathbf{V}}) - \tilde{\mathbf{Q}}^{MP2} \cdot \bar{\mathbf{V}} + \Delta\mu(\mathbf{R}, \tilde{\mathbf{Q}}^{MP2}), \quad (36)$$

where $\mathcal{E}_{QM}^{MP2}$ denotes MP2 energy calculated in the presence of $\bar{\mathbf{V}}$, and $\tilde{\mathbf{Q}}^{MP2}$ is the partial charge obtained with Eq. (35).

B. Protocol of CRK-MD simulation

The MD calculations were performed with the *NVT* condition at 298 K with the box size determined from separate *NPT* runs at 1 atm. The Ewald summation was used for calculating the Coulomb interactions. The system consisted of one solute molecule and 252 solvent molecules for all the systems considered. We set the mass of O and H atoms of water molecules to 10 amu in order to use a time step to 3 fs (Ref. 55). Note that varying the mass of solvent atoms does not affect the thermodynamic properties of interest. We performed the solute-solvent iterations in the mean-field calculation typically 4–6 times until convergence. The iterations were terminated when the residual free energy gradient converged to within 5×10^{-4} (in atomic units). In each iteration, we performed a MD sampling of 600 ps to calculate the mean solvent ESP, followed by QM optimization in the presence of the average solvent field.[34] The above procedure was performed for each grid point of ξ', and the PMF was obtained by integrating the free energy gradient as Eq. (32). Several independent calculations show that statistical error of the free energy barrier is on the order of 0.1 kcal/mol, indicating that the obtained PMF is statistically well converged (see Sec. IV and the supplementary material).[116] The calculation of a single PMF took several hours to several days depending on the type of solvent using 8 nodes of dual-cpu Xeon 2.5 GHz in parallel. For the reactions studied below, most of computer time was spent on the classical MD sampling of solvent rather than the QM calculation of the solute molecule. All the MD calculations were performed with a modified version of the DLPOLY program.[56]

The solvent was described both with polarizable and nonpolarizable models. In the nonpolarizable case, we used the TIP3P model for water,[4] the OPLS-AA model[5] for methanol, acetone, and *N,N*-dimethylformamide (DMF), the Böhm model[57] for acetonitrile, and the OPLS-UA model for cyclohexane.[58] In the polarizable case, we used the CRK model[44,45] obtained by attaching the ESP derived charges $\mathbf{q}_i^0$ and the CRK matrix $\mathbf{K}_i^0$ on each solvent molecule. The $\mathbf{q}_i^0$ was calculated at the B3LYP/aug-cc-pVTZ level, and the $\mathbf{K}_i^0$ was obtained by numerically differentiating the partial charge with respect to the external ESP values [cf. Eq. (A3)]. The geometries and LJ parameters of solvent were taken from the corresponding nonpolarizable models. In all the calculations, the solvent molecules were treated as rigid bodies. Regarding the solute molecule, the LJ parameters for the Finkelstein and Menshutkin reactions were taken from the AMBER94 force field,[59] except for the Cl parameter taken from Gao and Xia,[60] whereas the LJ parameters for the phosphoryl dissociation reaction were taken from the OPLS-AA model.[5] The validity of the present CRK model is further discussed in Appendix A.

In the CRK-MD calculation, all the electrostatic interactions (including the solute-solvent ones) were attenuated at short distances as follows:[45,61]

$$[\mathbf{D}_{i,j}]_{a,b} = \frac{1}{|\mathbf{R}_{ia} - \mathbf{R}_{jb}|} f\left(\frac{|\mathbf{R}_{ia} - \mathbf{R}_{jb}|}{s_{ia,jb}}\right), \quad (37)$$

where $f(v)$ is a Thole-type damping function[61]

$$f(v) = \begin{cases} v^4 - 2v^3 + 2v, & v \le 1 \\ 1, & v > 1, \end{cases} \quad (38)$$

and the $s_{ia,jb}$ in Eq. (37) is a characteristic distance given by

$$s_{ia,jb} = A(\alpha_{ia}\alpha_{jb})^{1/6}, \quad (39)$$

where α is the atomic polarizability and A is an adjustable parameter[45] set to 2.7 for water and 2.6 for organic solvents. The above damping function accounts for the charge distribution of individual atoms and is necessary for avoiding the so-called "polarization catastrophe." The values of α (in Å^3) are chosen as follows:[61–63] $\alpha_C = 1.405$, $\alpha_H = 0.514$, $\alpha_N = 1.105$, $\alpha_O = 0.862$, $\alpha_P = 3.63$, $\alpha_{Cl^-} = 3.25$, $\alpha_{CH_2} = 1.405$ for the united CH_2 atom of cyclohexane. In nonpolarizable simulations, all the Coulomb interactions were treated without attenuation [i.e., $f(v) \equiv 1$]. Other details of the CRK-MD (including the extrapolation of partial charges) are identical to those described in Ref. 48.

IV. RESULTS AND DISCUSSION

A. Finkelstein reaction

The first reaction that we study is the Type-I S_N2 (Finkelstein) reaction

$$Cl^- + CH_3Cl \longrightarrow ClCH_3 + Cl^-, \quad (40)$$

which exhibits charge displacement from the attacking chloride ion to the leaving one as the reaction proceeds. Since the reactants have a more localized charge distribution than the transition state, a polar solvent stabilizes the former more
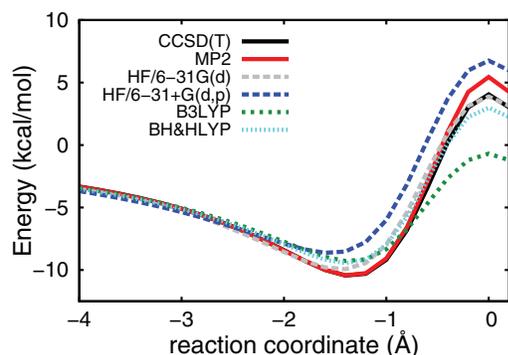
FIG. 1. Potential energy profile of the Finkelstein reaction ($Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$) in the gas phase. Zero of energy is set at the infinitely separated reactants. The basis sets are 6-311++G(3df,3p) for CCSD(T) and MP2 and 6-31+G(d,p) for B3LYP, BHHLYP, and HF methods. For comparison, the energy profile at the HF/6-31G(d) level is also shown.

FIG. 2. Potential of mean force (PMF) of the Finkelstein reaction $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$ in aqueous solution. The result obtained with the mean-field QM/MM method is compared to those obtained from direct QM/MM calculations (see text for details). The QM calculation is performed at the HF/6-31G(d) level. The water is represented by the TIP3P model. The circles depict raw data points obtained by integrating the free energy gradient in Eq. (30). The statistical error of the PMF is comparable to the width of the plotted curve (see the supplementary material[116] for more details).

strongly than the latter, making the free energy barrier higher in polar solvents. Indeed, it is known that the Finkelstein reaction proceeds by orders of magnitude slower in polar solvents than in the gas phase.[64] Because of this, the Finkelstein reaction has been studied extensively using a variety of theoretical methods.[9,32,65–74]

We first present in Fig. 1 the potential energy profile in the gas phase calculated at various QM levels. The energy profile was calculated by optimizing all the internal coordinates other than the reaction coordinate. The CCSD(T), MP2, and Becke's half-and-half Lee-Yang-Parr (BHHLYP) methods provide similar results with a complexation energy of 11 kcal/mol and a barrier height of 3–4 kcal/mol (as measured from the infinitely separated reactants), while the B3LYP method somewhat underestimates the reaction barrier. Interestingly, the HF/6-31G(d) method gives an energy profile close to the CCSD(T) result due essentially to error cancellation. This fact was utilized in the earlier work of Jorgensen and co-workers[65,66] in order to perform accurate *ab initio* calculation of PMF with minimal computational efforts.

Figure 2 displays the PMF in aqueous solution obtained from the mean-field QM/MM calculation. Here, the QM level is HF/6-31G(d) and the solvent is described with the TIP3P model. The PMF exhibits a free energy barrier of 25 kcal/mol, which is significantly higher than the potential energy barrier in the gas phase (3–4 kcal/mol). This difference arises from a large energy cost for desolvating the reactants (particularly the chloride ion) upon formation of the transition state. We note that the PMF thus obtained is statistically well converged, with the statistical error being comparable to the width of the plotted curve. We also note that the PMF is well converged with respect to the grid spacing in Eq. (32); see the supplementary material[116] for more details.

To check the accuracy of the mean field approximation, we performed a direct QM/MM calculation without mean field approximation. Specifically, we calculated the free energy gradient as

$$\frac{\partial A(\mathbf{R})}{\partial \mathbf{R}} = \left\langle \frac{\partial E_{tot}(\mathbf{R}, \mathbf{r})}{\partial \mathbf{R}} \right\rangle, \quad (41)$$

by running a direct QM/MM trajectory for 90 ps. Here, the solute geometry was fixed at one of the optimized
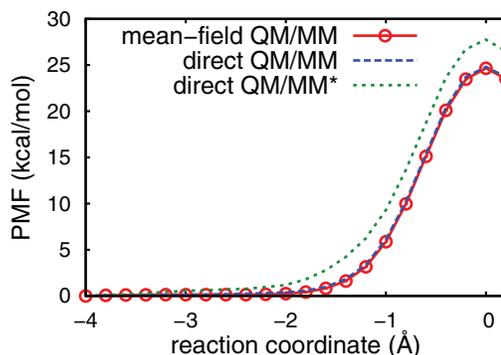
geometries $\{\mathbf{R}(\xi_k')\}$ obtained from the mean-field calculation. The free energy gradient was integrated to obtain the PMF in Fig. 2 (labeled as "direct QM/MM"). Note that the integration involves 22 grid points of ξ_k, thus performing a total of ~2000 ps QM/MM run to obtain a single PMF. Remarkably, the PMF thus obtained agrees quite well with that obtained from the mean-field QM/MM calculation, indicating that the self-consistent mean-field approximation works very well for the present system. This in turn suggests that statistical fluctuations of the QM wave function (about the self-consistent state) is of minor importance once the average distortion of the QM wave function is properly taken into account in the free energy calculation.

The above direct QM/MM calculation is not conventional in that the solute geometry is fixed in space during the trajectory calculation. A more standard approach is to move all the degrees of freedom other than the reaction coordinate $\xi(\mathbf{R})$. The corresponding PMF is given by

$$G(\xi') = -\frac{1}{\beta} \ln \int d\mathbf{R} \int d\mathbf{r}^N$$
$$\times \exp[-\beta E_{tot}(\mathbf{R}, \mathbf{r})]\delta[\xi(\mathbf{R}, \mathbf{r}) - \xi']. \quad (42)$$

For comparison, we also calculated the above $G(\xi')$ by performing a series of umbrella sampling calculations followed by the weighted histogram analysis. Here, the umbrella potential was defined for each grid point of $\{\xi_k'\}$ (with an equal spacing of 0.2 Å) and with the harmonic frequency of 30, 50, 80 kcal/mol/Å² for $\xi_k' \in [-4.0, -2.6]$, $[-2.4, -1.0]$, and $[-0.8, 0.2]$, respectively. For each umbrella potential, we performed a direct QM/MM sampling for 60 ps at the HF/6-31G(d) level, thus running a total of ~1300 ps QM/MM calculation to obtain a single PMF. The PMF thus obtained is shown in Fig. 2 (labeled as "direct QM/MM*"), whose barrier is higher by 3–4 kcal/mol than that of $A_{MF}(\mathbf{R})$. This difference arises from solute thermal/entropic contribution, which needs to be taken into account when making a quantitative comparison with experiment (see below).
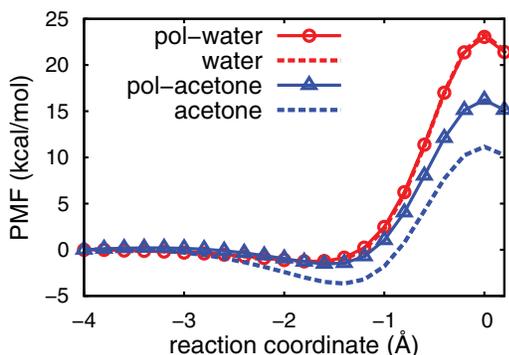
FIG. 3. PMF of the Finkelstein reaction $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$ in water and acetone solutions calculated with the mean-field QM/MM method. Solvents are described with the TIP3P and OPLS models in the nonpolarizable case (dashed lines) and with the CRK model in the polarizable case (solid line). The QM calculation is performed at the BHHLYP/6-31+G(d,p) level.

Figure 3 displays the profile of $A_{MF}(\mathbf{R})$ in aqueous and acetone solutions calculated at the BHHLYP level. In aqueous solution, the PMF obtained with the polarizable model is almost identical to that obtained with the TIP3P model, resulting in a barrier height of 23 kcal/mol. Indeed, the PMF in aqueous solution hardly depends on the type of water models employed, as seen from comparison among several different models (see the supplementary material).[116] In acetone solution, on the other hand, the free energy barrier calculated with the polarizable model (16 kcal/mol) is considerably higher than that obtained with the OPLS model (11 kcal/mol). To

obtain more insight, we depict in Fig. 4 the profile of average solvent ESP $\bar{V}$ acting on the solute molecule. In aqueous solution, the ESP profiles obtained with the polarizable and nonpolarizable models are almost identical, while in acetone solution the ESP profiles are rather different between the two models. Figure 4(c) displays the ES contribution to the solvation free energy calculated with the linear response approximation, namely, $1/2\tilde{Q} \cdot (\bar{V} - V_0)$, where $\mathbf{V}_0$ is the Wigner potential.[75–78] This figure shows that in acetone solution, the solvation free energy decreases by -15 kcal/mol for the reactant and by -8 kcal/mol for the transition state, resulting in the increased free energy barrier shown in Fig. 3. On the other hand, the solvation free energies in aqueous solution calculated with the polarizable and nonpolarizable models are almost identical with a slight overall shift of ~2 kcal/mol.

Figure 5(a) displays the profile of $A_{MF}(\mathbf{R})$ calculated for various solvents with the polarizable models. As seen, the PMFs exhibit different barrier heights depending on the type of solvents. Polar protic solvents (water and methanol) exhibit the highest barrier of 23–24 kcal/mol due to the presence of solute-solvent hydrogen bonds. Polar aprotic solvents (DMF, acetonitrile, and acetone) exhibit an intermediate barrier height (16–18 kcal/mol), while a nonpolar aprotic solvent (cyclohexane) gives a low barrier of 9 kcal/mol. The barrier heights obtained with different levels of QM theory are summarized in Table I. For comparison, we also calculated the PMF using the COSMO continuum model.[79] The obtained PMFs are plotted in Fig. 5(b), which shows that the COSMO model provides a reasonable estimate of the free energy barrier in aqueous solution (21 kcal/mol). However,
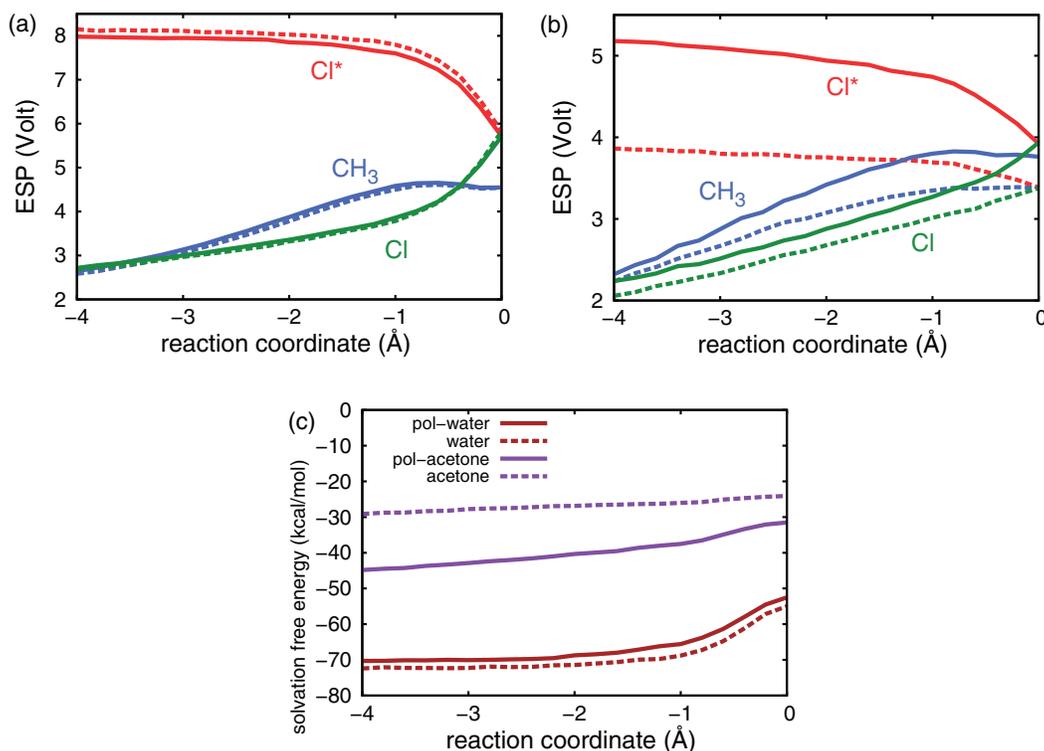

FIG. 4. Solvent electrostatic potentials (ESPs) acting on the solute atoms, $\langle V_\alpha \rangle$, for the Finkelstein reaction in (a) water and (b) acetone solutions. The results obtained with the polarizable and nonpolarizable solvent models are shown by solid and dashed lines, respectively. Cl* and Cl indicate the attacking and leaving chloride atoms, respectively. The curve labeled with CH_3 shows the mean value of ESP acting on the methyl group. Panel (c) displays the solvation free energy calculated with the linear response approximation (see the main text). The ESP values in panels (a) and (b) include the contribution of the Wigner potential.
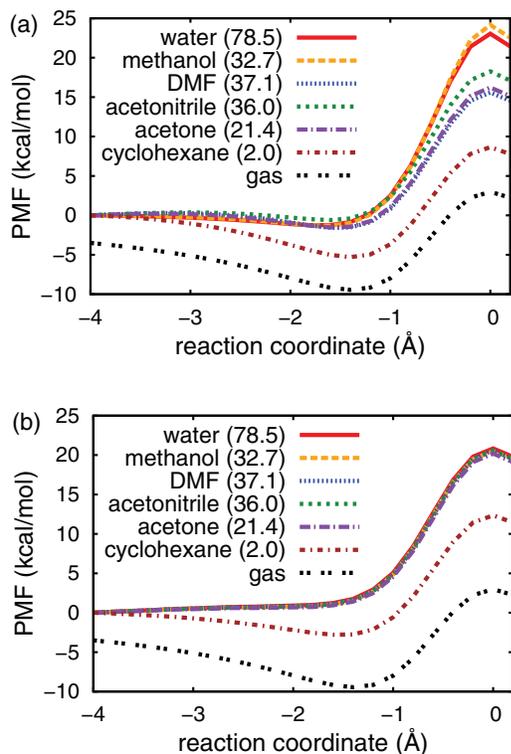
FIG. 5. PMF of the Finkelstein reaction in various solvents calculated with (a) the mean-field QM/MM method and (b) the COSMO continuum solvation model. In panel (a), all the solvents are described with the polarizable CRK model. The QM calculation is performed at the BHHLYP/6-31G+(d,p) level.

it is also seen that the PMF is rather insensitive to the value of dielectric constant, and as a result the COSMO model cannot distinguish solvents other than cyclohexane. A similar trend is observed for the PCM model (see the supplementary material)[116] and also observed previously for different reactions.[80,81] The above trend may be attributed to the inadequate description of solute-solvent specific interactions including hydrogen bonds. Another possible reason is that the apparent surface charge used in the continuum model is proportional to $f(\varepsilon) = (\varepsilon - 1)/(\varepsilon + 1/2)$, which is a rather flat function for $\varepsilon > 20$.

To compare the calculated value of $\Delta A^{\ddagger}$ with experiment, we need to make two further corrections for $\Delta A^{\ddagger}$. The first is the correction for solute thermal motions, which we estimated using the standard separable rotation/vibration approximation.[82,83] The necessary input for this is the Hessian matrix of $A_{MF}(\mathbf{R})$ at stationary points, which was obtained by

TABLE I. Free energy barrier $\Delta A^{\ddagger}$ (in kcal/mol) for the Finkelstein reaction (Cl$^-$ + CH$_3$Cl $\rightarrow$ ClCH$_3$ + Cl$^-$) obtained with the mean-field QM/MM method. $\Delta A^{\ddagger}$ is calculated as $A(\xi = 0.0) - A(\xi = -4.0)$. The solvents are described with the polarizable CRK model. Values in parentheses are obtained with the nonpolarizable solvent models. The basis sets are 6-311++G(3df,3p) for MP2 and 6-31+G(d,p) for BHHLYP method.

QM method	Water	Methanol	Acetone	DMF
MP2	25.6 (26.0)	25.8 (23.6)	17.1 (11.6)	16.3 (11.7)
BHHLYP	23.0 (23.3)	24.2 (21.9)	16.2 (11.2)	15.6 (11.1)
HF/6-31G(d)	24.2 (24.6)	25.3 (23.3)	17.5 (12.3)	17.0 (12.5)

TABLE II. Free energy correction for statistical fluctuations of the QM wave function [namely, the second term in Eq. (C1)] calculated for the Finkelstein reaction in solution (in kcal/mol). The QM level is BHHLYP/6-31+G(d,p) and the solvents are described with the polarizable CRK model. Values in parentheses are obtained with the nonpolarizable solvent models.

ξ (Å)	Water	Methanol	Acetone	DMF
−4.0	−0.15 (−0.36)	−0.12 (−0.34)	−0.17 (−0.29)	−0.22 (−0.36)
0.0	−0.34 (−0.97)	−0.30 (−0.95)	−0.11 (−0.16)	−0.09 (−0.14)

finite difference of the free energy gradient in Eq. (30). The solute thermal correction thus obtained with the standard state of 1 M is 4.0 kcal/mol in aqueous solution and 4.6 kcal/mol in the gas phase. Since those values are rather close, we used the average value (4.3 kcal/mol) for all the solvents considered. The second correction for $\Delta A^{\ddagger}$ is on the statistical fluctuations of QM wave function about the self-consistent state (see Appendix C). The latter correction was estimated using the CRK of the solute molecule and is provided in Table II. The result shows that the correction for $\Delta A^{\ddagger}$ is on the order of 0.1 kcal/mol and is thus rather minor compared to the solute thermal correction.

Table III summarizes the activation free energy $\Delta G^{\ddagger}$ thus obtained by adding the above two corrections for $\Delta A^{\ddagger}$ (hereafter we denote the corrected free energy as $\Delta G^{\ddagger}$). This table shows that in nonaqueous solutions the agreement with experiment[84,85] is significantly improved by using the polarizable model. For example, $\Delta G^{\ddagger}$ in acetone solution is calculated to be 20.6 and 15.6 kcal/mol with the polarizable and nonpolarizable models, respectively, where the former is in better agreement with experiment (21.8 kcal/mol). On the other hand, $\Delta G^{\ddagger}$ in aqueous solution obtained with the CRK and TIP3P models (27.1 and 27.0 kcal/mol) are both in excellent agreement with experiment (26.6 kcal/mol). The substantial improvement observed for nonaqueous solutions is partly attributed to the relatively large polarizability of solvent (see Table IV). For example, the molecular polarizability of acetone (6.4 Å^3) is considerably larger than that of water (1.5 Å^3), implying much stronger electronic polarization in acetone solution. This expectation is consistent with the profile of solvent ESP and solvation free energy shown in Fig. 4. Another likely reason for the different behavior of water and organic solvents is the parametrization of empirical solvent models; this point will be discussed in Sec. IV B.

TABLE III. Activation free energy $\Delta G^{\ddagger}$ (in kcal/mol) for the Finkelstein reaction Cl$^-$ + CH$_3$Cl $\rightarrow$ ClCH$_3$ + Cl$^-$ calculated with the mean-field QM/MM method. The QM calculation is performed at the BHHLYP/6-31+G(d,p) level, and the solvents are described with the polarizable CRK model. The values include the corrections for solute thermal motions (4.3 kcal/mol, see the main text) and statistical fluctuations of the QM wave function (Table II). Values in parentheses are the results obtained with the nonpolarizable models. $\Delta G^{\ddagger}_{expt}$ refers to the experimental estimate.[85]

Solvent	Water	Methanol	Acetone	DMF
$\Delta G^{\ddagger}$	27.1 (27.0)	28.3 (25.6)	20.6 (15.6)	20.0 (15.6)
$\Delta G^{\ddagger}_{expt}$	26.6	28.2	21.8	22.7

TABLE IV. Dielectric constants and molecular polarizability of water, methanol (MeOH), acetonitrile (MeCN), acetone, N,N-dimethylformamide (DMF), and cyclohexane (CHX).

	Water	MeOH	MeCN	Acetone	DMF	CHX
ε	78.5	32.7	36.0	21.4	37.1	2.0
α (Å^3)	1.5	3.3	4.4	6.4	11.5	10.9

B. Menshutkin reaction

We next study the Type-II S_N2 (Menshutkin) reaction

$$NH_3 + CH_3Cl \longrightarrow NH_3CH_3^+ + Cl^-. \qquad (43)$$
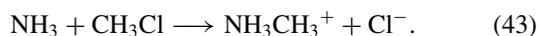
This reaction is known to exhibit significantly enhanced rates in polar solvents due to strong electrostatic stabilization of the products.[64] This is in contrast to the Finkelstein reaction discussed in Sec. IV A, which are decelerated by the electrostatic stabilization of the reactants. Because of this, a number of theoretical studies[25,34,60,86–97] have been performed for the model reaction in Eq. (43). Those studies suggest the free energy of activation in aqueous solution to be around 20–30 kcal/mol and the free energy of reaction to be −20 to −35 kcal/mol (including solute thermal correction),[98] although the results depend significantly on the detail of calculation. To our knowledge, no experimental result is available for the activation free energy, while the free energy of reaction in aqueous solution is known to be −34 ± 10 kcal/mol.[60]

We first present the potential energy profile in the gas phase at various QM levels (Fig. 6). The CCSD(T) and MP2 results agree well with each other, and thus they can be utilized as the reference. In comparison, the DFT methods underestimate the reaction barrier by 5–10 kcal/mol, and the HF method underestimates the potential energy at $\xi = 2.0$ Å by 10 kcal/mol. Since the BHHLYP method gives a well-balanced description of the entire profile, we will utilize the latter for the QM/MM calculation. The use of the BHHLYP method is also motivated by the previous studies that employ the same functional.[34,90,94] It should be noted, however, that the free energy barrier obtained with the CCSD(T) or MP2 theory would be greater by ∼5 kcal/mol than that obtained with the BHHLYP method.
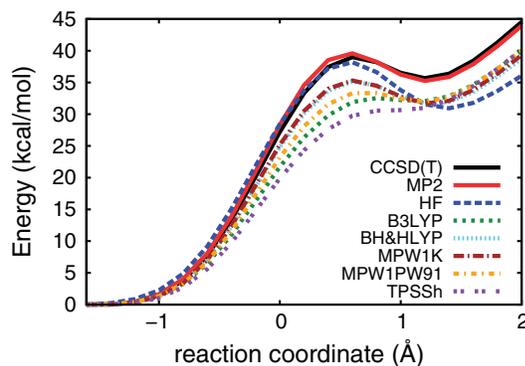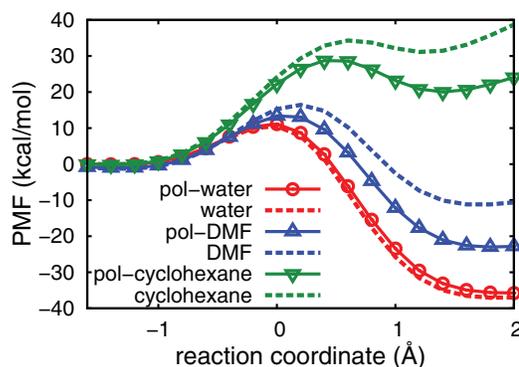

FIG. 7. PMF of the Menshutkin reaction $NH_3 + CH_3Cl \rightarrow NH_3CH_3^+ + Cl^-$ in water, DMF, and cyclohexane solutions calculated with the mean-field QM/MM method. Solvents are described with the TIP3P and OPLS models in the nonpolarizable case (dashed lines) and with the CRK model in the polarizable case (solid line). The QM calculation is performed at the BHHLYP/6-31+G(d,p) level.

Figure 7 displays the PMF in water, DMF, and cyclohexane solutions obtained with the mean-field QM/MM method. In aqueous solution, the PMF obtained with the CRK and TIP3P models are almost identical, as observed for the Finkelstein reaction. With the CRK model, the free energy of reaction (including solute thermal correction)[98] is calculated to be $-36.5 + 7.5 = -29.0$ kcal/mol, which falls within the error bar of the experimental result (-34 ± 10 kcal/mol).[60] In contrast, in DMF and cyclohexane solutions the PMF obtained with the CRK and OPLS models are rather different. In DMF, the free energy barrier $\Delta A^{\ddagger}$ and the reaction free energy ΔA_r [calculated as $A_{MF}(\xi = 2.0) - A_{MF}(\xi = -2.0)$] decrease by 3 and 12 kcal/mol when using the polarizable model. In cyclohexane, the $\Delta A^{\ddagger}$ and ΔA_r decrease by 6 and 18 kcal/mol when using the polarizable model. The significant decrease in the PMF suggests that the product ion pair is solvated more strongly by the polarizable model. Figure 8 depicts the solvent ESP for the Menshutkin reaction. In aqueous solution, the ESP profile is essentially identical between the CRK and TIP3P models, implying that the TIP3P model is doing a rather good job for describing the ionic solvation. On the other hand, in DMF and cyclohexane the ESP value increases by up to 1.0 V in the product region, which leads to a sizable (negative) increase in the solvation free energy [Fig. 8(d)]. A closer look at Fig. 8 also reveals that in aqueous solution the ESP acting on the chloride ion is greater (in magnitude) than that acting on the NH_3 moiety, while the opposite is true for DMF and cyclohexane solutions (that is, the ESP acting on NH_3 is greater than that acting on Cl^-). This is a reflection of the fact that water is a good solvator of anions due to hydrogen bonds, while aprotic solvents favor cations due to lone pairs/π-electrons in the solvent molecules.[64]

There are two likely reasons why the PMF in aqueous solution is not much different between polarizable and nonpolarizable models, while the PMF in nonaqueous solutions is modified significantly by explicit polarization. First, as noted in Sec. IV A, the polarizability of water is much smaller than organic solvents (see Table IV). The latter is thus expected to exhibit larger induced dipole in the vicinity of solvated ions. The second likely reason is that the empirical

FIG. 6. Potential energy profile of the Menshutkin reaction $NH_3 + CH_3Cl \rightarrow NH_3CH_3^+ + Cl^-$ in the gas phase. The basis sets are 6-311++G(3df,3p) for CCSD(T) and MP2 methods and 6-31+G(d,p) otherwise.

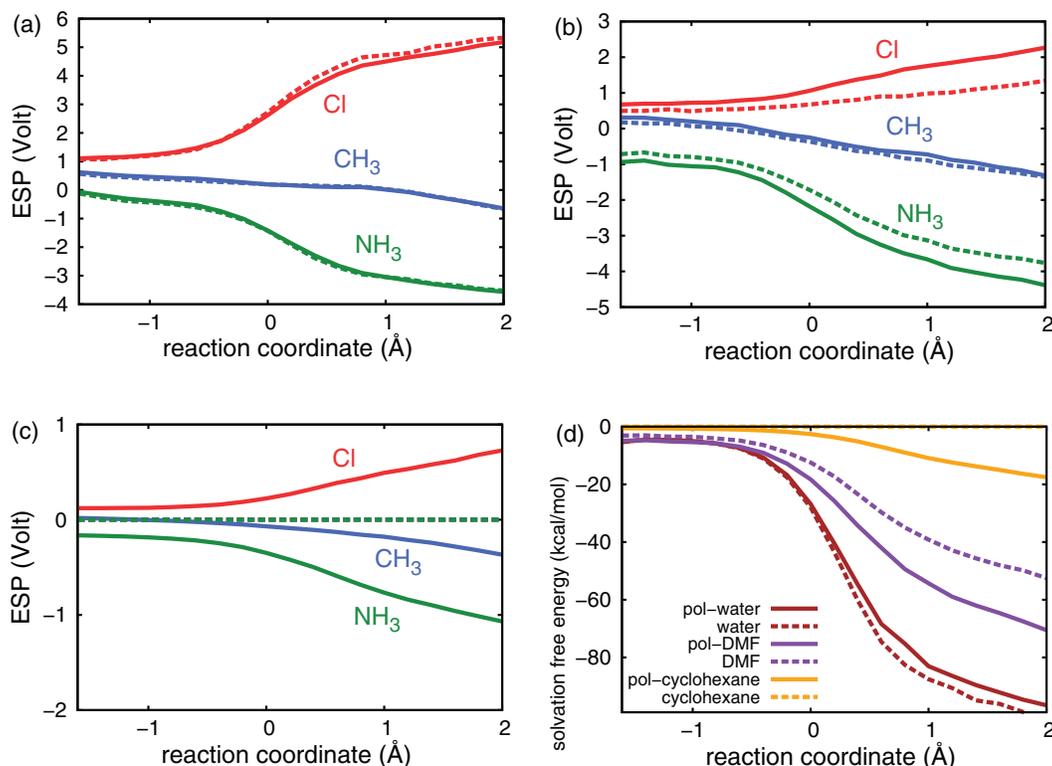FIG. 8. Solvent electrostatic potentials (ESPs) acting on the solute atoms, $\langle V_\alpha \rangle$, for the Menshutkin reaction in (a) water, (b) DMF, and (c) cyclohexane solutions. The results obtained with the polarizable and nonpolarizable solvent models are shown by solid and dashed lines, respectively. Panel (d) displays the solvation free energy calculated with the linear response approximation (see the main text). In panel (c), the ESP value calculated with the OPLS-UA model for cyclohexane is identically zero because the latter has no partial charge on the united CH_2 atoms.

nonpolarizable models are parametrized using the bulk properties of solvent. This means that implicit polarization included in the empirical models is highly dependent on the type of solvent. In protic solvents, the molecules interact strongly with each other via hydrogen bonds and thus their wave functions are distorted significantly by the surrounding molecules. Indeed, it has been suggested that the wave functions of water molecules are nearly maximally polarized in the bulk and thus no further polarization can be expected in the vicinity of ions.[99] The above observation is quite consistent with the present results for aqueous solution. On the other hand, the intermolecular interactions in nonaqueous solvents are relatively weak and they are probably insufficient for describing the ES interaction with a solvated ion. An extreme case of this is cyclohexane, for which the OPLS-UA model carries no partial charge and hence the model is not able to describe ES interactions at all.[58] We expect that the above two factors, namely, the relatively large polarizability of organic solvents and small implicit polarization included in the empirical models, are mainly responsible for the significant changes in the PMF observed above.

Figure 9 compares the PMF obtained with the mean-field QM/MM method and the COSMO continuum model. Here, all the QM/MM calculations are performed with the polarizable model. The PMF obtained with the QM/MM method reflects the type of solvents (i.e., protic versus aprotic), as expected, while the PMF obtained with the continuum model is rather insensitive to the value of dielectric constant. A similar trend is observed for the PCM model (see the supplementary

material).[116] We can also see from Fig. 9 that the reaction free energy in aqueous solution is underestimated by the continuum model (−26 kcal/mol) compared to the QM/MM method (−36 kcal/mol). This is probably because of the inadequate description of solute-solvent hydrogen bonds for the products by the continuum model.

C. Phosphoryl dissociation reaction

We now study the dissociation reaction of methyl phosphate dianion in solution

$$CH_3OPO_3{}^{2-} \longrightarrow CH_3O^- + PO_3^-. \qquad (44)$$

Since the solute has a large net charge of $-2e$, we can expect that the solvent effects are stronger than those observed for the Finkelstein and Menshutkin reactions. Experimentally, phosphoryl dissociation reactions, such as Eq. (44), have been studied as model reactions of phosphoryl transfer in biological systems, and it is known that the reaction rate depends dramatically on the polarity of solvents.[100–103] Theoretical work has also been performed for the above reaction, including a semiempirical QM/MM free energy calculation for aqueous solution[104] and a comparison of free energy profile in various solvents at the continuum level.[105]

We first calculated an approximate reaction path for Eq. (44) using the string method[50] at the B3LYP/6-31+G(d)/COSMO level. The string method determines an approximate minimum free energy path by optimizing a chain of replicas on the free energy surface.[50] Using the reaction
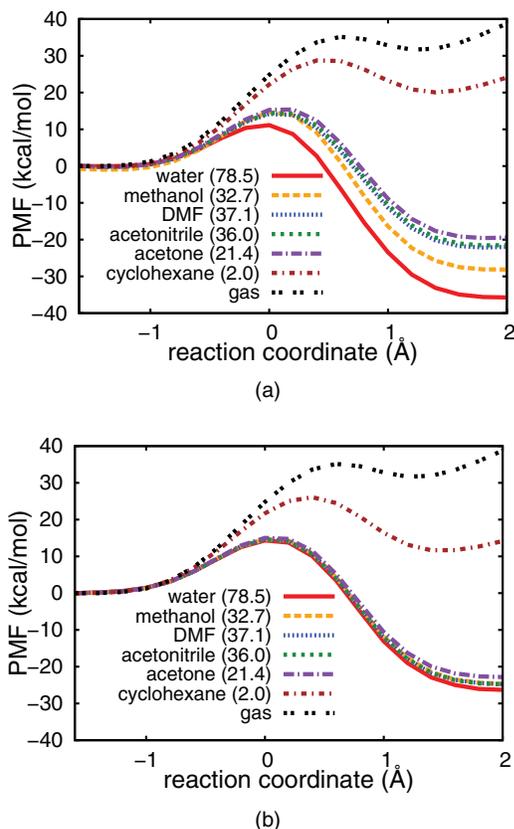
FIG. 9. PMF of the Menshutkin reaction in various solvents calculated with (a) the mean-field QM/MM method and (b) the COSMO continuum solvation model. In panel (a), all the solvents are described with the polarizable CRK model. The QM calculation is performed at the BHHLYP/6-31+G(d,p) level.

path thus obtained, the QM/MM PMF was calculated by integrating the free energy gradient in Eq. (30) along the reaction path. The reason for doing so is that a simple choice of distinguished reaction coordinate [e.g., $\xi = r(\text{P} - \text{O})$] failed to optimize the transition state on the QM/MM free energy surface. This problem may be due to the flatness of free energy surface in the transition state region.[105]

Figure 10 presents the potential energy profiles at various QM levels calculated along the reaction path obtained above. The CCSD(T) and MP2 methods again predict almost
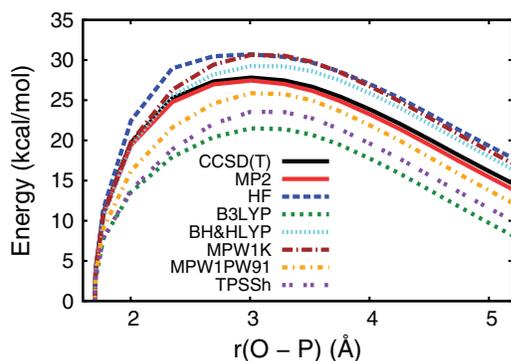

FIG. 10. Potential energy profiles of the phosphoryl dissociation reaction $\text{CH}_3\text{OPO}_3{}^{2-} \rightarrow \text{CH}_3\text{O}^- + \text{PO}_3{}^-$ in the gas phase. The basis sets are 6-311++G($3df,3p$) for CCSD(T) and MP2 methods and 6-31+G(d,p) otherwise.

identical results with a barrier height of 27 kcal/mol. Among the DFT methods employed, the MPW1PW91 and BHHLYP functionals give the best agreement with the CCSD(T) and MP2 results, while the B3LYP method somewhat underestimates the reaction barrier. We thus employ the MPW1PW91 method in the following calculations. (The latter functional has also been used in a recent study of phosphate hydrolysis reactions).[106]

Figure 11 displays the PMF in acetone and cyclohexane solutions obtained from the QM/MM calculation. In acetone, the free energy barrier obtained with the polarizable model (44 kcal/mol) is much higher than that obtained with the OPLS model (32 kcal/mol). A similar trend is observed for cyclohexane, in which the barrier height increases from 25 to 36 kcal/mol by using the polarizable model.[58] Thus, the polarization effect is significant for this reaction. Although experimental results for the activation free energy are not available for acetone and cyclohexane, the activation free energy for aqueous solution is known to be 44 kcal/mol (Ref. 101), which is close to the polarizable QM/MM result for acetone. For comparison, we also display in Fig. 11 the PMF obtained with the COSMO model. The latter gives reasonable PMF for both acetone and cyclohexane, although the excellent agreement for cyclohexane seems somewhat fortuitous considering the discrepancy between the COSMO and QM/MM results observed in Secs. IV A and IV B.

Figure 12 displays the ESP profiles in acetone and cyclohexane. Since the solute has a large net charge of $-2e$,

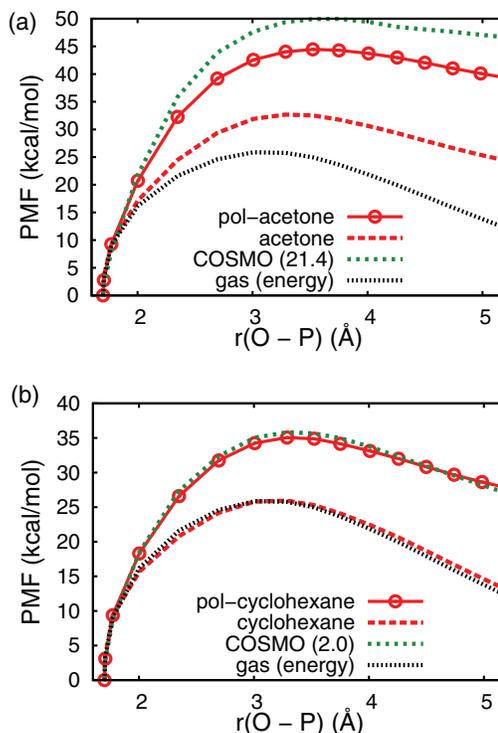FIG. 11. PMF of the phosphoryl dissociation reaction $\text{CH}_3\text{OPO}_3{}^{2-} \rightarrow \text{CH}_3\text{O}^- + \text{PO}_3{}^-$ in (a) acetone and (b) cyclohexane solutions calculated with the mean-field QM/MM method. For comparison, PMF obtained with the COSMO model and the potential energy profile in the gas phase are also shown. The QM calculation is performed at the MPW1PW91/6-31+G(d,p) level.
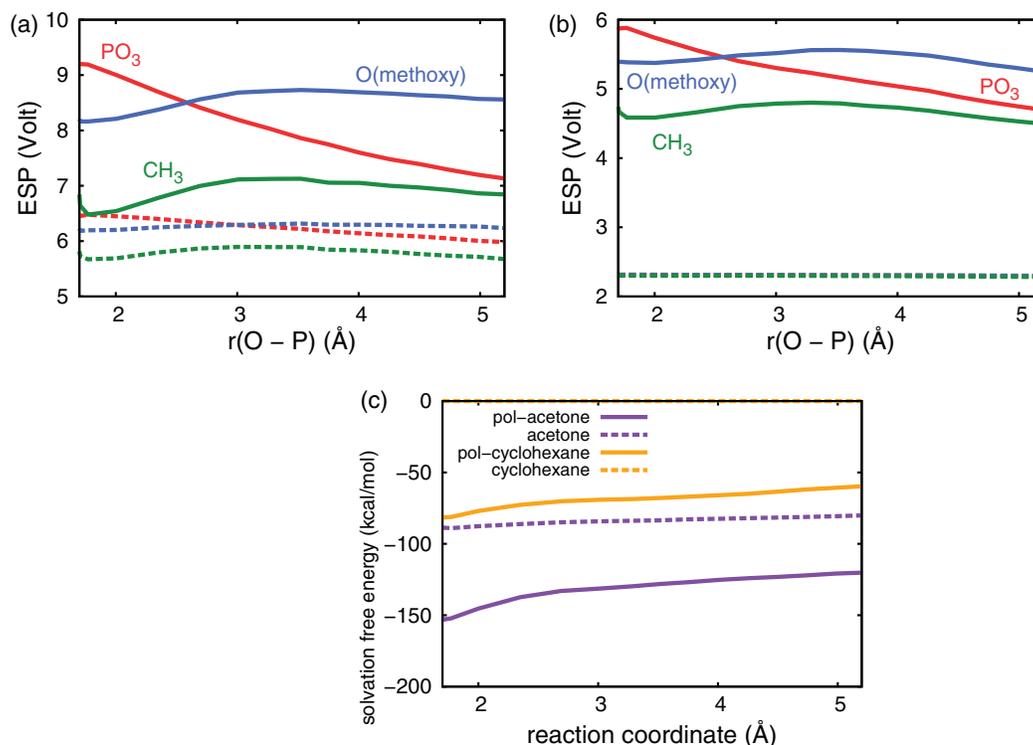
FIG. 12. Solvent electrostatic potentials (ESPs) acting on the solute atoms, $\langle V_\alpha \rangle$, for the phosphoryl dissociation reaction in (a) acetone and (b) cyclohexane solutions. The results obtained with the polarizable and nonpolarizable solvent models are shown by solid and dashed lines, respectively. Panel (c) displays the solvation free energy calculated with the linear response approximation (see the main text). The ESP values in panels (a) and (b) include the contribution of the Wigner potential.

the polarization effect is much more significant than observed for the Finkelstein and Menshutkin reactions. Indeed, the use of the polarizable models increases the solvent ESP by 2.0–2.5 V, which corresponds to an increase in the solvation free energy by >50 kcal/mol (evaluated with the linear response approximation). The solvation free energy for cyclohexane amounts to 60–80 kcal/mol in magnitude, which is produced solely by electronic polarization. It should be noted, however, that the PMF is not determined by the magnitude of solvation free energy but its variation along the reaction coordinate. Since the solute has a more localized charge distribution in the reactant state, the electronic polarization stabilizes the reactant more strongly than the transition state [Fig. 12(c)], and as a result, the electronic polarization increases the free energy barrier as observed in Fig. 11.

In addition to acetone and cyclohexane, we also attempted the mean-field QM/MM calculation for protic solvents (water and methanol). However, we find that the latter calculations are unstable when the TIP3P and OPLS models are used. Specifically, the partial charge on the oxygen atom of the methoxide ion (CH_3O^-) diverged during the solute-solvent iterations in the mean-field calculation. Similar divergence was observed with the polarizable CRK model. On one hand, this may be reasonable because the methoxide ion tends to remove a proton from the surrounding water molecules and is thus not very stable in aqueous solution. (Nevertheless, it should be noted that experimental free energies of reaction and activation are available for aqueous reaction).[101,107] On the other hand, the instability may be attributed in part to the ESP charge operator $\hat{Q}$ that tends to overestimate the solute-

solvent electrostatic interactions. To verify this, we have performed a preliminary calculation that employs a modified ESP charge operator that includes charge penetration effects at short range.[108] As a result, we find that the mean-field calculation can indeed be converged without severe difficulty. This indicates that the above instability is not an inherent limitation of the mean field approximation but rather due to insufficient accuracy of the (raw) charge operator for describing highly charged systems in protic solvents. To investigate the latter problem further, however, is somewhat beyond the scope of this paper and thus left for future study.[108]

V. CONCLUSIONS

In this paper, we have presented a mean-field QM/MM theory for electronically polarizable systems by employing a fragment based description of the total system and using a variational principle of free energy. The main advantage here is that it can reduce the number of QM calculations significantly while representing the molecular environment explicitly. We then applied the obtained method to prototypical reactions in solution and made a systematic comparison of polarization effects on PMF. The main conclusions can be summarized as follows:

- The PMF obtained from the mean field QM/MM calculation agrees quite well with that obtained from a direct QM/MM calculation (with fixed QM geometry). This indicates that statistical fluctuations of the QM wave function is of minor importance once the average

distortion of the QM wave function is properly taken into account in the free energy calculation.

- The empirical nonpolarizable models for water (such as the TIP3P) provide a PMF very close to that obtained with the polarizable model. This is presumably because water molecules in bulk condition are nearly maximally polarized and they cannot be further polarized in the vicinity of a charged molecule.

- The empirical models for organic solvents, on the other hand, underestimate the solvation of a charged species. This is because the empirical models are parametrized so as to reproduce the bulk condition, in which the organic molecules are often not significantly polarized. An extreme case of this is cyclohexane, which possesses no or little dipole moment in the bulk but it can solvate a charged molecule rather strongly via induced dipole moment.

- As a result, the PMF in nonaqueous solution is affected much more strongly by polarization than in aqueous solution. This trend is most evident for reactions with a large net charge, e.g., a phosphoryl dissociation reaction in Eq. (44). For the latter, the free energy barrier in acetone and cyclohexane solutions increases by >10 kcal/mol when using the polarizable model. This observation suggests that for quantitative comparison of some property between aqueous and nonaqueous environments, it is desirable to employ a polarizable force field at least for the latter environment, particularly when the solute has net charge.

- The PMF obtained with the continuum model is found to be rather insensitive to the dielectric constant of solvent for $\varepsilon > 20$. The obtained PMF is roughly in between the QM/MM results for polar protic solvents (e.g., water) and aprotic polar solvents (e.g., acetone).

- The ESP charge operator works well for the S_N2 reactions studied in Sec. IV, but has some difficulty in describing multiply charged ions in protic solvents. This is probably because the charge operator tends to overestimate the solute-solvent electrostatic interactions at short range. A more accurate description of the latter is thus desirable to study highly charged systems in aqueous solution.

ACKNOWLEDGMENTS

APPENDIX A: CRK SOLVENT MODEL

1. Derivation of Eq. (25)

To obtain Eq. (25), we expand the eigenenergy ϵ_i in Eq. (23) with respect to $\mathbf{v}_i$ such that

$$\epsilon_i(\mathbf{r}_i, \mathbf{v}_i) = \epsilon_i^0(\mathbf{r}_i) + \left(\frac{\partial \epsilon_i}{\partial \mathbf{v}_i}\right)^0 \cdot \mathbf{v}_i$$

$$+ \frac{1}{2}\mathbf{v}_i^T \left(\frac{\partial^2 \epsilon_i}{\partial \mathbf{v}_i \partial \mathbf{v}_i}\right)^0 \mathbf{v}_i + O(\mathbf{v}^3), \quad (A1)$$

where superscript 0 means that the quantity is evaluated at $\mathbf{v}_i = 0$. The first derivative of ϵ_i is

$$\left(\frac{\partial \epsilon_i}{\partial \mathbf{v}_i}\right)^0 = \langle \psi_i^0 | \hat{\mathbf{q}}_i | \psi_i^0 \rangle \equiv \mathbf{q}_i^0, \quad (A2)$$

which can be obtained from the Hellmann-Feynman theorem for Eq. (23). The second derivative of ϵ_i is given by

$$\left(\frac{\partial^2 \epsilon_i}{\partial \mathbf{v}_i \partial \mathbf{v}_i}\right)^0 = \left(\frac{\partial \mathbf{q}_i}{\partial \mathbf{v}_i}\right)^0 \equiv \mathbf{K}_i^0, \quad (A3)$$

where by the last equality we define the charge response matrix $\mathbf{K}_i^0$. Substitution of the above results into Eq. (A1) gives

$$\epsilon_i(\mathbf{r}_i, \mathbf{v}_i) \simeq \epsilon_i^0 + \mathbf{q}_i^0 \cdot \mathbf{v}_i + \frac{1}{2}\mathbf{v}_i \cdot \mathbf{K}_i^0 \cdot \mathbf{v}_i, \quad (A4)$$

where ϵ_i^0 and $\mathbf{q}_i^0$ are the energy and partial charge of the isolated solvent molecule. By differentiating the above equation with respect to $\mathbf{v}_i$, we obtain

$$\mathbf{q}_i = \mathbf{q}_i^0 + \mathbf{K}_i^0 \mathbf{v}_i + O(\mathbf{v}^2). \quad (A5)$$

Furthermore, by using the identity relation

$$\epsilon_i = \langle \psi_i | \hat{h}_i^0 | \psi_i \rangle + \mathbf{q}_i \cdot \mathbf{v}_i, \quad (A6)$$

which follows from Eq. (23), we obtain the expansion of the internal molecular energy

$$\langle \psi_i | \hat{h}_i^0 | \psi_i \rangle \simeq \epsilon_i^0 - \frac{1}{2}\mathbf{v}_i \cdot \mathbf{K}_i^0 \cdot \mathbf{v}_i + O(\mathbf{v}^3). \quad (A7)$$

Now inserting Eq. (A7) into Eq. (21), we obtain Eq. (25). In practical simulation, we model the first term in Eq. (25), [namely, $\sum_i \epsilon_i^0$] with a classical force field consisting of bonds/angles/dihedrals (for flexible molecules) or simply set it to zero (for rigid molecules). The gas-phase charge $\mathbf{q}_i^0$ and the response matrix $\mathbf{K}_i^0$ are obtained from *ab initio* calculation for an isolated molecule.[45] Polarizable force fields thus obtained have been utilized successfully in previous MD simulations.[45,48,109–113]

2. Validation of the CRK solvent model

In this paper, the CRK solvent model is obtained by replacing the permanent charge of an empirical nonpolarizable model with the gas-phase partial charge $\mathbf{q}_i^0$ and adding the charge response matrix $\mathbf{K}_i^0$. The latter are calculated with *ab initio* methods, while the molecular geometry and the LJ parameters are taken from the underlying nonpolarizable model. This approach has been utilized previously to model various solvents with reasonable success.[48,109,113] Since we have performed no further tuning of the CRK model (especially the LJ parameters), it is important to check the validity of the obtained model. For this purpose, we calculated the density and vaporization enthalpy of each solvent using the present CRK model (Table V). The radial distribution functions for water are also presented in the supplementary material.[116] Table V shows that the present CRK model provides reasonable agreement with experiment, although there is a slight tendency that density and vaporization enthalpy are overestimated for aprotic polar solvents. To check whether

TABLE V. Density ρ (in g/cm^3) and vaporization enthalpy ΔH_v (in kcal/mol) of bulk solvents calculated with the CRK and empirical (non-polarizable) models. Vaporization enthalpy was calculated as $\Delta H_v \simeq -E + RT$, where E is the average interaction energy per molecule and R is the gas constant.

	Water	MeOH	MeCN	Acetone	DMF	CHX
ρ(pol)	0.999	0.765	0.894	0.841	0.976	0.739
ρ(nonpol)	0.987	0.761	0.802	0.774	0.911	0.739
ρ(expt)	0.997	0.787	0.782	0.791	0.873	0.779
ΔH_v(pol)	9.1	9.0	8.8	8.5	12.0	8.3
ΔH_v(nonpol)	10.2	8.6	7.5	7.2	9.5	8.3
ΔH_v(expt)	10.5	8.4–8.9	7.9	7.5	10.4	7.6

this does not affect the main conclusions in this paper, we performed additional test calculations. Specifically, we calculated QM/MM PMF with a modified CRK model obtained by (i) using slightly different values of the A parameter in Eq. (39) that affects the strength of Coulomb interactions at short range; (ii) removing the diffuse basis functions from the calculation of $\mathbf{q}_i^0$ and $\mathbf{K}_i^0$ (to account for the suppression of polarizability in condensed phases);[113,114] or (iii) using the SPC (Ref. 115) model as the underlying water model. The obtained results are summarized in Table VI, which shows that the PMF is rather insensitive to the above modifications. More specifically, the variation of the free energy barrier is at

TABLE VI. Density ρ (in g/cm^3) and vaporization enthalpy ΔH_v (in kcal/mol) of bulk water and acetone at 298 K and 1 atm calculated with the polarizable and nonpolarizable models. "pol-TIP3P" refers to the CRK water model derived from the TIP3P model, while "pol-SPC" refers to the CRK water model derived from the SPC model. The CRK model for acetone is derived from the OPLS-AA model. "TIP3P," "SPC," and "OPLS-AA" refer to the standard empirical models. The value of Coulomb damping parameter A is given in parentheses. In the main text, A was set to 2.7 for water and 2.6 for other solvents. The $\Delta A^\ddagger$ stands for the free energy barrier (in kcal/mol) for the Finkelstein reaction $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$ calculated with the mean-field QM/MM method at the BHHLYP/6-31+G(d,p) level. The CRK matrix was calculated with the B3LYP/aug-cc-pVTZ method unless otherwise noted.

Model	ρ	ΔH_v	$\Delta A^\ddagger$
	Water		
TIP3P	0.987	10.17	23.3
pol-TIP3P (A = 2.6)[a]	0.992	9.23	23.3
pol-TIP3P (A = 2.6)	1.024	10.34	23.9
pol-TIP3P (A = 2.7)	0.999	9.08	23.0
pol-TIP3P (A = 2.8)	0.963	8.10	22.3
SPC	0.977	10.55	23.4
pol-SPC (A = 2.6)[a]	0.985	10.21	23.4
pol-SPC (A = 2.6)	0.993	11.98	24.2
pol-SPC (A = 2.7)	0.995	10.05	23.1
pol-SPC (A = 2.8)	0.967	8.73	22.1
	Acetone		
OPLS-AA	0.774	7.2	11.2
pol (A = 2.6)[a]	0.826	8.0	15.8
pol (A = 2.6)	0.841	8.5	16.2
pol (A = 2.7)	0.841	8.5	16.2
pol (A = 2.8)	0.840	8.5	16.2

[a]CRK matrix calculated with the B3LYP/cc-pVTZ method.

most ± 1 kcal/mol compared to the reference values in Table I, although the bulk properties are more sensitive to the above modifications. We thus expect that, although the present polarizable model is not of highest accuracy compared to other calibrated models, it is sufficient for making quantitative discussion of polarization effects on the PMF.

APPENDIX B: FREE ENERGY GRADIENT IN THE MEAN FIELD APPROXIMATION

Here we consider the nuclear derivative of $A_{\mathrm{MF}}(\mathbf{R})$, i.e.,

$$\frac{\partial}{\partial R} A_{\mathrm{MF}}(\mathbf{R}) = \frac{\partial}{\partial R}[\langle \tilde{\Psi} | \hat{H}^0 | \tilde{\Psi} \rangle + \Delta \mu(\mathbf{R}, \tilde{\mathbf{Q}})], \qquad (B1)$$

where R is an arbitrary element of $\mathbf{R}$. The derivative of the first term, $\langle \tilde{\Psi} | \hat{H}^0 | \tilde{\Psi} \rangle$, can be obtained as

$$\frac{\partial}{\partial R}\langle \tilde{\Psi} | \hat{H}^0 | \tilde{\Psi} \rangle = \frac{\partial}{\partial R}\{\mathcal{E}_{\mathrm{QM}}(\mathbf{R}, \bar{\mathbf{V}}) - \bar{\mathbf{V}} \cdot \tilde{\mathbf{Q}}\}$$

$$= \left. \frac{\partial \mathcal{E}_{\mathrm{QM}}(\mathbf{R}, \bar{\mathbf{V}})}{\partial R}\right|_{\bar{\mathbf{V}}} - \bar{\mathbf{V}} \cdot \frac{\partial \tilde{\mathbf{Q}}}{\partial R}, \quad (B2)$$

where we used the definition of $\mathcal{E}_{\mathrm{QM}}(\mathbf{R}, \bar{\mathbf{V}})$ in Eq. (18) and the relation

$$\left. \frac{\partial \mathcal{E}_{\mathrm{QM}}(\mathbf{R}, \bar{\mathbf{V}})}{\partial \bar{\mathbf{V}}}\right|_{\mathbf{R}} = \tilde{\mathbf{Q}}, \qquad (B3)$$

which follows from the Hellmann-Feynman theorem. The derivative of the second term, i.e., $\Delta \mu(\mathbf{R}, \tilde{\mathbf{Q}})$, can be written as

$$\frac{\partial}{\partial R}\Delta \mu(\mathbf{R}, \tilde{\mathbf{Q}}) = \left. \frac{\partial \Delta \mu(\mathbf{R}, \tilde{\mathbf{Q}})}{\partial R}\right|_{\tilde{\mathbf{Q}}} + \left. \frac{\partial \Delta \mu(\mathbf{R}, \tilde{\mathbf{Q}})}{\partial \tilde{\mathbf{Q}}}\right|_{\mathbf{R}} \cdot \frac{\partial \tilde{\mathbf{Q}}}{\partial R}. \tag{B4}$$

Using the definition of $\Delta \mu(\mathbf{R}, \tilde{\mathbf{Q}})$ in Eq. (29), we can express the partial derivative of $\Delta \mu(\mathbf{R}, \tilde{\mathbf{Q}})$ as follows:

$$\left. \frac{\partial \Delta \mu(\mathbf{R}, \tilde{\mathbf{Q}})}{\partial R}\right|_{\tilde{\mathbf{Q}}} = \left\langle \left. \frac{\partial \mathcal{E}_{\mathrm{MM}}(\mathbf{R}, \mathbf{r}, \tilde{\mathbf{Q}})}{\partial R}\right|_{\mathbf{r}, \tilde{\mathbf{Q}}} \right\rangle, \quad (B5a)$$

$$\left. \frac{\partial \Delta \mu(\mathbf{R}, \tilde{\mathbf{Q}})}{\partial \tilde{\mathbf{Q}}}\right|_{\mathbf{R}} = \left\langle \left. \frac{\partial \mathcal{E}_{\mathrm{MM}}(\mathbf{R}, \mathbf{r}, \tilde{\mathbf{Q}})}{\partial \tilde{\mathbf{Q}}}\right|_{\mathbf{R}, \mathbf{r}} \right\rangle, \quad (B5b)$$

where the ensemble average is defined by Eq. (20). The partial derivative of $\mathcal{E}_{\mathrm{MM}}(\mathbf{R}, \mathbf{r}, \tilde{\mathbf{Q}})$ can be readily obtained by considering virtual displacements of $\mathcal{E}_{\mathrm{MM}}$ with respect to $(\mathbf{R}, \mathbf{r}, \tilde{\mathbf{Q}})$, that is

$$\delta \mathcal{E}_{\mathrm{MM}} = \sum_i \delta \epsilon_i + \delta \tilde{\mathbf{Q}} \cdot \sum_i \mathbf{D}_i \mathbf{q}_i + \sum_i \tilde{\mathbf{Q}} \cdot (\delta \mathbf{D}_i) \cdot \mathbf{q}_i$$

$$+ \frac{1}{2}\sum_{i \neq j} \mathbf{q}_i \cdot (\delta \mathbf{D}_{ij}) \cdot \mathbf{q}_j + \delta U_{\mathrm{vdW}}, \qquad (B6)$$

which suggests

$$\left. \frac{\partial \mathcal{E}_{\mathrm{MM}}(\mathbf{R}, \mathbf{r}, \tilde{\mathbf{Q}})}{\partial R}\right|_{\mathbf{r}, \tilde{\mathbf{Q}}} = \sum_i \tilde{\mathbf{Q}} \cdot \frac{\partial \mathbf{D}_i}{\partial R} \cdot \mathbf{q}_i + \frac{\partial U_{\mathrm{vdW}}}{\partial R},$$

$$(B7a)$$

$$\left.\frac{\partial \mathcal{E}_{\text{MM}}(\mathbf{R}, \mathbf{r}, \tilde{\mathbf{Q}})}{\partial \tilde{\mathbf{Q}}}\right|_{\mathbf{R}, \mathbf{r}} = \sum_i \mathbf{D}_i \mathbf{q}_i = \mathbf{V}, \qquad \text{(B7b)}$$

where we have utilized Eqs. (24) and (26) to cancel several terms in Eq. (B6), and also assumed that $\mathbf{K}_i^0$ is independent of $\mathbf{r}_i$. The derivative of $\Delta\mu(\mathbf{R}, \tilde{\mathbf{Q}})$ then becomes

$$\frac{\partial}{\partial R} \Delta\mu(\mathbf{R}, \tilde{\mathbf{Q}}) = \left\langle \sum_i \tilde{\mathbf{Q}} \cdot \frac{\partial \mathbf{D}_i}{\partial R} \cdot \mathbf{q}_i + \frac{\partial U_{\text{vdW}}}{\partial R} \right\rangle + \langle \mathbf{V} \rangle \cdot \frac{\partial \tilde{\mathbf{Q}}}{\partial R}. \tag{B8}$$

Combining Eqs. (B2) and (B8), we obtain the gradient expression for $A_{\text{MF}}(\mathbf{R})$ in Eq. (30). Importantly, the latter expression is identical to that obtained previously for nonpolarizable models.[34] This is because the function $\mathcal{E}_{\text{MM}}(\mathbf{R}, \mathbf{r}, \tilde{\mathbf{Q}})$ is stationarized with respect to the solvent wave function $\{\psi_i\}$, and hence all the derivatives with respect to $\mathbf{q}_i$ or $\mathbf{v}_i$ vanish in the final expression.

APPENDIX C: FLUCTUATION OF THE SOLUTE WAVE FUNCTION

In the mean-field QM/MM calculation, the solute wave function is fixed at the self-consistent state $\tilde{\Psi}(\mathbf{R})$. This means that we neglect statistical fluctuations of the true wave function $\Psi(\mathbf{R}, \mathbf{r})$ about the self-consistent state $\tilde{\Psi}(\mathbf{R})$. Here, we consider how to evaluate the latter effect by using the CRK of the solute molecule. Specifically, we first write the exact QM/MM free energy as

$$A(\mathbf{R}) = A_{\text{MF}}(\mathbf{R}) - \frac{1}{\beta} \ln\langle \exp(-\beta \Delta E_{\text{tot}}) \rangle, \qquad \text{(C1)}$$

where $\Delta E_{\text{tot}} = E_{\text{tot}} - E_{\text{tot}}^{(\text{MF})}$, and the ensemble average is defined by Eq. (20). Here, $E_{\text{tot}}(\mathbf{R}, \mathbf{r})$ denotes the exact total energy given by

$$E_{\text{tot}}(\mathbf{R}, \mathbf{r}) = \langle \Psi' | \hat{H}^0 | \Psi' \rangle + \mathbf{Q}' \cdot \mathbf{V}' + E_{\text{MM}}(\mathbf{q}', \mathbf{v}') \qquad \text{(C2)}$$

with $\mathbf{Q}' = \langle \Psi' | \hat{Q} | \Psi' \rangle$ and $\mathbf{V}' = \sum_i \mathbf{D}_i \mathbf{q}_i'$, and E_{MM} is defined by

$$E_{\text{MM}}(\mathbf{q}', \mathbf{v}') = \sum_i \epsilon_i^0 - \frac{1}{2} \sum_i \mathbf{v}_i' \cdot \mathbf{K}_i^0 \cdot \mathbf{v}_i' + \frac{1}{2} \sum_{i \neq j} \mathbf{q}_i' \cdot \mathbf{D}_{ij} \cdot \mathbf{q}_j' + U_{\text{vdW}}. \qquad \text{(C3)}$$

Note that the prime symbol is attached on quantities appearing in the exact QM/MM calculation. The $E_{\text{tot}}^{(\text{MF})}(\mathbf{R}, \mathbf{r})$ represents the mean-field approximation to $E_{\text{tot}}(\mathbf{R}, \mathbf{r})$ given by

$$E_{\text{tot}}^{(\text{MF})}(\mathbf{R}, \mathbf{r}) = \langle \tilde{\Psi} | \hat{H}^0 | \tilde{\Psi} \rangle + \tilde{\mathbf{Q}} \cdot \mathbf{V} + E_{\text{MM}}(\mathbf{q}, \mathbf{v}) \qquad \text{(C4)}$$

with $\tilde{\mathbf{Q}} = \langle \tilde{\Psi} | \hat{Q} | \tilde{\Psi} \rangle$ and $\mathbf{V} = \sum_i \mathbf{D}_i \mathbf{q}_i$. The QM wave functions Ψ' and $\tilde{\Psi}$ are calculated in the presence of $\mathbf{V}'$ and $\bar{\mathbf{V}} = \langle \mathbf{V} \rangle$, respectively. Using the eigenenergy corresponding to Ψ' and $\tilde{\Psi}$, one can rewrite Eqs. (C2) and (C4) as follows:

$$E_{\text{tot}}(\mathbf{R}, \mathbf{r}) = \mathcal{E}_{\text{QM}}(\mathbf{R}, \mathbf{V}') + E_{\text{MM}}(\mathbf{q}', \mathbf{v}'), \qquad \text{(C5a)}$$

$$E_{\text{tot}}^{(\text{MF})}(\mathbf{R}, \mathbf{r}) = \mathcal{E}_{\text{QM}}(\mathbf{R}, \bar{\mathbf{V}}) + \tilde{\mathbf{Q}} \cdot (\mathbf{V} - \bar{\mathbf{V}}) + E_{\text{MM}}(\mathbf{q}, \mathbf{v}). \qquad \text{(C5b)}$$

As noted earlier, it is expensive to calculate $\mathcal{E}_{\text{QM}}(\mathbf{R}, \mathbf{V}')$ for many solvent configurations. To avoid this, we expand $\mathcal{E}_{\text{QM}}(\mathbf{R}, \mathbf{V}')$ in terms of $\mathbf{V}'$ up to second order, namely,

$$\mathcal{E}_{\text{QM}}(\mathbf{R}, \mathbf{V}') \simeq \mathcal{E}_{\text{QM}}(\mathbf{R}, \bar{\mathbf{V}}) + \tilde{\mathbf{Q}} \cdot (\mathbf{V}' - \bar{\mathbf{V}}) + \frac{1}{2}(\mathbf{V}' - \bar{\mathbf{V}}) \cdot \bar{\mathbf{K}}_{\text{QM}} \cdot (\mathbf{V}' - \bar{\mathbf{V}}), \qquad \text{(C6)}$$

where

$$\bar{\mathbf{K}}_{\text{QM}} = \left.\frac{\partial^2 \mathcal{E}_{\text{QM}}(\mathbf{R}, \mathbf{V})}{\partial \mathbf{V} \partial \mathbf{V}}\right|_{\mathbf{V} = \bar{\mathbf{V}}} = \left.\frac{\partial \mathbf{Q}}{\partial \mathbf{V}}\right|_{\mathbf{V} = \bar{\mathbf{V}}} \qquad \text{(C7)}$$

is the CRK of the solute molecule. By using the above approximation, we obtain

$$\Delta E_{\text{tot}} \simeq \frac{1}{2}(\mathbf{V}' - \bar{\mathbf{V}}) \cdot \bar{\mathbf{K}}_{\text{QM}} \cdot (\mathbf{V}' - \bar{\mathbf{V}}) + \tilde{\mathbf{Q}} \cdot (\mathbf{V}' - \mathbf{V}) + E_{\text{MM}}(\mathbf{q}', \mathbf{v}') - E_{\text{MM}}(\mathbf{q}, \mathbf{v}). \qquad \text{(C8)}$$

The remaining problem is how to evaluate $\mathbf{q}'$, $\mathbf{v}'$, and $\mathbf{V}'$, given that $\bar{\mathbf{V}}$ and $\tilde{\mathbf{Q}}$ are already available from the mean-field calculation. While the $\mathbf{q}'$, $\mathbf{v}'$, $\mathbf{Q}'$, and $\mathbf{V}'$ depend implicitly on Ψ', we want to avoid any repeated calculation of Ψ'. To achieve this, we expand $\mathbf{Q}'$ up to first order in terms of $\mathbf{V}'$. One can then obtain those variables by solving the following coupled equations:

$$\mathbf{q}_i' = \mathbf{q}_i^0 + \mathbf{K}_i^0 \mathbf{v}_i', \qquad \text{(C9a)}$$

$$\mathbf{v}_i' = \sum_{j(\neq i)} \mathbf{D}_{ij} \mathbf{q}_j' + \mathbf{D}_i^T \mathbf{Q}', \qquad \text{(C9b)}$$

$$\mathbf{Q}' = \tilde{\mathbf{Q}} + \bar{\mathbf{K}}_{\text{QM}} \cdot (\mathbf{V}' - \bar{\mathbf{V}}), \qquad \text{(C9c)}$$

$$\mathbf{V}' = \sum_i \mathbf{D}_i \mathbf{q}_i'. \qquad \text{(C9d)}$$

In this way, one can evaluate ΔE_{tot} in Eq. (C1) without explicitly calculating Ψ' for each solvent configuration. We note that when the solvent is described with a nonpolarizable model, the ΔE_{tot} takes a much simpler form[34]

$$\Delta E_{\text{tot}} \simeq \frac{1}{2}(\mathbf{V}' - \bar{\mathbf{V}}) \cdot \bar{\mathbf{K}}_{\text{QM}} \cdot (\mathbf{V}' - \bar{\mathbf{V}}). \qquad \text{(C10)}$$

[1] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[2] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1989).
[3] A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and solutions* (Wiley, New York, 1991).
[4] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[5] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).
[6] W. L. Jorgensen, J. Chem. Theory Comput. **3**, 1877 (2007).
[7] A. Warshel, M. Kato, and A. V. Pisliakov, J. Chem. Theory Comput. **3**, 2034 (2007).
[8] D. P. Geerke, S. Thiel, W. Thiel, and W. F. van Gunsteren, J. Chem. Theory Comput. **3**, 1499 (2007).
[9] Z. Lu and Y. Zhang, J. Chem. Theory Comput. **4**, 1237 (2008).
[10] W. L. Jorgensen, N. A. McDonald, M. Selmi, and P. R. Rablen, J. Am. Chem. Soc. **117**, 11809 (1995).
[11] O. Acevedo and W. L. Jorgensen, J. Phys. Chem. B **114**, 8425 (2010).
[12] O. Acevedo and K. Armacost, J. Am. Chem. Soc. **132**, 1966 (2010).
[13] C. J. R. Illingworth, K. E. Parkes, C. R. Snell, S. Marti, V. Moliner, and C. A. Raynolds, Mol. Phys. **106**, 1511 (2008).
[14] V. Tipmanee, H. Oberhofer, M. Park, K. S. Kim, and J. Blumberger, J. Am. Chem. Soc. **132**, 17032 (2010).

[15] D. Jiao, P. A. Golubkov, T. A. Darden, and P. Ren, Proc. Natl. Acad. Sci. U.S.A. **105**, 6290 (2008).

[16] G. A. Kaminski, J. Phys. Chem. B **109**, 5884 (2005).

[17] C. M. MacDermaid and G. A. Kaminski, J. Phys. Chem. B **111**, 9036 (2007).

[18] E. Rosta, M. Haranczyk, Z. T. Chu, and A. Warshel, J. Phys. Chem. B **112**, 5680 (2008).

[19] I. Fdez. Galvan, M. L. Sanchez, M. E. Martin, F. J. Olivares del Valle, and M. A. Aguilar, J. Chem. Phys. **118**, 255 (2003).

[20] I. Fdez. Galvan, M. L. Sanchez, M. E. Martin, F. J. Olivares del Valle, and M. A. Aguilar, Comput. Phys. Commun. **155**, 244 (2003).

[21] I. Fdez. Galvan, M. E. Martin, M. A. Aguilar, and M. F. Ruiz-Lopez, J. Chem. Phys. **124**, 214504 (2006).

[22] F. F. Garcia-Prieto, I. Fdez. Galvan, M. A. Aguilar, and M. Elena Martin, J. Chem. Phys. **135**, 194502 (2011).

[23] N. Okuyama-Yoshida, M. Nagaoka, and T. Yamabe, Int. J. Quantum Chem. **70**, 95 (1998).

[24] N. Okuyama-Yoshida, K. Kataoka, M. Nagaoka, and T. Yamabe, J. Chem. Phys. **113**, 3519 (2000).

[25] H. Hirao, Y. Nagae, and M. Nagaoka, Chem. Phys. Lett. **348**, 350 (2001).

[26] J. Tomasi and M. Persico, Chem. Rev. **94**, 2027 (1994).

[27] J. Tomasi, B. Mennucci, and R. Cammi, Chem. Rev. **105**, 2999 (2005).

[28] S. Ten-no, F. Hirata, and S. Kato, J. Chem. Phys. **100**, 7443 (1994).

[29] H. Sato, F. Hirata, and S. Kato, J. Chem. Phys. **105**, 1546 (1996).

[30] *Molecular Theory of Solvation*, edited by F. Hirata (Kluwer, New York, 2004).

[31] H. Hu, Z. Lu, and W. Yang, J. Chem. Theory Comput. **3**, 390 (2007).

[32] H. Hu, Z. Lu, J. M. Parks, S. K. Burger, and W. Yang, J. Chem. Phys. **128**, 034105 (2008).

[33] H. Hu and W. Yang, Annu. Rev. Phys. Chem. **59**, 573 (2008).

[34] T. Yamamoto, J. Chem. Phys. **129**, 244104 (2008).

[35] By "analytical" we mean that no coupled equations need to be solved for the derivative of self-consistent charges or mean solvent ESP to evaluate the free energy gradient.

[36] D. Borgis and A. Staib, Chem. Phys. Lett. **238**, 187 (1995).

[37] M. J. Field, Mol. Phys. **91**, 835 (1997).

[38] J. Gao, J. Chem. Phys. **109**, 2346 (1998).

[39] B. D. Bursulaya and H. J. Kim, J. Chem. Phys. **108**, 3277 (1998).

[40] A. E. Lefohn, M. Ovchinnikov, and G. A. Voth, J. Phys. Chem. B **105**, 6628 (2001).

[41] W. Xie and J. Gao, J. Chem. Theory Comput. **3**, 1890 (2007).

[42] W. Xie, M. Orozco, D. G. Truhlar, and J. Gao, J. Chem. Theory Comput. **5**, 459 (2009).

[43] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).

[44] A. Morita and S. Kato, J. Am. Chem. Soc. **119**, 4021 (1997).

[45] A. Morita and S. Kato, J. Chem. Phys. **108**, 6809 (1998).

[46] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, J. Phys. Chem. **97**, 10269 (1993).

[47] In the variational procedure, one does not need to explicitly consider the variation of $\{\psi_i\}$ because the latter is defined so as to minimize the $\mathcal{E}_{MM}$ function for each given $(\mathbf{R}, \mathbf{r}, \tilde{\mathbf{Q}})$.

[48] H. Nakano, T. Yamamoto, and S. Kato, J. Chem. Phys. **132**, 044106 (2010).

[49] We also evaluated the integral over ξ' with the cubic spline method, but the result was essentially the same as that obtained with Eq. (32).

[50] W. E, W. Ren, and E. Vanden-Eijnden, Phys. Rev. B **66**, 052301 (2002).

[51] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su *et al.*, J. Comput. Chem. **14**, 1347 (1993).

[52] M. A. Spackman, J. Comput. Chem. **17**, 1 (1996).

[53] C. M. Aikens, S. P. Webb, R. L. Bell, G. D. Fletcher, M. W. Schmidt, and M. S. Gordon, Theor. Chem. Acc. **110**, 233 (2003).

[54] J. L. Bentz, R. M. Olson, M. S. Gordon, M. W. Schmidt, and R. A. Kendall, Comput. Phys. Commun. **176**, 589 (2007).

[55] H. Z. S. Wang and Y. Zhang, J. Comput. Chem. **30**, 2706 (2009).

[56] W. Smith, T. R. Forester, and I. T. Todorov, DLPOLY 2.20 User Manual, CCLRC, Daresbury Laboratory, Daresbury, Warrington, 2009.

[57] H. J. Böhm, I. R. McDonald, and P. A. Madden, Mol. Phys. **49**, 347 (1983).

[58] We also performed the PMF calculation with the OPLS-AA model for cyclohexane, but the result was essentially identical to that obtained with the OPLS-UA model.

[59] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. G. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).

[60] J. Gao and X. Xia, J. Am. Chem. Soc. **115**, 9667 (1993).

[61] B. T. Thole, Chem. Phys. **59**, 341 (1981).

[62] J. K. Nagle, J. Am. Chem. Soc. **112**, 4741 (1990).

[63] J. Caldwell, L. X. Dang, and P. A. Kollman, J. Am. Chem. Soc. **112**, 9144 (1990).

[64] C. Reichardt, *Solvents and Solvent Effects in Organic Chemistry* (Wiley-VCH, Germany, 2003).

[65] J. Chandrasekhar, S. F. Smith, and W. L. Jorgensen, J. Am. Chem. Soc. **107**, 154 (1985).

[66] J. Chandrasekhar and W. L. Jorgensen, J. Am. Chem. Soc. **107**, 2974 (1985).

[67] T. N. Truong and E. V. Stefanovich, J. Phys. Chem. **99**, 14700 (1995).

[68] C. S. Pomelli and J. Tomasi, J. Phys. Chem. A **101**, 3561 (1997).

[69] H. Sato and S. Sakaki, J. Phys. Chem. A **108**, 1629 (2004).

[70] G. Vayner, K. N. Houk, W. L. Jorgensen, and J. I. Brauman, J. Am. Chem. Soc. **126**, 9054 (2004).

[71] H. Freedman and T. N. Truong, J. Phys. Chem. B **109**, 4726 (2005).

[72] D. Ardura, R. López, and T. L. Sordo, J. Phys. Chem. B **109**, 23618 (2005).

[73] T. Bucko, J. Phys.: Condens. Matter **20**, 064211 (2008).

[74] M. Higashi and D. G. Truhlar, J. Chem. Theory Comput. **4**, 790 (2008).

[75] G. Hummer, L. R. Pratt, and A. E. García, J. Phys. Chem. **100**, 1206 (1996).

[76] T. Darden, D. Pearlman, and L. G. Pedersen, J. Chem. Phys. **109**, 10921 (1998).

[77] C. Sagui and T. Darden, Annu. Rev. Biophys. Biomol. Struct. **28**, 155 (1999).

[78] The Wigner potential $\mathbf{V}_0$ is nearly constant as a function of the reaction coordinate and takes about 1.9 and 1.3 V for the Finkelstein reaction in water and acetone solvents and 2.6 and 2.3 V for the phosphoryl dissociation reaction in acetone and cyclohexane solvents. The $\mathbf{V}_0$ is determined by the net charge of the solute molecule and the size of the simulation box.

[79] A. Klamt and G. Schuumann, J. Chem. Soc., Perkin Trans. **2**, 799 (1993).

[80] O. Acevedo and W. L. Jorgensen, Acc. Chem. Res. **43**, 142 (2010).

[81] M. Irani, M. Haqgu, A. Talebi, and M. R. Gholami, J. Mol. Struct.: THEOCHEM **893**, 73 (2009).

[82] C. J. Cramer, *Essentials of Computational Chemistry* (Wiley, New York, 2002).

[83] Y. Kim, J. R. Mohrig, and D. G. Truhlar, J. Am. Chem. Soc. **132**, 11071 (2010).

[84] W. J. Albery and M. M. Kreevoy, Adv. Phys. Org. Chem. **16**, 87 (1978).

[85] *Theoretical Aspects of Physical Organic Chemistry*, edited by S. S. Shaik, H. B. Schlegel, and S. Wolfe (Wiley, 1992).

[86] J. Gao, J. Am. Chem. Soc. **113**, 7796 (1991).

[87] S. P. Webb and M. S. Gordon, J. Phys. Chem. A **103**, 1265 (1999).

[88] S. Shaik, A. Ioffe, A. C. Reddy, and A. Pross, J. Am. Chem. Soc. **116**, 262 (1994).

[89] V. Dillet, D. Rinaldi, J. Bertran, and J.-L. Rivail, J. Chem. Phys. **104**, 9437 (1996).

[90] T. N. Truong, T. T. Truong, and E. V. Stefanovich, J. Chem. Phys. **107**, 1881 (1997).

[91] C. Amovilli, B. Mennucci, and F. M. Floris, J. Phys. Chem. B **102**, 3023 (1998).

[92] K. Naka, H. Sato, A. Morita, F. Hirata, and S. Kato, Theor. Chem. Acc. **102**, 165 (1999).

[93] S. P. Webb and M. S. Gordon, J. Phys. Chem. A **103**, 1265 (1999).

[94] I. Fdez. Galvan, M. E. Martin, and M. A. Aguilar, J. Comput. Chem. **25**, 1227 (2004).

[95] J. J. Ruiz-Pernia, E. Silla, I. Tunon, S. Marti, and V. Moliner, J. Phys. Chem. B **108**, 8427 (2004).

[96] M. Higashi, S. Hayashi, and S. Kato, J. Chem. Phys. **126**, 144503 (2007).

[97] Y. Komeiji, T. Ishikawa, Y. Mochizuki, H. Yamataka, and T. Nakano, J. Comput. Chem. **30**, 40 (2009).

[98] According to Refs. 90 and 91, the solute thermal correction for the Menshutkin reaction is 13.1 and 7.5 kcal/mol for the free energy of activation and reaction, respectively.

[99] P. E. M. Lopes, B. Roux, and A. D. MacKerell, Theor. Chem. Acc. **124**, 11 (2009).

[100] K. W. Y. Abell and A. J. Kirby, Tetrahedron Lett. **27**, 1085 (1986).

[101] C. Lad, N. H. Williams, and R. Wolfenden, Proc. Natl. Acad. Sci. U.S.A. **100**, 5607 (2003).

[102] R. Wolfenden, Chem. Rev. **106**, 3379 (2006).

[103] R. B. Stockbridge and R. Wolfenden, J. Am. Chem. Soc. **131**, 18248 (2009).

[104] K. Nam, J. Gao, and D. M. York, J. Chem. Theory Comput. **1**, 2 (2005).

[105] T. Yamamoto, Chem. Phys. Lett. **500**, 263 (2010).

[106] S. C. L. Kamerlin, M. Haranczyk, and A. Warshel, Chem. Phys. Chem. **10**, 1125 (2009).

[107] J. P. Guthrie, J. Am. Chem. Soc. **99**, 3991 (1977).

[108] H. Nakano and T. Yamamoto, "Including charge penetration effects into the ESP derived partial charge operator" (unpublished).

[109] S. Iuchi, A. Morita, and S. Kato, J. Phys. Chem. B **106**, 3466 (2002).

[110] M. Isegawa and S. Kato, J. Chem. Theory Comput. **5**, 2809 (2009).

[111] T. Ishiyama and A. Morita, J. Chem. Phys. **131**, 244714 (2009).

[112] T. Ishiyama, V. V. Sokolov, and A. Morita, J. Chem. Phys. **134**, 024509 (2011).

[113] M. Isegawa and S. Kato, J. Chem. Phys. **127**, 244502 (2007).

[114] A. Morita, J. Comput. Chem. **23**, 1466 (2002).

[115] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), pp. 331–342.

[116] See supplementary material at http://dx.doi.org/10.1063/1.3699234 for additional comparison of PMF.