JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

A Modified QM/MM Hamiltonian with the Self-Consistent-Charge Density-Functional-Tight-Binding Theory for Highly Charged QM Regions

Guanhua Hou,[†] Xiao Zhu,[†,§] Marcus Elstner,[‡] and Qiang Cui*[,†]

[†]Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, 1101 University Ave., Madison, Wisconsin 53706, United States

[‡]Institute of Physical Chemistry, Karlsruhe Institute of Technology, Kaiserstr. 12, 76131 Karlsruhe, Germany

ABSTRACT: To improve the description of electrostatic interaction between QM and MM atoms when the QM is SCC-DFTB, we adopt a Klopman−Ohno (KO) functional form which considers the finite size of the QM and MM charge distributions. Compared to the original implementation that used a simple Coulombic interaction between QM Mulliken and MM point charges, the KO-based QM/MM scheme takes the charge penetration effect into consideration and therefore significantly improves the description of QM/MM interaction at short range, especially when the QM region is highly charged. To be consistent with the third-order formulation of SCC-DFTB, the Hubbard parameter in the KO functional is dependent on the QM charge. As a result, the effective size of the QM charge distribution naturally adjusts as the QM region undergoes chemical transformations, making the KO-based QM/MM scheme particularly attractive for describing chemical reactions in the condensed phase. Together with the van der Waals parameters for the QM atom, the KO-based QM/MM model introduces four parameters for each element type. They are fitted here based on microsolvation models of small solutes, focusing on negatively charged molecular ions, for elements O, C, H, and P with a specific version of SCC-DFTB (SCC-DFTBPR). Test calculations confirm that the KO-based QM/MM scheme significantly improves the interactions between QM and MM atoms over the original point charge based model, and it is transferable due to the small number of parameters. The new form of QM/MM Hamiltonian will greatly improve the applicability of SCC-DFTB based QM/MM methods to problems that involve highly charged QM regions, such as enzyme catalyzed phosphoryl transfers.

I. INTRODUCTION

Hybrid QM/MM simulations have become an important tool for the investigation of reaction mechanisms in complex environments, such as in water and biomolecules.[1−6] Numerous careful studies have indicated that, once carefully calibrated, QM/MM methods are able to produce reliable results for not only reaction free energy profiles but also other experimental observables, such as infrared spectra[7] and kinetic isotope effects.[8] Nevertheless, additional developments are needed to further increase the efficiency and robustness of QM/MM methods, especially for highly heterogeneous and flexible systems such as biomolecules.

In the past decade, a promising QM/MM approach based on an approximate density functional theory, the self-consistent-charge density-functional-tight-binding (SCC-DFTB) method,[9] has emerged. In terms of computational efficiency, SCC-DFTB is comparable to the widely used semiempirical methods such as AM1 and PM3, i.e., being 2−3 orders of magnitude faster than popular DFT methods. In terms of accuracy, fairly extensive benchmark calculations have indicated that it is particularly reliable for structural properties, while energetics are generally comparable to AM1 and PM3.[10−12] There are several recent developments of SCC-DFTB[13−15] for metal ions[16−19] and a few other elements that require d orbitals for a reliable description (e.g., phosphorus[20]). Although there are still many areas for improvement, due to its reasonable balance of efficiency and speed, SCC-DFTB has become an attractive choice as the QM method in many QM/MM simulations of biomolecules.[3,7,21−28]

We have successfully applied the SCC-DFTB based QM/MM simulations to a fairly broad set of enzyme problems. For reactions that involve highly charged species, such as the hydrolysis of phosphate monoesters, we have demonstrated that the conventional SCC-DFTB/MM approach tends to significantly overestimate the interaction between QM and MM groups, leading to considerable errors in the reaction free energy profiles.[20] This observation motivated us to re-examine the Hamiltonian used to describe the interaction between the QM and MM regions in hybrid QM/MM simulations.[29]

The total Hamiltonian for the molecular system under consideration in a hybrid QM/MM framework is generally given by

$$\hat{H} = \hat{H}^{\mathrm{QM}} + \hat{H}^{\mathrm{QM/MM}} + \hat{H}^{\mathrm{MM}} \tag{1}$$

where $\hat{H}^{\mathrm{QM/MM}}$ describes the interaction between the QM and MM atoms governed by $\hat{H}^{\mathrm{QM}}$ and $\hat{H}^{\mathrm{MM}}$, respectively. Typically, $\hat{H}^{\mathrm{QM/MM}}$ is written as the sum of electrostatic, van der Waals (vdW), and bonded components:

$$\hat{H}^{\mathrm{QM/MM}} = \hat{H}^{\mathrm{QM/MM}}_{\mathrm{vdW}} + \hat{H}^{\mathrm{QM/MM}}_{\mathrm{elec}} + \hat{H}^{\mathrm{QM/MM}}_{\mathrm{bonded}} \tag{2}$$

Received: July 25, 2012
Published: September 21, 2012

Here, $\hat{H}_{elec}^{QM/MM}$ describes the electrostatic interactions between the QM electrons and nuclei with the MM partial charges (or higher moments[30]), which are variable if a polarizable MM model is used; with a simple fixed-charge MM model, it is given by

$$\hat{H}_{elec}^{QM/MM} = \sum_{i\in QM}\sum_{A\in MM}\frac{-Q_A}{|\mathbf{r}_i - \mathbf{R}_A|} + \sum_{\alpha\in QM}\sum_{A\in MM}\frac{Z_\alpha Q_A}{|\mathbf{R}_\alpha - \mathbf{R}_A|} \quad (3)$$

where Q_A is the partial charge on the MM atom A and Z_α is the nuclear charge for QM atom α. $\hat{H}_{vdW}^{QM/MM}$ qualitatively describes the Pauli repulsion between the QM and MM atoms at short range and dispersion interactions at long range. In popular implementations, it is given as an empirical term that does not depend on the QM wave function; i.e.,

$$\hat{H}_{vdW}^{QM/MM} = U_{vdW}^{QM/MM} = \sum_{A\in MM}\sum_{\alpha\in QM}\varepsilon_{A\alpha}\left[\left(\frac{\sigma_{A\alpha}}{R_{A\alpha}}\right)^{12} - 2\left(\frac{\sigma_{A\alpha}}{R_{A\alpha}}\right)^{6}\right] \quad (4)$$

in which $\varepsilon_{A\alpha}$ and $\sigma_{A\alpha}$ are defined by the standard combination rules: $\varepsilon_{A\alpha} = (\varepsilon_A\varepsilon_\alpha)^{1/2}$ and $\sigma_{A\alpha} = 1/2(\sigma_A + \sigma_\alpha)$. The parameters for the QM atoms ε_α and σ_α are usually fitted empirically[31,32] based on atom or element type and not allowed to change during a chemical reaction.

It should be emphasized that, due mainly to the approximate nature of MM force fields, the physical meaning of "electrostatic" and "van der Waals" QM/MM interactions is not as rigorously defined as in perturbative theories for intermolecular interactions;[33] i.e., it is usually not possible to directly compare the "electrostatic" and "van der Waals" QM/MM interaction energies to results from energy decomposition analysis,[34,35] unless a very physical[36] and accurate MM model is used. Finally, the $\hat{H}_{bonded}^{QM/MM}$ term is required when the QM/MM partitioning is across a covalent bond, such as between the C_α and C_β of an amino acid.[37−39]

For reactions that occur in a polar environment, such as water and biomolecules, electrostatic interactions between the QM and MM atoms tend to dominate. Therefore, much consideration has been given to an accurate description of QM/MM electrostatics, both at short and long ranges. At short range, charge penetration effects[33] become important, and therefore treating the MM atom as a simple point charge tends to be problematic. Smearing the MM charges into a continuous distribution (e.g., a Gaussian) has been proposed by several authors and shown to be effective.[40,41] At long range, both Ewald based[42] (for periodic boundary condition) and reaction field based (for finite-size boundary condition) techniques have been adapted to describe QM/MM electrostatic interactions. Specifically for SCC-DFTB, we have focused on the treatment of long-range QM/MM interactions in the context of both Ewald[43] and the generalized solvent boundary potential[44,45] (GSBP) frameworks. In the current work, we examine the limitations of the originally implemented SCC-DFTB/MM[46] electrostatic interactions at short-range, which we believe are responsible for the significant overestimate of QM/MM interactions observed for systems that contain highly charged QM regions.[20] Specifically, we propose to revise the form of the SCC-DFTB/MM electrostatic Hamiltonian based on an approximation widely used in the semiempirical QM literature;[47] we then show that the revised form considerably improves the accuracy of SCC-DFTB/MM interactions, especially for systems that feature high charges (e.g., phosphate esters).

Although the magnitude of the QM/MM van der Waals interactions tends to be substantially smaller than electrostatics for polar systems, the QM van der Waals parameters have a significant impact on the spatial distribution of the MM atoms around the QM atoms and thus indirectly influence the magnitude of QM/MM electrostatics.[32] Therefore, although error cancellation occurs in many cases, a critical consideration of the QM(/MM) van der Waals parameters is likely important when the QM region undergoes significant changes, such as during a chemical reaction. This issue has been addressed in the pioneering work of York and co-workers, who have developed a charge-dependent van der Waals model for QM/MM interactions.[48] In the reported model, the van der Waals term is decoupled from the determination of the QM wave function/density and computed in a post-SCF fashion; this complicates the computation of gradients, which are required in molecular dynamics simulations. Therefore, it is still valuable to develop alternative models, and here we briefly discuss the issue in the context of SCC-DFTB/MM simulations.

The paper is organized as follows: In section II, we present our new SCC-DFTB/MM interaction model and summarize the relevant computational details for parametrization and benchmark calculations. In section III, we first present results for QM/MM interaction energies in simple cluster models and then demonstrate the value of our new model using phosphate monoester dianion hydrolysis reactions in solution. Finally, we draw several conclusions.

II. THEORY AND METHODS

A. A Simple but Effective Model for SCC-DFTB/MM Electrostatics. In the original implementation of SCC-DFTB/MM,[46] the QM atoms interact with the MM sites electrostatically in the form of Mulliken partial charges:

$$\hat{H}_{elec}^{QM/MM} = \sum_{A\in MM}\sum_{\alpha\in QM}\frac{Q_A\Delta q_\alpha}{|\mathbf{R_A} - \mathbf{R}_\alpha|} \quad (5)$$

where Q_A and Δq_B represent the MM partial charge and Mulliken partial charge, respectively. We note that although other definitions of charges[49] in SCC-DFTB can in principle be used instead of the simple Mulliken charges, important parameters in SCC-DFTB (e.g., repulsive potentials) were optimized within the Mulliken framework.

Treating both QM and MM atoms as point charges is clearly a major approximation, especially when they approach each other and charge penetration becomes important.[33] The observation that QM/MM interactions are significantly overestimated (as compared to full QM calculations) when the QM region is highly charged further supports this consideration. Therefore, it is imperative to seek an alternative formulation that is physical and applicable for all ranges of QM/MM distances.

One possible avenue starts with reviewing the formulation of SCC-DFTB. The second-order expansion of the total energy takes the form

$$E = \sum_{i}^{occ} \langle \Psi_i | \hat{H}^0 | \Psi_i \rangle + \frac{1}{2} \iint' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \left. \frac{\delta^2 E_{xc}}{\delta\rho\delta\rho'} \right|_{\rho_0} \right) \delta\rho\delta\rho' - \frac{1}{2} \iint' \frac{\rho_0' \rho_0}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0]\rho_0 + E_{cc} \quad (6)$$

With the monopole approximation and spherical charges,

$$\delta\rho \approx \sum_{\alpha} \delta\rho_\alpha \approx \Delta q_\alpha F_{00}^{\alpha} Y_{00} \quad (7)$$

the second-order term becomes

$$E^{2nd} = \frac{1}{2} \sum_{\alpha\beta} \Delta q_\alpha \Delta q_\beta \iint' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \left. \frac{\delta^2 E_{xc}}{\delta\rho\delta\rho'} \right|_{\rho_0} \right) F_{00}^{\alpha} F_{00}^{\beta} Y_{00}^2 \approx \frac{1}{2} \sum_{\alpha\beta} \Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta} \quad (8)$$

To obtain explicit expressions for $\gamma_{\alpha\beta}$, Elstner et al. used Slater-like charge distributions to evaluate the approximate expression of the second-order term

$$F_{00}^{\alpha} = \frac{\tau_\alpha}{8\pi} \exp(-\tau_\alpha |\mathbf{r} - \mathbf{R}_\alpha|) \quad (9)$$

in which the exponent is related to the Hubbard parameter (U_α)

$$\tau_\alpha = \frac{16}{5}\gamma_{\alpha\alpha} = \frac{16}{5}U_\alpha = \frac{32}{5}\eta_\alpha \quad (10)$$

where η_α is the chemical hardness parameter for element α. In other words, the Hubbard parameter is closely related to the effective size of the atom (charge distribution). In recent studies,[13−15] it has been shown that including the charge dependence of the Hubbard parameter, which effectively extends the DFTB energy to third order in density (charge) fluctuation, significantly improves properties such as proton affinities. Physically, this highlights that the size of an atom (charge distribution) changes considerably as it deprotonates.

In principle, we can adopt $\gamma_{\alpha\beta}$ for QM/MM interactions. However, its functional form is rather complex. As discussed in ref 13, $\gamma_{\alpha\beta}$ is closely related to other approximate expressions for two-center two-electron integrals developed in the NDDO literature.[47,50] For example, the Klopman−Ohno (KO) expression for the two-center two-electron integral, $\langle \mu_A \nu_A | \lambda_B \sigma_B \rangle$, is given as

$$\langle \mu_A \nu_A | \lambda_B \sigma_B \rangle = \sum_{i\in A} \sum_{j\in B} \frac{cQ_i Q_j}{\sqrt{R_{ij}^2 + \frac{1}{4}(d_i^{-1} + d_j^{-1})^2}} \quad (11)$$

where the summations are over the multipole charges ($Q_{i,j}$) used to mimic atomic orbitals and $d_{i,j}$ represents parameters determined on the basis of the condition that as the distance R_{AB} approaches zero the expression reduces to known one-center two-electron integrals. For the interactions between two spherical charges (s orbitals), the expression can be used by setting the $d_{i,j}$ as nothing but the corresponding Hubbard parameters.

Putting the two lines of thoughts together, it is evident that a physically motivated QM/MM Hamiltonian appropriate for SCC-DFTB is

$$\hat{H}_{elec,KO}^{QM/MM} = \sum_{\alpha\in QM} \sum_{A\in MM} \frac{\Delta q_\alpha Q_A}{\sqrt{R_{\alpha A}^2 + 0.25[1/U_\alpha(\Delta q_\alpha) + 1/U_A]^2}} \quad (12)$$

where we emphasize that $U_\alpha(\Delta q_\alpha)$ is dependent on the Mulliken charge of the QM atom; i.e., $U_\alpha(\Delta q_\alpha) = U_\alpha^0 + \Delta q_\alpha U_\alpha^d$, in which the Hubbard derivative with respect to atomic charge (U_α^d) is either calculated from atomic properties or fitted on the basis of properties of interest[14,15] (e.g., proton affinity[15]). The the MM "Hubbard" parameter (U_A) can be taken from atomic electronic structure calculations or treated as a parameter similar to the width of the "Gaussian blur" in the approach introduced by Brooks and co-workers,[40] or simply set to ∞ (the point charge limit).

Adopting eq 12 for the SCC-DFTB/MM electrostatics has several advantages over the simple point charge model. First, it properly considers the damping of the charge−charge interaction at a short QM/MM distance and therefore avoids overpolarization of the QM region in the presence of nearby MM atoms. In fact, in one of the first implementations of semiempirical QM/MM methods,[37] the QM/MM electrostatic terms are described using essentially the same approach, i.e., using approximate $\langle ss|ss \rangle$ two-electron integrals. Second, using the charge-dependent Hubbard parameter naturally considers the continuous variation of atomic (charge distribution) size during a chemical reaction. Although this does not replace including charge-dependence in the QM(/MM) van der Waals parameters (see below), it provides a more physical description of QM/MM electrostatics during a chemical transformation.

In this work, the KO functional form is slightly modified to introduce additional flexibility:

$$H_{elec,KO}^{QM/MM} = \sum_{\alpha\in QM} \sum_{A\in MM} \frac{\Delta q_\alpha Q_A}{\sqrt{R_{\alpha A}^2 + a_\alpha \left(\frac{1}{U_\alpha(\Delta q_\alpha)} + \frac{1}{U_A} \right)^2 e^{-b_\alpha R_{\alpha A}}}} = \sum_{\alpha\in QM} \sum_{A\in MM} \gamma_{KO} \Delta q_\alpha Q_A \quad (13)$$

To avoid overfitting, the parameters a_α and b_α are set to be based on element rather than atom type; thus our model only introduces two extra parameters for each element. With the modified QM/MM electrostatic Hamiltonian, the SCC-DFTB Kohn−Sham matrix element also needs to be modified correspondingly; e.g., with $U_{MM} = \infty$

Table 1. Optimized Parameters for Different QM/MM Interaction Schemes

	point-charge[a]		KO-∞[b]				KO-MM[c]			
	ε_α (kcal/mol)	σ_α (Å)	ε_α (kcal/mol)	σ_α (Å)	a_α –	b_α (Å^{-1})	ε_α (kcal/mol)	σ_α (Å)	a_α –	b_α (Å^{-1})
O	−0.18	1.92	−0.03	2.05	0.068	0.017	−0.06	1.88	0.042	0.021
C	−0.06	2.15	−0.07	2.11	0.046	0.059	−0.05	2.15	0.026	0.069
P	−1.23	2.36	−0.52	2.39	0.060	0.001	−0.26	2.42	0.054	0.001
H	−0.02	0.82	−0.04	0.76	0.211	0.055	−0.02	0.81	0.066	0.053

[a]The original point-charge-based SCC-DFTB/MM Hamiltonian[46] with newly optimized QM van der Waals parameters. [b]The Hubbard parameter for the MM atoms in the KO scheme (see eq 13) is set to ∞. [c]The Hubbard parameters for the MM atoms in the KO scheme are taken from the calculated values for the respective element.[15]

$$H_{\mu\upsilon} = H^0_{\mu\upsilon} + \frac{1}{2}S_{\mu\upsilon}\sum_{B\in \mathrm{QM}}(\gamma_{CB}+\gamma_{DB})\Delta q_B + \frac{1}{2}S_{\mu\upsilon}\sum_{A\in \mathrm{MM}}\left[\gamma_{\mathrm{KO},AC}+\gamma_{\mathrm{KO},AD} + \frac{\Delta q_C \gamma^3_{\mathrm{KO},AC}U^d_C a_C\, e^{-b_C R_{AC}}}{U^3_C(\Delta q_C)} + \frac{\Delta q_D \gamma^3_{\mathrm{KO},AD}U^d_D a_D\, e^{-b_D R_{AD}}}{U^3_D(\Delta q_D)}\right]Q_A + \frac{1}{2}S_{\mu\upsilon}\sum_{A\in \mathrm{QM}}\frac{\partial U_A}{\partial q_A}\Delta q_A^2 \quad (14)$$

where $\mu \in C$, $\upsilon \in D$.

As discussed in the Introduction, the QM(/MM) van der Waals interactions in principle should also be made charge dependent; both effective atomic radius and polarizability for the reactive site(s) are expected to change during a chemical transformation. In the framework of DFT, this can be accomplished by noting the relationship between atomic charge, effective atomic size, and atomic polarizability.[14,51−54] For example, since the Hubbard parameter is directly related to the chemical hardness (eq 10), including charge dependence in the Hubbard parameter would also introduce charge-dependence into the chemical hardness, which is inversely related to atom size.[14,51−54] Therefore, it is possible to use this relationship to make the QM van der Waals radii charge dependent. Similar arguments can be made for the atomic polarizability,[55] which determines the effective well depth for the QM(/MM) van der Waals terms; this is similar in spirit to the environment (density)-dependent dispersion coefficients introduced by several authors in the literature.[56,57] In practical terms, however, introducing those charge dependences into the QM van der Waals parameters in the framework of existing QM/MM programs would require considerable modification of the SCC-DFTB matrix elements and the SCF procedures; whether such an increase in the computational cost is worthwhile for practical applications requires systematic analysis in the future. In the current work, we show that adopting the KO scheme for the QM/MM electrostatics along with a single set of optimized element-type dependent QM van der Waals parameters already leads to significant improvement over the original SCC-DFTB/MM model[46] for the description of highly charged systems in polar environments.

B. Parameterization of the New KO Models. To summarize, our new SCC-DFTB/MM model Hamiltonian contains four parameters for each element: a_α and b_α in eq 13 for QM/MM electrostatics and ε_α and σ_α for QM/MM van der Waals interactions (see Table 1). In principle, the MM Hubbard parameters can also be optimized to allow additional flexibility. In this work, we test two models: simply set the MM Hubbard parameters to ∞ (referred to as the KO-∞ model) or use the calculated Hubbard from atoms[15] (referred to as the KO-MM model).

For parametrization of the KO models, small molecule−water interaction energy is used as the target property. Since we are mainly interested in condensed phase processes, the training and test sets contain small clusters that involve solutes interacting with multiple water molecules. Test calculations indicate that including multiple water molecules is essential to a successful parametrization; including only pairwise models as in previous studies[32] does not capture the complexity of interactions in the condensed phase and therefore does not lead to as transferrable parameters. The training set includes 23 molecules (see Table 2) that mimic protein side chains and phosphate species with various charge states; we focus on negatively charged solutes here because their charge distribution tends to be more diffuse, and therefore it is essential to describe them with a finite size. Each solute in the training set is first solvated with a water sphere of 25 Å radius, then 50 ps MD at 300 K is carried out with the solute treated as SCC-DFTB and water treated as TIP3P.[58] The original QM/MM Hamiltonian[46] and QM van der Waals parameters[32] are used for the sampling, and test calculations show that the optimized parameters are not very sensitive to this choice. Ten snapshots are taken from each trajectory, and water molecules beyond 4 Å from the solute are deleted, leaving about 15 water molecules surrounding the solute. The binding energies between the solute and water molecules at the full SCC-DFTB level are used as reference values for the parametrization; higher-level QM calculations can be used as references, but we leave this to future work following the completion of parametrizing DFTB3 for more element types. In this study, which mainly serves to demonstrate the value of the revised SCC-DFTB/MM Hamiltonian, we use a specific version of SCC-DFTB parametrized for phosphate hydrolysis (SCC-DFTBPR[20]). The parameters are optimized using a genetic algorithm (GA)[59] in which the "fitness" (ξ) is defined as the inverse of a weighted sum of difference between binding energies determined from full SCC-DFTB and SCC-DFTB/MM calculation:

$$\xi^{-1} = \frac{\sum_{i=1} w_i[\Delta E^b_i(\mathrm{SCC}) - \Delta E^b_i(\mathrm{QM/MM})]^2}{\sum_{i=1} w_i} \quad (15)$$

where i is the index of species in the training set, and the sum is over all molecules in the training set. During optimization, a micro-GA technique with a population of 10 chromosomes is

Table 2. Error (in kcal/mol) Analysis of Binding Energies for the Training Set[a]

solute	E_{SCC}	unsigned error[b] KO-∞	KO-MM	point charge
propane	−0.6	4.3	3.8	3.4
isobutene	−1.6	5.4	5.0	4.6
butane	−1.1	5.4	4.7	4.5
toluene	−3.5	5.6	4.5	4.5
4-cresol	−10.5	2.6	2.5	2.4
methanol	−9.2	2.1	1.9	2.5
ethanol	−9.0	1.5	1.3	2.0
acetaldehyde	−6.7	1.3	1.1	1.8
methylacetate	−9.3	1.9	1.6	2.8
acetic acid	−7.6	1.5	1.2	2.7
propanic acid	−15.8	4.4	3.7	5.6
dimethyl ether	−4.2	1.6	1.3	2.0
methylphosphate	−21.9	7.4	5.3	8.6
dimethylphosphate	−15.9	2.8	3.3	5.2
acetate (−1)	−78.4	3.0	2.8	6.8
propanate (−1)	−88.3	2.5	2.7	2.1
4-cresol (−1)	−83.2	6.5	5.8	7.0
methoxide (−1)	−94.0	5.3	3.7	7.4
ethoxide (−1)	−95.5	6.7	4.1	10.0
hydroxide (−1)	−68.6	8.8	6.9	5.4
methylphosphate (−1)	−84.1	6.4	5.3	5.5
dimethylphosphate (−1)	−79.5	3.6	1.8	5.5
methyl phosphate (−2)	−249.6	7.6	5.6	8.6
	error analysis[c]			
MUE		4.3	3.3	4.8
MSE		−0.5	−0.8	−0.8

[a]The reference values (E_{SCC}) are calculated with SCC-DFTBPR; for the different QM/MM schemes (see the footnote of Table 1 for notation), negative values indicate that the strength of interaction is overestimated relative to the reference. [b]The unsigned error is averaged over 10 snapshots for each solute. [c]MUE, mean unsigned error; MSE, mean signed error.

Table 3. Error (in kcal/mol) Analysis of Binding Energies for the Test Set[a]

solute	E_{SCC}	unsigned error KO-∞	KO-MM	point charge
methane	−0.9	0.9	0.9	0.8
phenol	−6.3	4.8	4.2	4.7
propanol	−10.8	4.2	3.7	4.5
formaldehyde	−3.1	1.2	1.0	1.6
formic acid	−21.6	2.0	1.6	3.2
trimethyl phosphate	−21.0	7.2	6.4	12.0
formate (−1)	−93.6	5.6	1.7	10.8
benzoate (−1)	−94.1	10.7	5.3	15.1
propanoate (−1)	−96.5	6.8	3.1	10.9
dihydrogen phosphate (−1)	−105.4	2.6	2.6	9.0
methyl phenyl phosphate (−1)	−110.5	9.0	5.8	17.0
hydrogen phosphate (−2)	−271.5	2.1	5.9	17.1
	error analysis			
MUE		4.7	3.5	8.9
MSE		−4.2	−2.0	−8.8

[a]See Table 2 for notations.

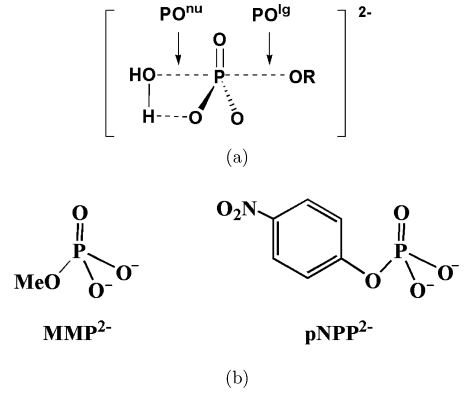

Figure 1. The phosphate monoester dianions hydrolysis reactions studied in this work.

allowed to operate for 500 generations with uniform cross-overs; see ref 60 for detailed descriptions and recommendations for GA options.

Finally, to test the transferability of the parametrization, a test set of 12 different clusters (see Table 3) is constructed in a similar fashion. The KO-∞ and KO-MM models are compared to the original QM/MM interaction scheme;[46] for a fair comparison, we also reparametrize the QM van der Waals parameters for the original QM/MM interaction scheme with the same training set used for parametrizing the KO models.

C. Potential of Mean Force (PMF) Simulations for Aqueous Phosphate Hydrolysis Reactions. To explicitly evaluate the performance of different QM/MM interaction schemes for condensed phase reactions, we study the aqueous hydrolysis reactions of two phosphate monoesters, methyl monophosphate^{2-} (MMP^{2-}) and p-nitrophenyl phosphate^{2-} ($pNPP^{2-}$) (see Figure 1), with a water molecule as the nucleophile. These reactions serve as ideal benchmarks because there are extensive experimental[61] and computational[62,63] studies. In addition, these phosphate monoesters are the typical substrates of phosphatases;[64,65] thus the results also provide important references for future enzyme studies.

The solute (MMP^{2-} or $pNPP^{2-}$) is solvated by the standard protocol of superimposing the molecule with a water droplet of 25 Å radius and removing water molecules within 2.8 Å from any solute atoms.[66] Water molecules are described using the TIP3P model[58] with modified van der Waals parameters for hydrogen. The QM region includes the solute and the nucleophilic water. The generalized solvent boundary potential (GSBP)[44,45] is used to treat long-range electrostatics in MD simulations. To be consistent with the GSBP protocol, the extended electrostatic model[67] is used to treat the electrostatic interactions among inner-region atoms for which interactions beyond 12 Å are treated with multipolar expansions, including the dipolar and quadrupolar terms. The deformable boundary forces[68] are added in the boundary region to constrain water molecules within the sphere. An additional weak GEO type of potential is added to the QM region to keep it in the center of the water sphere. An angular constraining potential is added to the nucleophilic water, the phosphate atom and the leaving group oxygen to favor the "in line attack" configuration. All bonds involving hydrogen in MM water are constrained using the SHAKE algorithm,[69] and the time step is set to 1 fs.

The 2D PMF calculations are carried out for the aqueous reactions. The system is energy minimized and slowly heated to 300 K and equilibrated for 50 ps. The reaction coordinate is defined as PO^{lg}–PO^{nu} and HO^{wat}–HO^{po}. Here, $O^{lg/nu}$ indicates the oxygen in the leaving group/nucleophile, and O^{po} indicates

the phosphate oxygen that the proton transfers to. The umbrella sampling approach[70] is used to constrain the system along the reaction coordinates. In total, more than 250 windows are used for each PMF, and 50 ps simulations are performed for each window. The first 10 ps trajectories are discarded and only the last 40 ps are used for data analysis. Convergence of the PMF is monitored by examining the overlap of reaction coordinate distributions sampled in different windows and by evaluating the effect of leaving out segments of trajectories. The probability distributions are combined together by the weighted histogram analysis method (WHAM)[71] to obtain the PMF along the reaction coordinates.

As additional benchmarks that focus on the quality of the QM method rather than other technical details such as QM/MM coupling and sampling, we use a previously developed implicit solvent model[72] to study these aqueous reactions of phosphate monoesters. In this model, the solute radii are dependent on the charge distribution, and therefore the model is particularly useful for studying solution reactions that involve highly charged species. Our previous benchmark calculations suggest that the method has comparable accuracy to the SM6 model for a fairly broad range of neutral and charged solutes,[73] while being much more efficient (due to the use of SCC-DFTB) and having only a small number of parameters. The energetics of the reactions are studied using adiabatic mapping, and the reaction coordinates are the same as those in QM/MM PMF simulations. Each point in the 2D adiabatic map is obtained by starting the constrained optimization from several different initial structures and taking the lowest energy value. The initial grid size is 0.2 Å due to the large number of structures that need to be optimized. Later, a finer grid size (0.1 Å) is used to scan the transition state (TS) region and to locate the TS structure. Finally, frequency calculations are carried out to confirm the nature of the stationary points and to compute the vibrational entropy and zero point energies to obtain approximate activation free energy; although using a harmonic approximation to estimate activation entropy is known to be of limited accuracy, previous studies of phosphate diester hydrolysis found that activation entropy does not differ much between different diesters.[74]

To account for intrinsic errors of SCC-DFTBPR energies, we explore corrections based on gas phase single-point energy calculations with MP2/6-311++G** at SCC-DFTBPR geometries. As discussed in the literature,[75] such a simple correction may not always improve the energetics for semiempirical methods given the errors in geometry; however, our previous tests[20,27,72] indicated that this correction scheme appears useful for SCC-DFTBPR since the method gives fairly reliable structures, even for transition states.

III. RESULTS AND DISCUSSION

A. Solute−Water Binding Energies in Training and Test Sets. Since performance in the condensed phase is of main concern, we adopt the cluster type of models to explicitly sample multibody interactions. As shown in Table 2 for the training set, the total interactions between the "first solvation shell" water and the solute cover a great range, from a few kilocalories per mole for aliphatic molecules to almost 250 kcal/mol for a methyl phosphate dianion. With the original point-charge based QM/MM electrostatic Hamiltonian, the error from full SCC-DFTB results is not overwhelmingly large once the QM van der Waals parameters are optimized (based on element type). The mean unsigned error (MUE) is 4.8 kcal/mol for *total* solute−water interactions; even for charged solutes, such as ethoxide and methyl phosphate, the error is about 10 kcal/mol, which is merely a few percent of the total interaction. We note that using the van der Waals parameters optimized in ref 32 based on solute−water dimers leads to substantially larger errors, with a MUE of 14.1 kcal/mol (data not shown). This is partly because the van der Waals parameters in ref 32 were optimized for the second-order SCC-DFTB, while the SCC-DFTBPR[20] used herein includes diagonal third-order contributions; more importantly, however, using the cluster models better captures the multibody interactions featured in the condensed phase.

For the training set, the two KO models give slightly better results; the KO-∞ model has a MUE of 4.3 kcal/mol while KO-MM has a value of 3.3 kcal/mol. For all three QM/MM models, the interaction between nonpolar solute and nearby water has notable errors of a few kilocalories per mole, which are comparable or even larger in magnitude than the total interaction at the full QM level. This is because the penalty function is largely dictated by charged molecules, which have much stronger interactions with water. Since for most applications the solute tends to be substantially charged, this is not expected to be a major issue in practice. Nevertheless, a more balanced description for broader classes of solutes is desirable, and we leave this to future work in which we also plan to use higher level QM results as the reference for parametrization.

To examine the transferability of the parametrized models, we construct a test set that contains 12 molecules of biological relevance but not covered in the training set (see Table 3). The performance of the KO models is quite comparable to that observed for the training set; the MUE is 4.7 kcal/mol for KO-∞ and 3.5 kcal/mol for KO-MM. With the original point-charge based SCC-DFTB/MM Hamiltonian,[46] however, the errors become quite large for several molecules, leading to a substantially larger MUE of 8.9 kcal/mol. Therefore, it seems that the KO models are substantially more robust than the original SCC-DFTB/MM Hamiltonian.

Finally, to examine the performance of the models for not only stable structures but also transition states, we have selected structures for 10 model phosphate hydrolysis reactions in RNA from the QCRNA database established by the York group.[76] These include 16 stable states and 24 transition states. Solvated cluster models for these structures are constructed in a similar fashion to that for the training and test sets, and we again compare the solute−water interaction calculated by different SCC-DFTBPR/MM models to full SCC-DFTBPR results. As shown in Table 4, the errors are rather large with the original SCC-DFTB/MM Hamiltonian; the values are often about 10% of the total interaction energy. Fortunately, it appears that the errors are fairly comparable for the stable and transition states; the MUEs are 14.2 and 17.6 kcal/mol, respectively. This explains why SCC-DFTBPR/MM has been found to give fairly reasonable results for several phosphoryl transfer reactions in solution[77] and enzymes.[24,27,78] With the KO-∞ model, the errors are smaller than the original SCC-DFTB/MM model by almost a factor of 2, with a MUE of 7.1 kcal/mol. Most impressively, the KO-MM model continues to perform well. The MUE is 4.3 kcal/mol, which is rather similar to both the training and test sets; the MUEs for stable states and transition states are also comparable, 3.5 and 4.8 kcal/mol, respectively. Therefore, the KO-MM model appears to be transferrable and

Table 4. Energetics Benchmark Calculations for Different QM/MM Interaction Schemes Based on 10 Phosphate Reactions from the QCRNA Database[a]

reactions	conf.	E_{SCC}	errors		
			point-charge	KO-∞	KO-MM
CH3O...P(O)(O)(OH)(OCH3)	ts12	−255.0	28.0	14.0	2.4
HO...P(O)(O)(OH)(OCH3)	ts12	−252.3	25.8	11.8	4.2
HO...P(O)(OH)(OH)(OCH3)	ts12	−96.7	12.0	3.5	5.8
	min2	−100.5	10.5	3.1	3.6
	ts23	−98.2	10.6	3.1	3.4
HOH...P(O)(O)(OCH3)(OCH3)	min1	−99.3	11.2	3.6	5.2
	ts12	−102.1	12.0	3.4	6.6
	min2	−98.7	12.0	3.9	7.3
	ts23	−100.2	10.8	2.8	6.2
	min3	−109.1	10.9	3.1	2.5
	ts34	−96.5	13.3	5.1	2.2
	min4	−98.4	11.3	3.8	2.0
	ts45	−98.6	14.4	6.4	1.9
	min1	−94.5	11.6	4.0	2.3
HO...P(O)(O)(OCH3)(OCH3)	ts12	−257.6	23.0	7.6	2.1
	min2	−274.0	26.3	11.4	2.3
	ts23	−267.2	24.5	10.4	1.7
CH3O...P(O)(O)(OCH3)(OCH3)	ts12	−244.1	32.1	18.9	2.8
CH3O...P(O)(OH)(OH)(OCH3)	ts12	−102.0	13.1	4.8	1.9
	min2	−93.3	13.2	4.9	3.9
	ts23	−99.6	14.4	6.0	2.6
CH3O...P(O)(OH)(OCH3)(OCH3)	ts12	−94.8	16.6	9.1	3.3
	min2	−98.1	13.2	5.2	6.0
	ts23	−107.4	16.1	7.6	4.0
	min3	−96.4	13.0	5.5	12.1
	ts34	−103.5	14.4	6.5	2.4
CH3O...P(O)(OCH3)(OCH3)(OCH3)	ts12	−97.9	18.9	10.5	2.6
	min2	−106.2	16.5	8.7	3.6
	ts23	−103.5	17.2	9.4	6.4
	min3	−99.4	16.9	9.0	2.6
	ts34	−92.0	17.4	10.0	4.5
	min4	−103.3	14.6	7.1	3.4
	ts45	−102.8	19.5	11.3	3.9
HO...P(O)(OCH3)(OCH3)(OCH3)	ts12	−113.8	19.3	2.7	6.9
	min2	−94.6	15.4	7.2	5.4
	ts23	−96.9	16.5	8.5	6.1
	min3	−101.0	14.6	6.3	5.6
	ts34	−96.6	13.8	6.0	6.7
	min4	−98.3	16.5	8.0	4.9
	ts45	−96.5	17.6	9.4	7.4
	error analysis[b]				
overall performance	MUE		16.2	7.1	4.3
stable states performance	MUE		14.2	5.9	3.5
transition states performance	MUE		17.6	7.9	4.8

[a]The unsigned error is averaged over 10 snapshots for each solute. [b]MUE: mean unsigned error.

gives robust QM/MM interactions for a fairly broad set of polar and charged solutes.

B. PMF for Phosphate Monoester Reactions. *1. MMP^{2-} Hydrolysis Reaction.* MMP^{2-} is a simple phosphate monoester, and its solution reaction has been extensively studied using experimental and computational methods. In aqueous solution, the nucleophile has been determined as a water molecule, and the experimental free energy barrier is 44.3 kcal/mol at 298 K (converted from rate constants with transition state theory).[61] The barrier was well reproduced as 47 kcal/mol at 312 K using the B3LYP/COSMO model.[63] The calculated transition state structure indicates that the water first transfers a proton to MMP^{2-} to become a hydroxide, which in turn attacks the protonated phosphate monoester. The $P-O^{lg}$ and $P-O^{nu}$ bond lengths are 1.8 and 2.0 Å, respectively, in the rate-limiting transition state.

Before studying this reaction using different QM/MM interaction models, it is useful to establish the intrinsic error of the QM method. For this purpose, we use a recently developed implicit solvent model that combines the SCC-DFTB method and Poisson−Boltzmann (PB) equation with a set of charge dependent radii.[72] In the calculated potential energy surface shown in Figure 2a, the reactant state corresponds to the bottom left corner while the product state corresponds to the upper right corner. The first step involves an exothermic proton transfer from water to MMP^{2-}, followed by the nucleophilic attack; this is consistent with the results from previous studies.[63] The calculated rate-limiting transition state has the reaction coordinate of $PO^{lg}-PO^{nu}$ as 0.0 Å and the proton transfer coordinate as 1.1 Å. The PO^{lg} and PO^{nu} bond lengths are both 1.95 Å, also close to previous computational results. Rescanning the TS region by a finer grid leads to a reaction barrier of 30.5 kcal/mol. After including vibrational entropy and zero point energy corrections at 300 K, the free energy barrier is 39.5 kcal/mol; further including MP2 single point correction (see Table 5) leads to the best estimated barrier of 45.7 kcal/mol. As shown in Figure 2c, the overall potential energy surface after the MP2 correction remains similar to that at the SCC-DFTBPR/PB level, although the latter systematically underestimates the relative energy compared with infinitely separated reactants, especially for the upper left corner, which corresponds to the proton transfer step. This is not unexpected because significant imbalance of proton affinities for phosphate and nonphosphate species remains in SCC-DFTBPR.[20] Our recent studies indicate that this deficiency is largely removed when the complete third-order contributions are included[15] (Gaus and Cui, work in progress).

With QM/MM simulations, the results depend on the model for the QM/MM interactions. With the original point-charge based SCC-DFTB/MM Hamiltonian, the free energy barrier is grossly overestimated compared to both experimental and SCC-DFTBPR/PB results; the value of 55 kcal/mol is about 10 kcal/mol higher than the experimental result. With the KO models, the barriers are 41 (KO-∞) and 40 kcal/mol (KO-MM), respectively. These are close to the SCC-DFTBPR/PB results (without the MP2 correction). The transition state region calculated by the KO schemes is at reaction coordinate $PO^{lg}-PO^{nu}$ slightly less than 0 Å and the proton transfer coordinate at 1.2 Å. The averaged PO^{lg} and PO^{nu} bond lengths are 1.94 and 2.04 Å, respectively, similar to those from previous calculations[63] and the SCC-DFTBPR/PB model.

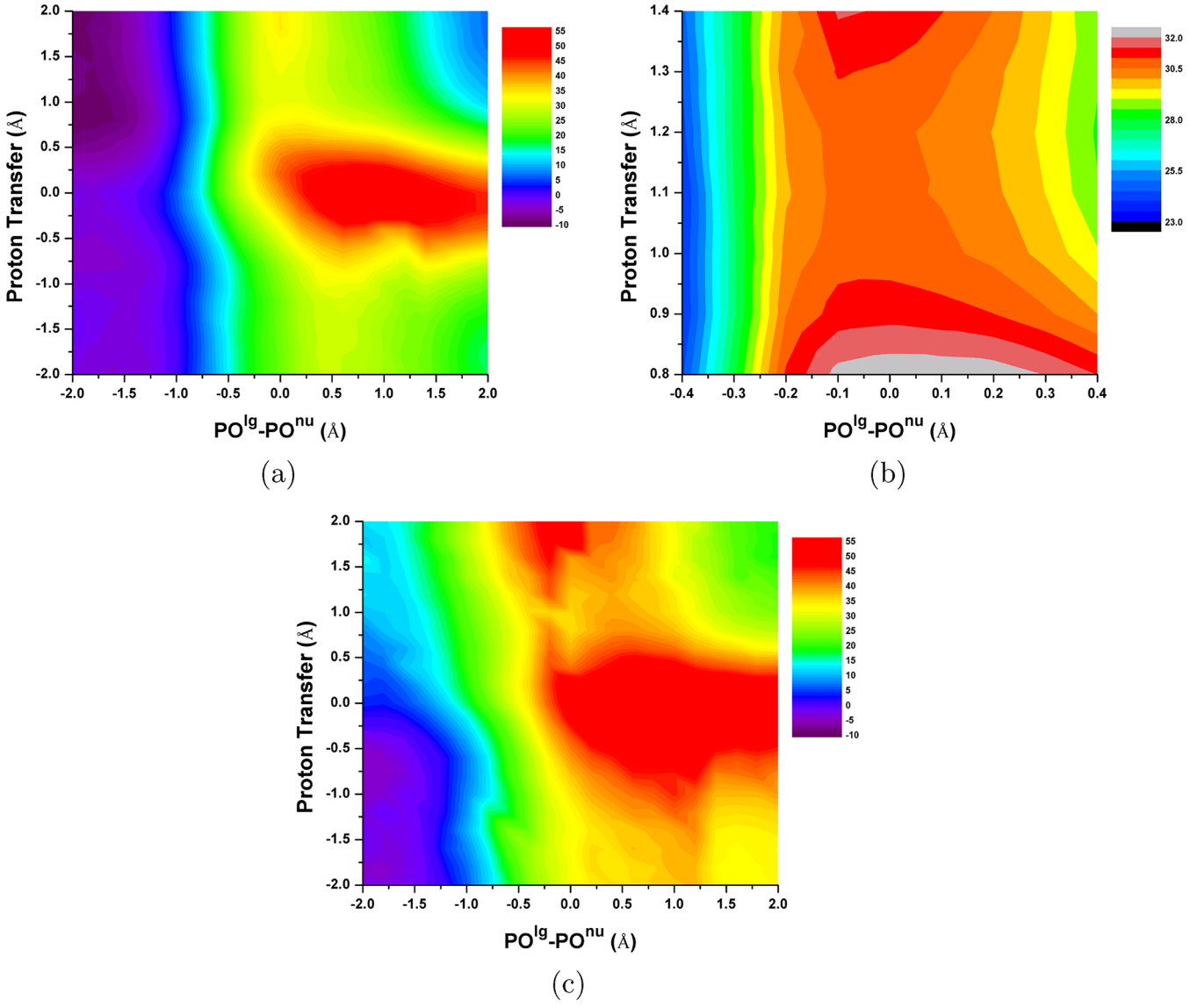

Figure 2. Potential energy surface (PES, in kcal/mol) of MMP^{2-} hydrolysis reaction in water. (a) 2D PES by SCC-DFTBPR/PB;[72] (b) 2D PES of the TS region with a finer grid size by SCC-DFTBPR/PB; (c) 2D PES that includes the gas-phase MP2/6-311++G** single point energy corrections over SCC-DFTBPR.

It is worth noting that the first proton transfer step is exothermic with the original SCC-DFTBPR/MM model but endothermic by about 10 kcal/mol with the KO models (see Figure.3). Warshel and co-workers studied this step[62] using the MP2/LD method and obtained an endothermicity of ~9 kcal/mol. As noted above, the SCC-DFTBPR/PB model has notable errors in this region of the potential energy surface (Figure 2) due to proton affinity errors in SCC-DFTBPR. These errors are partially canceled by the remaining deficiencies in QM/MM interactions when the KO models are used, leading to fortuitous agreement with MP2/LD results. The substantial errors in the QM/MM interactions by the KO models for this region of the potential energy surface are not unexpected because the interaction between the nucleophile (hydroxide) and nearby water are difficult to describe with a QM/MM model (see Table 2); the significant charge transfer between the hydroxide and nearby water in principle requires a full QM description.

2. $pNPP^{2-}$ Hydrolysis Reaction. In addition to MMP^{2-}, we also study the hydrolysis of another phosphate monoester, $pNPP^{2-}$, which has a rather different ester group. $pNPP^{2-}$ is an important substrate for phosphatase; therefore the aqueous results provide an important reference for enzyme studies. Similar to the MMP^{2-} case, we first use SCC-DFTBPR/PB to estimate the inherent error in the QM method. The overall potential energy landscape (Figure 4a) is similar to that of MMP^{2-}. The phosphate nonbridging oxygen first abstracts a proton from the water, and then the hydroxide acts as the nucleophile to attack the monoanionic phosphate. The free energy barrier is calculated to be 29.3 kcal/mol after adding entropic effects and zero point energy corrections; the value becomes 27.0 kcal/mol after adding MP2 single point energy corrections. The experimental value is 31.8 kcal/mol at 298 K, slightly higher than our results. The transition state structure (Figure 4b) has the reaction coordinate PO^{lg}–PO^{nu} as −0.3 Å and the proton transfer coordinate as 1.0 Å. The PO^{lg} bond length is 1.95 Å, similar to that for MMP^{2-}. However, the PO^{nu} bond length increases to 2.26 Å, and the overall transition state structure becomes looser than that for MMP^{2-}. These observations are consistent with the trend found in previous

dx.doi.org/10.1021/ct300649f | J. Chem. Theory Comput. 2012, 8, 4293−4304

Table 5. Free Energy Barriers (kcal/mol) of Phosphate Monoester Hydrolysis in Water Estimated by Different Methods[a]

solute	method	$\Delta G^{\ddagger}$
MMP^{2-}	exptl.[b]	44.3 (298K)
	MP2/LD[c]	43 (312K)
	SCC-DFTBPR/PB[d]	39.5/45.7
	point-charge QM/MM	55
	KO-∞	41
	KO-MM	40
$pNPP^{2-}$	exptl.[e]	31.8 (298K)
	SCC-DFTBPR/PB[d]	29.3/27.0
	KO-∞	33
	KO-MM	32

[a]All results are for 300 K unless noted otherwise. [b]Results taken from ref 61. [c]Results taken from ref 63. [d]The number before the slash is an SCC-DFTBPR/PB result that includes entropic and ZPE corrections; the number after the slash further includes the gas phase MP2/6-311++G** single point corrections. [e]Results taken from ref 80.

studies that the transition state changes from associative to dissociative in nature as the pK_a of the leaving group decreases.[63] SCC-DFTBPR appears to be able to capture the substituent effect on the hydrolysis reaction.

With QM/MM simulations, the calculations are problematic with the original point-charge based SCC-DFTBPR/MM Hamiltonian; the grossly overestimated QM/MM interactions overpolarize the QM region, leading to highly distorted structures and energetics. With the KO models, QM/MM PMF simulations are able to converge without much difficulty. Both lead to free energy barriers of 33 kcal/mol, similar to the SCC-DFTBPR/PB results; if the MP2 gas-phase correction estimated for SCC-DFTBPR/PB (~2 kcal/mol) is transferrable to the QM/MM results, the QM/MM barriers would be in good agreement with the experimental value. The rate-limiting transition state structure from the QM/MM simulations is also consistent with the SCC-DFTBPR/PB results. The average PO^{lg} and PO^{nu} bond lengths are 1.84 and 2.14 Å, respectively, also in qualitative agreement with previous theoretical studies of Warshel and co-workers.[63]

IV. CONCLUDING REMARKS

QM/MM methods have become indispensable in the analysis of reaction mechanisms of complex systems. Once specific QM and MM models are chosen for a given application, the remaining factor that determines the accuracy of the calculation is the coupling between the QM and MM models. In the current work, we analyze and improve the QM/MM coupling Hamiltonian when SCC-DFTB is the QM method. In the original implementation,[46] the electrostatic interaction between SCC-DFTB and MM atoms is computed using a simple point charge model in which the QM charge density is represented as Mulliken charges. This becomes a poor approximation when the MM and QM atoms approach each other, making the original SCC-DFTB/MM method less reliable when the QM region is highly charged. Motivated by models used to

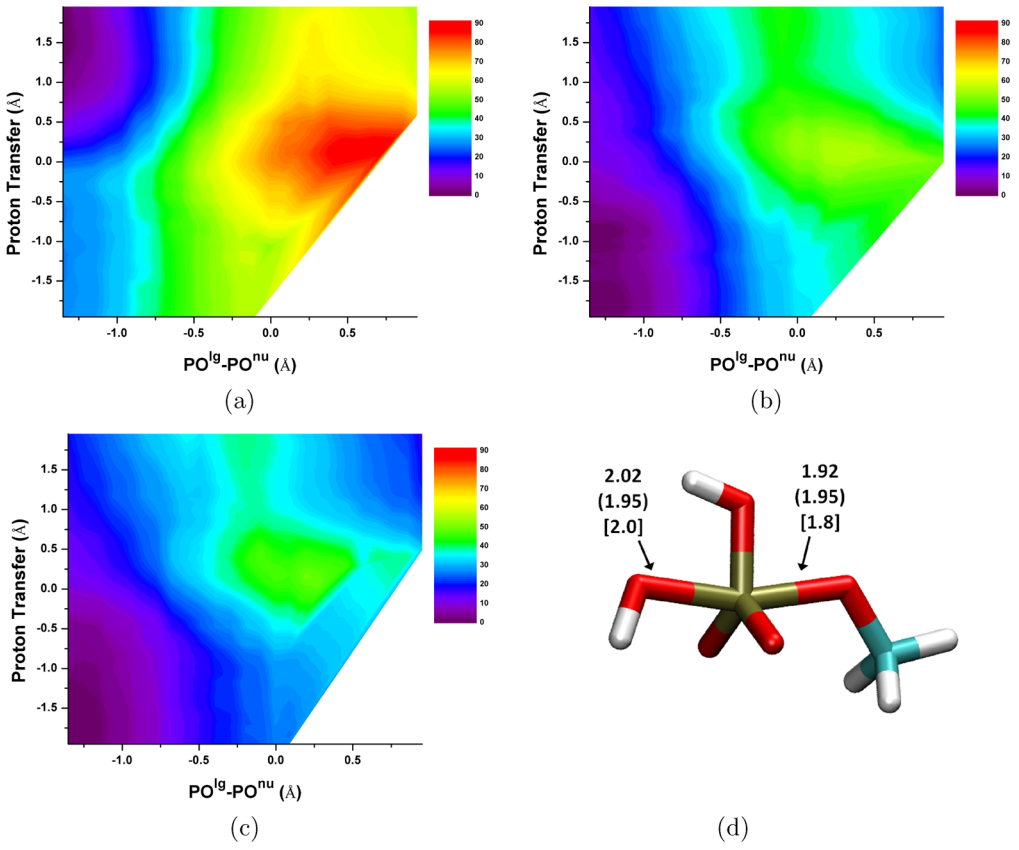

Figure 3. 2D PMF (in kcal/mol) of MMP^{2-} hydrolysis reaction in water calculated with different SCC-DFTB/MM schemes. (a) The original QM/MM scheme[46] with newly optimized QM van der Waals parameters (summarized in Table 1). (b) The KO-∞ scheme. (c) The KO-MM scheme. (d) The transition state structure. Numbers without parentheses are calculated with KO-MM; those with parentheses are calculated with SCC-DFTBPR/PB; those with brackets are taken from ref 63.

dx.doi.org/10.1021/ct300649f | J. Chem. Theory Comput. 2012, 8, 4293−4304

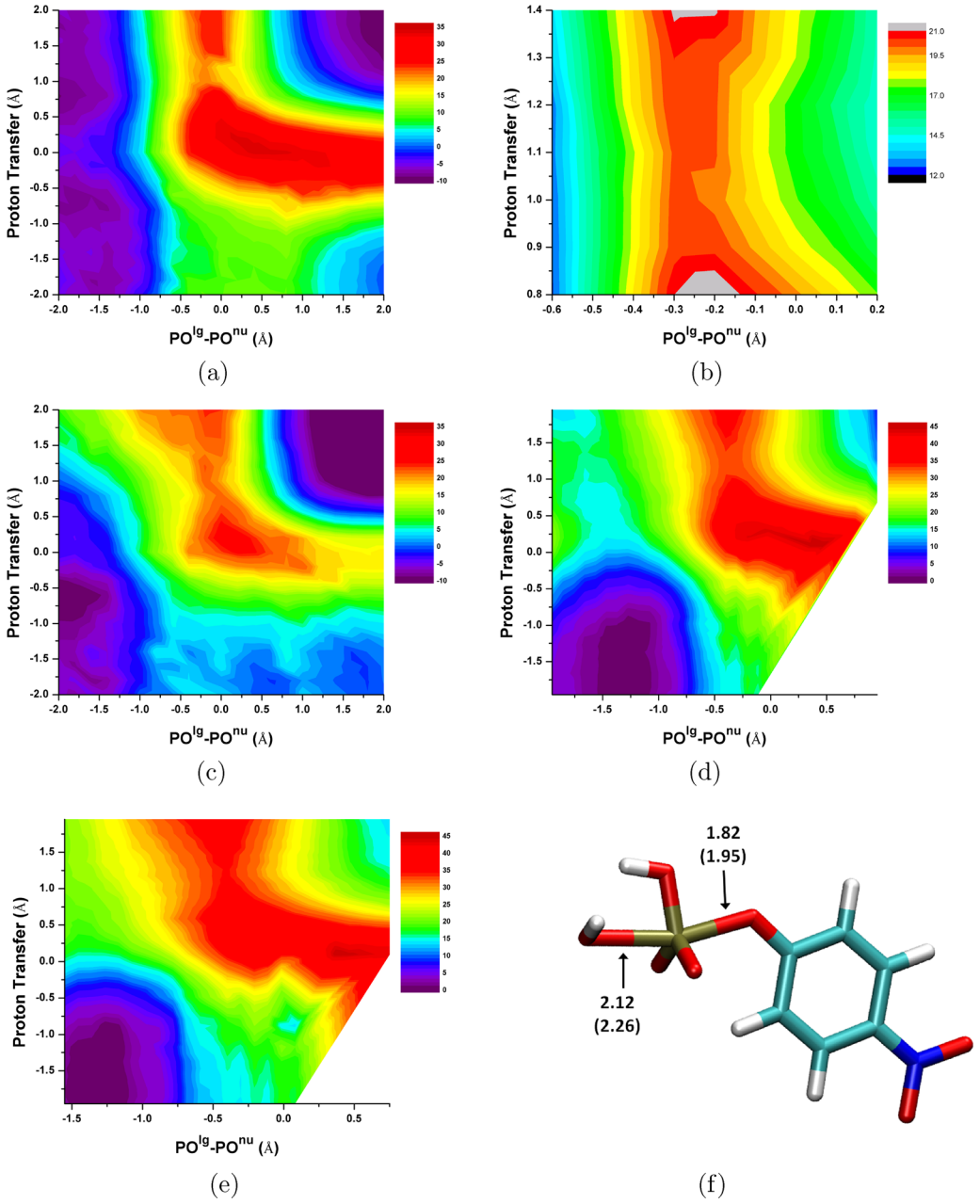

Figure 4. 2D potential energy surface (PES) and potential of mean force (PMF) of the $pNPP^{2-}$ hydrolysis reaction in water calculated by SCC-DFTBPR/PB and SCC-DFTBPR/MM simulations, respectively. (a) 2D PES by SCC-DFTBPR/PB. (b) 2D PES for the transition state region with a finer grid size by SCC-DFTBPR/PB. (c) 2D PES including the gas phase MP2/6-311++G** single point energy corrections over SCC-DFTBPR. (d) 2D PMF by the KO-∞ scheme. (e) 2D PMF by the KO-MM scheme. (f) The transition state structure. Numbers without parentheses are calculated with the KO-MM scheme; those with parentheses are calculated with SCC-DFTBPR/PB.

approximate two electron integrals in semiempirical QM methods, we propose to revise the electrostatic part of the SCC-DFTB/MM interaction to adopt a Klopman−Ohno (KO) functional form, which mimics the interaction between two spherical charges. Similar to the Gaussian-blur models[40,41] for ab initio QM/MM, the KO model takes charge penetration effects into consideration and therefore significantly improves the description of charge−charge interactions at short range; at the same time, the computation remains efficient. To be consistent with the third-order formulation of SCC-DFTB,[15] the Hubbard parameter in the KO functional is dependent on the QM charge. In this way, the effective size of the QM charge distribution naturally adjusts as the QM region undergoes chemical transformations, making the KO based QM/MM scheme particularly attractive for describing chemical reactions in the condensed phase.

The KO based QM/MM scheme introduces two parameters for each element; together with the QM van der Waals parameters, there are only four QM/MM parameters for each element. As a proof of concept, we have optimized these parameters for O, C, H, and P using a specific version of SCC-DFTB (SCC-DFTBPR). Test calculations using microsolvation clusters indicate that the KO based QM/MM scheme significantly improves the interactions between QM and MM atoms over the original point-charge implementation.[46] Since the number of parameters is small, the parametrized model

appears highly transferrable, as demonstrated by test calculations on molecules in the QCRNA database, which are rather different from the ones used for fitting the parameters. Finally, we apply the KO based QM/MM scheme to study the hydrolysis of two phosphate monoesters, MMP^{2-} and $pNPP^{2-}$. These were challenging to study using the original point-charge based SCC-DFTB/MM model, while the KO based SCC-DFTBPR/MM calculations led to results consistent with experimental data and previous theoretical calculations using approximate solvent models.

As future work, we will develop QM/MM-KO parameters for more elements following the complete parametrization of DFTB3,[15] which is more robust and generally more applicable than SCC-DFTBPR. For the new parametrization, solute−solvent interactions calculated at a higher level of QM method will be used as a reference, rather than using SCC-DFTBPR as done here. Along this line, it is worth noting that SCC-DFTB is a minimal basis set method and therefore expected to have only limited polarizability;[79] as a result, to be fully consistent with high-level QM/MM calculations, it is likely that additional improvements such as including a response density contribution[48] are necessary. Finally, it is worth exploring the practical benefit of better describing the van der Waals component of QM/MM coupling by making the QM atomic radii and polarizabilities fully consistent with the QM charge distribution. For reactions that involve highly charged species in a polar environment, it is likely that the electrostatic component dominates, although the contribution from the van der Waals component may still be non-negligible.

AUTHOR INFORMATION

Corresponding Author

*E-mail: cui@chem.wisc.edu.

Present Address

[§]HPC Application group, Texas Advanced Computing Center (TACC), 10100 Burnet Road (R8700), Austin, Texas 78758.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work is supported in part by the NIH (R01-GM084028), NSF (CHE-0957285), and by the WARF of UW—Madison. Computational resources from the National Center for Supercomputing Applications at the University of Illinois and the Center for High Throughput Computing (CHTC) at UW—Madison are greatly appreciated; computations are also supported in part by the National Science Foundation through a major instrumentation grant (CHE-0840494).

REFERENCES

(1) Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389−427.

(2) Gao, J. L.; Ma, S. H.; Major, D. T.; Nam, K.; Pu, J. Z.; Truhlar, D. G. *Chem. Rev.* **2006**, *106*, 3188−3209.

(3) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, H.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458−6469.

(4) Hu, H.; Yang, W. T. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573−601.

(5) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(6) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *J. Phys. Chem. B* **2009**, *113*, 1253−1272.

(7) Goyal, P.; Ghosh, N.; Phatak, P.; Clemens, M.; Gaus, M.; Elstner, M.; Cui, Q. *J. Am. Chem. Soc.* **2011**, *133*, 14981−14997.

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

(9) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260−7268.

(10) Kruger, T.; Elstner, M.; Schiffels, P.; Frauenheim, T. *J. Chem. Phys.* **2005**, *122*, 114110.

(11) Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. *J. Phys. Chem. A* **2006**, *110*, 13551−13559.

(12) Otte, N.; Scholten, M.; Thiel, W. *J. Phys. Chem. A* **2007**, *111*, 5751−5755.

(13) Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 5614−5621.

(14) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 10861−10873.

(15) Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2011**, *7*, 931−948.

(16) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565−581.

(17) Cai, Z.; Lopez, P.; Reimers, J. R.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 5743−5750.

(18) Zheng, G. S.; Witek, H. A.; Bobadova-Parvanova, P.; Irle, S.; Musaev, D. G.; Prabhakar, R.; Morokuma, K. *J. Chem. Theory Comput.* **2007**, *3*, 1349−1367.

(19) Moreira, N. H.; Dolgonos, G.; Aradi, B.; da Roasa, A. L.; Frauenheim, T. *J. Chem. Theory Comput.* **2009**, *5*, 605−614.

(20) Yang, Y.; Yu, H.; York, D.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2008**, *4*, 2067−2084.

(21) Elstner, M.; Frauenheim, T.; Suhai, S. *J. Mol. Struct.* **2003**, *632*, 29−41.

(22) Elstner, M. *Theor. Chem. Acc.* **2006**, *116*, 316−325.

(23) Bondar, A. N.; Smith, J. C.; Elstner, M. *Theor. Chem. Acc.* **2010**, *125*, 353−363.

(24) Yang, Y.; Yu, H.; Cui, Q. *J. Mol. Biol.* **2008**, *381*, 1407−1420.

(25) Riccardi, D.; Koenig, P.; Guo, H.; Cui, Q. *Biochemistry* **2008**, *47*, 2369−2378.

(26) Xu, D. G.; Guo, H. *J. Am. Chem. Soc.* **2009**, *131*, 9780−9788.

(27) Hou, G. H.; Cui, Q. *J. Am. Chem. Soc.* **2012**, *134*, 229−246.

(28) Chakravorty, D. K.; Wang, B.; Lee, C. W.; Giedroc, D. P.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **2012**, *134*, 3367−3376.

(29) Riccardi, D.; Zhu, X.; Goyal, P.; Yang, S.; Hou, G.; Cui, Q. *Sci. China: Chem.* **2012**, *55*, 3−18.

(30) Ponder, J. W.; Wu, C. J.; Ren, P. Y.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.

(31) Freindorf, M.; Gao, J. L. *J. Comput. Chem.* **1996**, *17*, 386−395.

(32) Riccardi, D.; Li, G. H.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467−6478.

(33) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, U.K., 1997.

(34) Morokuma, K. *Acc. Chem. Res.* **1977**, *10*, 294−300.

(35) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887−1930.

(36) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piguemal, J.-P. *J. Chem. Theory Comput.* **2007**, *3*, 1960−1986.

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

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

(39) Gao, J. L.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714−4721.

(40) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodoscek, M.; Brooks, B. R. *J. Chem. Phys.* **2002**, *117*, 10534−10547.

(41) Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2010**, *6*, 3330−3342.

(42) Nam, K.; Gao, J. L.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2−13.
(43) Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 17715−17733.
(44) Im, W.; Bernèche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924−2937.
(45) Schaefer, P.; Riccardi, D.; Cui, Q. *J. Chem. Phys.* **2005**, *123*, 014905.
(46) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569−585.
(47) Thiel, W. *Adv. Chem. Phys.* **1996**, *93*, 703−757.
(48) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2007**, *127*, 194101.
(49) Kalinowski, J. A.; Lesyng, B.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2545−2549.
(50) Kolb, M.; Thiel, W. *J. Comput. Chem.* **1993**, *14*, 775−789.
(51) Politzer, P.; Parr, R.; Murphy, D. *J. Chem. Phys.* **1983**, *79*, 3859−3861.
(52) Pearson, R. *Inorg. Chem.* **1988**, *27*, 734−740.
(53) Ghosh, D.; Biswas, R. *Int. J. Mol. Sci.* **2003**, *4*, 379−407.
(54) Politzer, P.; Murray, J.; Lane, P. *J. Comput. Chem.* **2003**, *24*, 505−511.
(55) Politzer, P. *J. Chem. Phys.* **1987**, *86*, 1072−1073.
(56) Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 073005.
(57) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *123*, 154101.
(58) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926−935.
(59) Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addison-Wesley: Reading, MA, 1989.
(60) Carroll, D. L. http://cuaerospace.com/carroll/ga.html (accessed Jan 2009).
(61) Lad, C.; Williams, N. H.; Wolfenden, R. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 5607−5610.
(62) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 719−734.
(63) Klhn, M.; Rosta, E.; Warshel, A. *J. Am. Chem. Soc.* **2006**, *128*, 15310−15323.
(64) O'Brie, P.; Herschlag, D. *Biochemistry* **2002**, *41*, 3207−3225.
(65) Zalatan, J.; Fenn, T.; Brunger, A.; Herschlag, D. *Biochemistry* **2006**, *45*, 9788−9803.
(66) Boorks, C.; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312−6325.
(67) Steinbach, P.; Brooks, B. *J. Comput. Chem.* **1994**, *15*, 667−683.
(68) Brooks, C. L.; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312−6325.
(69) Rychaert, J.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327−341.
(70) Torrie, G. M.; Valleau, J. P. *J. Chem. Phys.* **1977**, *23*, 187−199.
(71) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(72) Hou, G. H.; Zhu, X.; Cui, Q. *J. Chem. Theory Comput.* **2010**, *6*, 2303−2314.
(73) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133−1152.
(74) Rosta, E.; Kamerlin, S. C. L.; Warshel, A. *Biochemistry* **2008**, *47*, 3725−3735.
(75) Cui, Q.; Karplus, M. *J. Phys. Chem. B* **2002**, *106*, 1768−1798.
(76) Giese, T. J.; Gregersen, B. A.; Liu, Y.; Nam, K.; Mayaan, E.; Moser, A.; Range, K.; Faza, A. N.; Lopez, C. S.; de Lera, A. R.; Schaftenaar, G.; Lopez, X.; Lee, T. S.; Karypis, G.; York, D. M. *J. Mol. Graphics Modell.* **2006**, *25*, 423−433.
(77) Yang, Y.; Cui, Q. *J. Phys. Chem. B* **2009**, *113*, 4930−4933.
(78) Yang, Y.; Cui, Q. *J. Phys. Chem. A* **2009**, *113*, 12439−12446.
(79) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2005**, *123*, 164108.
(80) Lassila, J.; Zalatan, J.; Herschlag, D. *Annu. Rev. Biochem.* **2011**, *80*, 669−702.