JCTC Journal of Chemical Theory and Computation

An Implicit Solvent Model for SCC-DFTB with Charge-Dependent Radii

Guanhua Hou, Xiao Zhu, and Qiang Cui*

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

Received April 5, 2010

Abstract: Motivated by the need to rapidly explore the potential energy surface of chemical reactions that involve highly charged species, we have developed an implicit solvent model for approximate density functional theory, SCC-DFTB. The solvation free energy is calculated using a popular model that employs Poisson−Boltzmann for electrostatics and a surface-area term for nonpolar contributions. To balance the treatment of species with different charge distributions, we make the atomic radii that define the dielectric boundary and solute cavity depend on the solute charge distribution. Specifically, the atomic radii are assumed to be linearly dependent on the Mulliken charges and solved self-consistently together with the solute electronic structure. Benchmark calculations indicate that the model leads to solvation free energies of comparable accuracy to the SM6 model (especially for ions), which requires much more expensive DFT calculations. With analytical first derivatives and favorable computational speed, the SCC-DFTB-based solvation model can be effectively used, in conjunction with high-level QM calculations, to explore the mechanism of solution reactions. This is illustrated with a brief analysis of the hydrolysis of monomethyl monophosphate ester (MMP) and trimethyl monophosphate ester (TMP). Possible future improvements are also briefly discussed.

I. Introduction

Many chemical reactions take place in solution, so a proper description for the solvation effect is one of the most important challenges in computational chemistry. Although major progress has been made in QM/MM[1−6] and *ab initio* molecular dynamics[7] methods in which the solvent molecules are treated explicitly, the cost of such calculations is still rather high. Therefore, implicit solvent models remain an attractive choice for many studies. In the context of studying chemical reactions, the most commonly used framework for treating solvent implicitly is the dielectric continuum model,[8,9] in which the solvent is replaced by a homogeneous dielectric medium. More sophisticated treatments based on integral equations have also been developed, such as (MC)SCF-RISM,[10] although they tend to be computationally more expensive than dielectric continuum models.

Over the past few decades, many different dielectric solvent models have been developed in the quantum chemistry community, such as the self-consistent reaction field (SCRF) model,[11,12] polarized continuum model (PCM),[13−25] generalized Born (GB) model,[26−32] conductor-like screening model (COSMO),[33−38] and the Langevin dipole model.[39] For the application to chemical reactions involving large solutes, there are two practical issues. First, the computational cost of implicit solvent model calculations is still rather high, especially when used with a high-level QM method. Therefore, it is fairly common to perform gas-phase optimization for stationary points and then carry out single-point energy calculations in solution using a dielectric continuum model. This can be problematic when there is significant difference between the gas phase and solution potential energy landscape,[40] a scenario which is not uncommon when the solute is highly charged or zwitterionic. The second problem is that most implicit solvent models employ a set of fixed atomic radii to define the solvent/solute dielectric boundary, and these radii are typically preoptimized on the basis of the experimental solvation free energies of a set of small molecules[8,9,41] and are therefore limited by the diversity of

* Corresponding author e-mail: cui@chem.wisc.edu.

10.1021/ct1001818
Published on Web 07/07/2010

the training set. The use of fixed atomic radii causes additional errors in the application to chemical reactions, as the description of transition states is rarely included during parametrization; moreover, the atom type for atoms[41] explicitly involved in the reaction is often ill-defined in a transition state. Methods have been developed in which the molecular cavity is determined on the basis of the electron isodensity surface,[42,43] although an optimal value for the electron density cutoff is not always straightforward to determine.[44]

Motivated by these considerations, we have implemented a dielectric solvent model for an approximate density functional theory, the self-consistent-charge density-functional-tight-binding (SCC-DFTB) method.[45] SCC-DFTB is an approximation to density functional theory (DFT) based on a second-order expansion of the DFT total energy around a reference electron density. With respect to computational efficiency, SCC-DFTB is comparable to 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.[46−48] With recent developments of SCC-DFTB[49,50] for metal ions[51−54] and a few other elements that require d orbitals for a reliable description (e.g., phosphorus[55]), an effective implicit solvent model for SCC-DFTB will be very useful and complementary to existing models based on other semiempirical methods.[26,56,57] Our model takes advantage of the finite difference Poisson−Boltzmann approach[58,59] implemented in CHARMM[60] and has analytic first derivatives.[61] This makes it possible to perform geometry optimization, reaction path searchers, and vibrational frequency calculations (based on the numerical finite difference of first derivatives).

Our main aim is to use SCC-DFTB for quickly exploring minimum energy paths for reactions in solution and then refine selected results on the basis of higher level theories. To be able to describe transition state and stable structures on equal footing, it is desirable to determine the atomic radii in a self-consistent fashion on the basis of the electronic structure of the solute. The simple model we have adopted is to make the atomic radii depend on the Mulliken charges, which are fundamental to SCC-DFTB[45] and are solved self-consistently via an iterative procedure (see the Methods). A similar idea was explored in the context of an implicit solvent model for PM3.[62] More recently, as this work was in progress, charge-dependent radii have been developed for a DFT-based COSMO approach,[63,64] and much improved results (solvation free energies and chemical reactions) compared to fixed-radii models have been reported for small ions.

We have developed two sets of solvation radii parameters for SCC-DFTB. The first set is for the standard second-order SCC-DFTB[45] with parameters for C, H, O, and N. We recommend using this set for general applications to molecules consisting of these elements. The second set is for SCC-DFTBPR,[55] which is a specific version parametrized for the phosphate hydrolysis reaction and includes third-order on-site terms for C, H, O, and P; this set can be useful for studying phosphate hydrolysis reactions, although we caution that SCC-DFTBPR has been parametrized mainly for monoanionic phosphates and a limited set of hydrolysis reactions. Two rather large training sets for solvation free energy with an emphasis on biorelated molecules (including 103 and 57 solutes for SCC-DFTB and SCC-DFTBPR, respectively) are used to develop the solvation radii parameters. Calculations on two additional sets of test molecules show that the performance for neutral and charged species is rather well balanced, and the error is comparable to the SM6 model,[31] which is more sophisticated yet also much more expensive computationally. To illustrate the applicability of our model to chemical reactions in solution, we briefly study the hydrolysis of monomethyl monophosphate ester (MMP) and trimethyl monophosphate ester (TMP). The results from the current implicit solvent model are generally consistent with previous *ab initio* calculations in conjunction with PCM[65,66] or the Langevin dipole solvation models,[67] as well as with our explicit solvent simulations using SCC-DFTBPR/TIP3P.[55] Compared to the latter, however, the significant overstabilization of the zwitterionic intermediate is avoided, which highlights the complementary value of implicit solvent models to explicit solvent methods for studying reactions that involve highly charged species.

The paper is organized as follows: In section II, we summarize the key theoretical foundation for our implicit solvent model for SCC-DFTB; details for the parametrization and benchmark calculations are also included. In section III, we present results and discussions of the parametrization and benchmark data, including the overall performance for both the training and test sets of molecules, and results for the hydrolysis of MMP/TMP. Finally, we summarize our findings in section IV.

II. Methods

A. SCC-DFTB. Here, we briefly recall the basic elements of SCC-DFTB[45,51] that are important to the development of an implicit solvent model. The SCC-DFTB approach is based on a second-order expansion of the DFT total energy around a reference density, ρ_0,

$$E = \sum_i^{occ} \langle \Psi_i | \hat{H}^0 | \Psi_i \rangle + \frac{1}{2} \int\int \left(\frac{1}{|\vec{r} - \vec{r}'|} + \frac{\delta^2 E_{xc}}{\delta\rho\delta\rho'}|_{\rho_0} \right) \delta\rho\delta\rho' - \frac{1}{2} \int\int \frac{\rho'_0 \rho_0}{|\vec{r} - \vec{r}'|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0]\rho_0 + E_{cc} \quad (1)$$

where $\hat{H}^0 = \hat{H}[\rho_0]$ is the effective Kohn−Sham Hamiltonian evaluated at the reference density ρ_0, and the Ψ_i's are the Kohn−Sham orbitals. E_{xc} and V_{xc} are the exchange-correlation energy and potential, respectively, and E_{cc} is the core−core repulsion energy. With a minimal basis set, a monopole approximation for the second-order term and the two-center approximation to the integrals, the SCC-DFTB total energy is given in the following form:

$$E = \sum_{i\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0 + \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_\alpha \Delta q_\beta + \frac{1}{2} \sum_{\alpha\beta} U[R_{\alpha\beta}; \rho_0^\alpha, \rho_0^\beta] \quad (2)$$

where $c^i_{\mu/\nu}$ represents orbital coefficients, $\Delta q_{\alpha/\beta}$ represents the Mulliken charges on atom α/β, and $\gamma_{\alpha\beta}$ is the approximate second-order kernel derived on the basis of two interacting spherical charges. The last pairwise summation gives the so-called repulsive potential term, which is the core−core repulsion plus double counting terms and is defined relative to infinitely separated atomic species.

As discussed in our recent work,[2,49,50] it was found that further including the third-order contribution can substantially improve calculated proton affinity; for a set of biologically relevant small molecules, significant improvements were observed even with only the on-site terms included. The corresponding expression for the SCC-DFTB total energy is

$$E = \sum_{i\mu\nu} c^i_\mu c^i_\nu H^0_{\mu\nu} + \frac{1}{2}\sum_{\alpha\beta}\gamma_{\alpha\beta}\Delta q_\alpha \Delta q_\beta + \frac{1}{2}\sum_{\alpha\beta} U[R_{\alpha\beta};\rho^\alpha_0,\rho^\beta_0] + \frac{1}{6}\sum_\alpha U^d_\alpha \Delta q^3_\alpha \quad (3)$$

where U^d_α is the derivative of the Hubbard parameter of atom α with respect to atomic charge. For the development of SCC-DFTBPR for phosphorus-containing systems,[55] we found it was useful to adopt an empirical Gaussian functional form for the Hubbard charge derivative, i.e.

$$U^d_\alpha(q) = U^d_{0\alpha} + D_0 \exp[-\Gamma_0(\Delta q_\alpha - Q_0)^2] \quad (4)$$

where the charge-independent parameter ($U^d_{0\alpha}$) is dependent on the element type, whereas the three parameters associated with the Gaussian (D_0, Γ_0, Q_0) are taken to be independent of the element type to minimize the number of parameters.

B. The Solvation Model Based on Surface Area and Poisson−Boltzmann. The implicit solvent framework that we adapt is based on the popular formulation[68] that includes a surface-area-dependent nonpolar component and an electrostatic component:

$$\Delta G_{sol} = \Delta G_{np} + \Delta G_{elec} \quad (5)$$

where

$$\Delta G_{np} = \gamma S \quad (6)$$

Here, S is the solvent accessible surface area (SASA), which is dependent on atomic radii,[69] and γ is a phenomenological surface tension coefficient.

The electrostatic solvation free energy ΔG_{elec} for a given charge distribution $\rho(\mathbf{r})$ is generally given by

$$\Delta G_{elec} = \frac{1}{2}\int\int d\mathbf{r}\, d\mathbf{r}'\, \rho(\mathbf{r})\, G(\mathbf{r},\mathbf{r}')\, \rho(\mathbf{r}') \quad (7)$$

where 1/2 reflects the linearity of the dielectric medium[70] and the reaction field Green's function $G(\mathbf{r},\mathbf{r}')$ corresponds to the reaction field potential at $\mathbf{r}$ due to a unit charge at $\mathbf{r}'$:[71]

$$\phi_{rf}(\mathbf{r}) = \int d\mathbf{r}'\, G(\mathbf{r},\mathbf{r}')\, \rho(\mathbf{r}') \quad (8)$$

For a set of point charges, $\rho(\mathbf{r}) = {}_\alpha q_\alpha \delta(\mathbf{r} - \mathbf{r}_\alpha)$, ΔG_{elec} is simplified to

$$\Delta G_{elec} = \frac{1}{2}\sum_\alpha q_\alpha \phi_{rf}(\mathbf{r}_\alpha) \quad (9)$$

The reaction-field potential $\phi_{rf}(\mathbf{r})$ is obtained by subtracting a reference electrostatic potential computed in a vacuum, $\phi_v(\mathbf{r})$, from the electrostatic potential computed in the dielectric solvent medium, $\phi_s(\mathbf{r})$. The electrostatic potentials are determined as solutions of the (linearized) Poisson−Boltzmann (PB) equation:[59,72]

$$\nabla\cdot[\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})] - \bar{\kappa}^2(\mathbf{r})\,\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \quad (10)$$

with the appropriate dielectric boundary ($\varepsilon(\mathbf{r})$) and charge distributions in finite difference (FD) form using iterative numerical techniques. The solution yields the electrostatic potential at every grid point, and the total electrostatic solvation free energy is given by

$$\Delta G_{elec} = \frac{1}{2}\sum_i q_i(\phi_{s,i} - \phi_{v,i}) \quad (11)$$

where q_i and ϕ_i are the charge and calculated potential at the ith gridpoint, for the cases of a vacuum (v) and a solution (s).

In SCC-DFTB, ΔG_{elec} in eq 7 is also simplified by the fact that the charge (electrons plus nuclei) density is represented by a collection of atom-centered Mulliken charges:[45,73]

$$\rho(\mathbf{r}) = \sum_\alpha \Delta q_\alpha \delta(\mathbf{r} - \mathbf{R}_\alpha) \quad (12)$$

where Δq_α is the Mulliken charge of atom α. Thus, calculating ΔG_{elec} is a straightforward extension of the classical expression

$$\begin{aligned}\Delta G_{elec} &= \frac{1}{2}\int\int d\mathbf{r}\, d\mathbf{r}'\rho(\mathbf{r})\, G(\mathbf{r},\mathbf{r}')\, \rho(\mathbf{r}') \\ &= \frac{1}{2}\int d\mathbf{r}\, \rho(\mathbf{r})\, \phi_{rf}(\mathbf{r}) \\ &= \frac{1}{2}\sum_\alpha \Delta q_\alpha\, \phi_{rf}(\mathbf{R}_\alpha)\end{aligned} \quad (13)$$

Using variational principles, the solvation contribution to the total solute energy leads to additional terms in the SCC-DFTB matrix elements during SCF iterations:

$$\frac{1}{2}S_{\mu\nu}[\phi_{rf}(\mathbf{R_C}) + \phi_{rf}(\mathbf{R_D})]\mu \in C, \nu \in D \quad (14)$$

where μ and ν run over a minimal set of localized pseudoatomic Slater orbitals located on atoms C and D, respectively, and $S_{\mu\nu}$ is the overlap integral associated with the two basis functions.

Additional analytical gradient components from the solvation are calculated on the basis of the finite difference force proposed by Im et al.[61] They used a continuous, spline-based dielectric boundary, which has been shown to give accurate and numerically stable forces for PB calculations. The total solvation force acting on atom α is given by

$$\begin{aligned}\mathbf{F}_\alpha^{\mathrm{sol}} &= -\frac{\partial \Delta G_{\mathrm{sol}}}{\partial \mathbf{R}_\alpha} \\ &= -\frac{\partial \Delta G_{\mathrm{elec}}}{\partial \mathbf{R}_\alpha} - \frac{\partial \Delta G_{\mathrm{np}}}{\partial \mathbf{R}_\alpha} \\ &= \mathbf{F}_\alpha^{\mathrm{RF}} + \mathbf{F}_\alpha^{\mathrm{DB}} + \mathbf{F}_\alpha^{\mathrm{IB}} + \mathbf{F}_\alpha^{\mathrm{NP}}\end{aligned} \tag{15}$$

This method calculated the electrostatic solvation force as a sum of individual terms:[61] the reaction field force ($\mathbf{F}_\alpha^{\mathrm{RF}}$) arising from the variation of atomic positions assuming the dielectric boundary remains constant, the dielectric boundary force ($\mathbf{F}_\alpha^{\mathrm{DB}}$) caused by the spatial variations of the dielectric function $\varepsilon(\mathbf{r})$ from the solvent to the solute interior, and the ionic boundary force ($\mathbf{F}_\alpha^{\mathrm{IB}}$) resulting from spatial variations of the modified Debye−Hückel screening factor $\bar{\kappa}(\mathbf{r})$. In SCC-DFTB/PB approach, for the atom α located at position $\mathbf{R}_\alpha$, the three terms in the limit of infinitesimal grid spacing are

$$\begin{aligned}\mathbf{F}_\alpha^{\mathrm{RF}} &= -\int_V \mathrm{d}\mathbf{r}\left[(\phi_{\mathrm{s}} - \phi_{\mathrm{v}})\frac{\partial \Delta q_\alpha}{\partial \mathbf{R}_\alpha}\right] \\ \mathbf{F}_\alpha^{\mathrm{DB}} &= -\frac{1}{8\pi}\int_V \mathrm{d}\mathbf{r}\phi_{\mathrm{s}}\nabla\cdot\left[\left(\frac{\partial \varepsilon}{\partial \mathbf{R}_\alpha} + \frac{\partial \varepsilon}{\partial \Delta q_\alpha}\frac{\partial \Delta q_\alpha}{\partial \mathbf{R}_\alpha}\right)\nabla\phi_{\mathrm{s}}\right] \\ \mathbf{F}_\alpha^{\mathrm{IB}} &= \frac{1}{8\pi}\int_V \mathrm{d}\mathbf{r}(\phi_{\mathrm{s}})^2\frac{\partial \bar{\kappa}^2}{\partial \mathbf{R}_\alpha}\end{aligned} \tag{16}$$

Calculations for the derivative of the converged Mulliken charge, dielectric function, and modified Debye−Hückel screening factor have been discussed in previous studies;[61] note that the derivatives for the Mulliken charge with respect to the orbital coefficients are not needed for the reaction field force ($\mathbf{F}_\alpha^{\mathrm{RF}}$) because the reaction field contribution to the solute energy is included in a variational fashion.[74] As preliminary tests indicate, the contribution from the second term in $\mathbf{F}_\alpha^{\mathrm{DB}}$ is rather small; therefore, we omit it to simplify calculation (i.e., to avoid solving the coupled-perturbed KS equations[74] for the derivative of the MO coefficients).

C. Charge-Dependent Radii Scheme. To establish a simple relationship between the dielectric boundary and the electronic structure of the solute, we take the atomic radius of a solute atom α to be linearly dependent on its Mulliken charge, Δq_α

$$R_\alpha = A_{i(\alpha)} + B_{i(\alpha)}\Delta q_\alpha \tag{17}$$

where $A_{i(\alpha)}$ and $B_{i(\alpha)}$ are element-type-dependent parameters that need to be determined on the basis of a training set (see below). Higher-order polynomials have also been tested, although no systematic improvement in the results is observed.

Since the atomic radii have an impact on the solvation free energy and therefore on the solute wave function and the Mulliken charges, R_α and Δq_α need to be determined self-consistently through an iterative scheme:

1. Perform a gas-phase SCC-DFTB energy calculation to obtain the initial solute wave function and Mulliken charges.

2. Substitute Mulliken charges into eq 17 to obtain the atomic radii and establish the dielectric boundary.

3. Solve the PB equation (eq 10) to obtain the reaction field, $\phi_{\mathrm{rf}}(\mathbf{R}_\alpha)$.

4. Resolve SCC-DFTB in the presence of reaction field perturbation (eq 14) to obtain a new set of Mulliken charges.

5. Check the convergence of energy (0.001 kcal/mol used for this work); if the convergence criterion is not met, return to step 2.

6. On the basis of converged atomic radii, calculate SASA,[69] the nonpolar contribution, and the total energy of the solute in solution. For most molecules tested here, it requires less than 10 iterations (typically 4−8) of atomic radii/Mulliken charges to update each geometry.

D. Parameter Optimization. The new parameters in the SCC-DFTB/PB-based solvation model are the $A_{i(\alpha)}$ and $B_{i(\alpha)}$ in eq 17, which are dependent only on the element type. Although in principle the surface tension parameter in eq 6 can also be optimized, we have not done so (see below) because, for the systems of interest, the nonpolar contribution tends to be overwelmed by the electrostatic component; the value of γ adopted is 0.005 kcal/(mol·Å^2), which is commonly used in protein simulations using implicit solvent models.[75] For optimizing $A_{i(\alpha)}$ and $B_{i(\alpha)}$, two training sets with molecules of broad chemical compositions have been constructed (see the Supporting Information), for which the experimental solvation free energies are taken from refs 26, 31, and 76. Set 1 is used for parametrizing the solvation model with the standard (second-order) SCC-DFTB method and includes 103 species that contain C, H, O, and N; the list includes alkane, alkene, alkyne, arene, alcohol, aldehyde, carboxylic acid, ketone, ester, amine, amide, and other biorelated molecules and ions. Set 2 is used for parametrizing the solvation model with SCC-DFTBPR and includes 57 species that contain C, H, O, and P; the list includes representative species from set 1 plus phosphorus-containing molecules. Both sets contain a large number of charged species (57 in set 1 and 24 in set 2), which is essential for parametrizing the charge dependence of atomic radii.

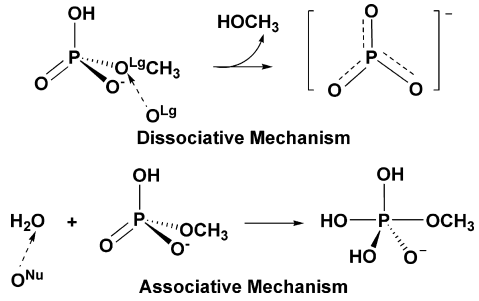
The parameters are optimized using a genetic algorithm (GA)[77] in which the "fitness" (ξ) is defined as the inverse of a weighted sum of difference between solvation free energies determined from calculation and experimentation:

$$\xi^{-1} = \frac{\sum_{i=1} w_i[\Delta G_i^{\mathrm{solv}}(\mathrm{exp}) - \Delta G_i^{\mathrm{solv}}(\mathrm{calc})]^2}{\sum_{i=1} w_i} \tag{18}$$

where i is the index of species in the training set, and the sum is over all molecules in the training set. For the weighting factors (w_i), 1.0 and 0.1 are used for the neutral molecules and ions according to the typical uncertainties in the experimental values; as analyzed by Kelly et al.,[31] the typical uncertainties in *experimental data* for neutral molecules and ions are 0.2 and 3 kcal/mol, respectively. During optimization, a micro-GA technique with a population of 10 chromosomes was allowed to operate for 500 generations with uniform crossovers; see ref 78 for detailed descriptions and recommendations for GA options.

In principle, geometry change upon solvation should be taken into consideration for a meaningful comparison to

Scheme 1



experimental results. In practice, this is very time-consuming for parameter fitting, even with the semiempirical QM method (SCC-DFTB) we employ here. Several authors discussed this point[31,63] and concluded that the change in geometry is generally small. However, in several cases, such as alcohol anions, we have observed significant structural changes upon solvation that have a substantial influence on the calculated solvation free energy. Therefore, a compromise is adopted: the gas-phase geometries are used to obtain the initial set of solvation parameters ($A_{i(\alpha)}$, $B_{i(\alpha)}$). With this set of parameters, solutes that have solvation free energy changes larger than 5 kcal/mol upon geometry optimization in solution are identified, and their geometries in solution are updated for the optimization of a new set of $A_{i(\alpha)}$, $B_{i(\alpha)}$. This cycle continues until all cases with major structural changes upon solvation have been taken into account.

Finally, it is worth noting that systematic optimization of surface tension coefficient γ (eq 6) results in negligible improvements for both neutral molecules alone and the complete training sets. A possible reason is that the nonpolar contribution to solvation free energy in the current scheme is also charge-dependent due to the use of charge-dependent atomic radii in SASA calculations. Therefore, compared with the fixed-radii scheme, the dependence of solvation free energy on γ is substantially smaller.

E. Additional Benchmark Calculations and Studies of (H)MMP/TMP Hydrolysis. To test the transferability of the optimized parameters, test sets are constructed (see the Supporting Information), which contain 32 for SCC-DFTB and 22 for SCC-DFTBPR. The calculated solvation free energies (including full geometry optimization in solution) are compared to the experimental values; similar to the training sets, the test cases contain a significant number of ionic species. As a comparison to popular and well-established solvation models, we also studied the same sets of molecules with the SM6 model of Cramer and co-workers.[31]

In addition, we have studied the mechanism[79,80] (first steps of both dissociative and associative pathways, see Scheme 1) of monomethyl monophosphate ester (MMP) hydrolysis using the SCC-DFTBPR/PB model. The potential energy surface is first explored by adiabatic mapping; the reaction coordinates include the P−$O^{Lg/Nu}$ distance (where O^{Lg} is the oxygen atom of the leaving group, methanol, and O^{Nu} is the oxygen in the nucleophilic water) and the antisymmetric stretch that describes the relevant proton transfers that involve $O^{Lg/Nu}$. The antisymmetric stretch is defined as the donor−proton distance minus the acceptor−proton distance. 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. Following the adiabatic mapping calculations, the structures along the approximate reaction path are examined carefully to ensure that the change of geometry is continuous along the path; in addition, the saddle point is optimized by conjugated peak refinement (CPR).[81] Finally, frequency calculations are carried out to confirm the nature of the stationary points and to compute the vibrational entropy and zero point energies. The results are compared to previous calculations with *ab initio* QM-based implicit solvent model calculations,[65,66,82] SCC-DFTBPR/MM calculations by us,[55] and available experimental data. To correct for intrinsic errors of SCC-DFTBPR, we also explore corrections based on single-point energy calculations with B3LYP/6-311++G(d,p) at SCC-DFTBPR geometries in the gas phase; this level of theory was found to give very similar results for the reactions of interest compared to MP2 and large basis sets.[55] As discussed in the literature,[83] such a simple correction may not always improve the energetics for semiempirical methods given the errors in the geometries; however, our previous tests[55] indicated that this correction scheme appears useful for SCC-DFTBPR since the method gives fairly reliable structures, even for transition states.

Finally, we briefly compare the energetics of protonated MMP (HMMP) and trimethyl monophosphate ester (TMP) hydrolysis with OH^- as the nucleophile (see Scheme 2). This is motivated by the previous work of Warshel and Florián,[84] who discussed the roles of neutral water vs OH^- as the nucleophile in MMP hydrolysis. Since SCC-DFTBPR was developed on the basis of MMP hydrolysis with water as the nucleophile,[55] this study helps to gain initial insights into the transferability of SCC-DFTBPR and lays the groundwork for possible future developments. For better comparison to previous calculations,[67,84] we follow the same two-dimensional adiabatic mapping calculations with the bond lengths for the forming and breaking P−O bonds as the reaction coordinates. Single-point B3LYP/6-311++G(d,p) calculations in the gas phase are used as an attempt to correct for intrinsic errors of SCC-DFTBPR.

III. Results and Discussions

A. Performance for the Training and Test Sets. The trends in optimized atomic radii (see Table 1) are consistent with other implicit solvent models and chemical intuition. For example, P has the largest charge-independent radius ($A_{i(\alpha)}$), while C, O, and N have comparable values, leaving H as the smallest. The absolute values are larger than those in SM6 and also the Bondi radii.[85] Compared with the atom-type-based charge-dependent radii in CD-COSMO by Dupuis et al.,[63] comparable values are found for nitrogen and oxygen in our model and the "internal −N", "terminal oxygen", and "internal −O" in CD-COSMO. The hydrogen radius (∼1.4 Å) in our model is larger than that (polar hydrogen) in CD-COSMO (1.202 Å). In terms of the charge dependence, the typical $B_{i\alpha}$ values are around −0.10, although they are substantially larger (∼−0.2) for C in SCC-DFTB and H in

Scheme 2

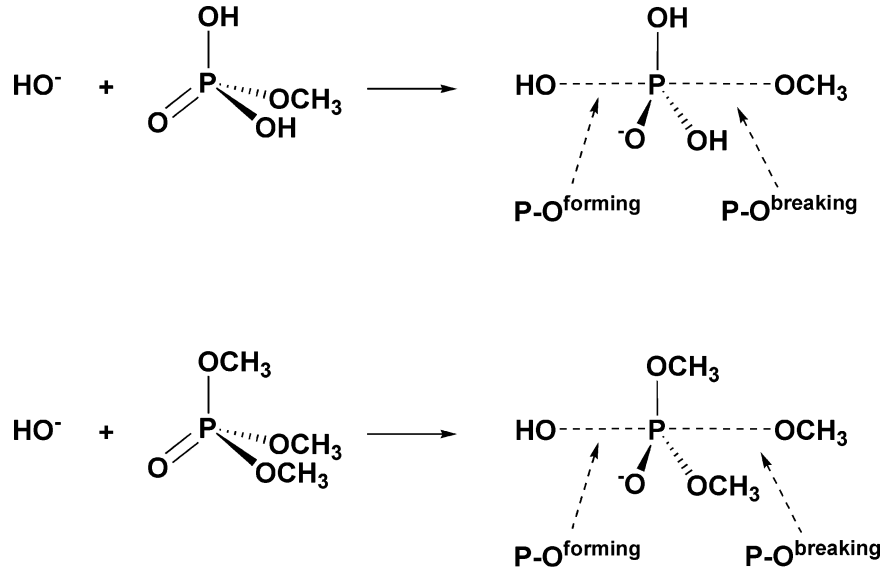


Table 1. Optimized Atomic Radii Parameters and Comparison to Other Values from the Literature[a]

	SCC-DFTB		SCC-DFTBPR			
element	$A_{i(\alpha)}$	$B_{i(\alpha)}$	$A_{i(\alpha)}$	$B_{i(\alpha)}$	SM6[31]	Bondi[85]
C	1.85	−0.24	2.07	−0.05	1.57	1.70
O	1.70	−0.11	1.87	−0.07	1.52	1.52
N	1.94	−0.01	N/A	N/A	1.61	1.55
P	N/A	N/A	2.47	−0.10	1.80	1.80
H	1.47	−0.11	1.41	−0.25	1.02	1.20

[a] $A_{i(\alpha)}$ in Å, $B_{i(\alpha)}$ in Å per charge. The values shown are fitted with solution geometry optimization (see the Methods).

SCC-DFTBPR. Even the latter are nearly half of the values in CD-COSMO, which is probably due to the use of different charges in SCC-DFTB (Mulliken) and CD-COSMO (CHELPG). It is worth emphasizing that the parameters in our model depend only on element type, rather than atom type as in CD-COSMO; therefore, CD-COSMO probably tends to be more accurate (see below for some comparison) while our scheme tends to be less problematic for studying transition states, which likely involve change in atom types.

As shown in the Supporting Information, the absolute value of solvation free energy is usually less than 10 kcal/mol for neutral molecules but larger than 60 kcal/mol for ions. Therefore, it is generally challenging to reproduce the solvation free energy of ions in a reliable fashion. Nevertheless, as shown in Table 2, the overall performance of our SCC-DFTB(PR)-based solvation model is very encouraging. For example, for ions, the mean unsigned error (MUE) for SCC-DFTB is ∼3 kcal/mol either without or with geometry optimization in solution. For SCC-DFTBPR, the error is slightly larger, with the corresponding MUE values of 5 and 4 kcal/mol. These values can be compared to results from the SM6 model,[31] which is one of the most sophasticated and well-calibrated models developed with *ab initio* DFT methods; the MUE values are 4 and 5 kcal/mol for the first (for SCC-DFTB) and second (for SCC-DFTBPR) training sets, respectively, which are even slightly larger than the values for our SCC-DFTB(PR)-based solvation model.

The level of performance deteriorates slightly for the test sets. As shown in Table 3, for example, the MUE for the ions in the first and second test sets is 3 and 5 kcal/mol, respectively, when geometry optimization in solution is carried out; without solution geometry optimization, the MUE values are 4 and 6 kcal/mol. By comparison, the SM6MUE values are 5 and 7 kcal/mol, again slightly larger than the SCC-DFTB(PR) values. These benchmark calculations indicate that the good performance of our model is fairly transferrable. This is very encouraging since the SCC-DFTB(PR)-based calculations are much faster than the DFT (MPW1PW91/6-31+G(d,p))-based SM6 calculations. Compared with CD-COSMO,[64] which is also DFT-based and involves more elaborate parametrization of charge dependence of atomic radii, it is again encouraging to see that for the three ions tested by both models, the performance is comparable. For example, for hydroxide, SCC-DFTB with or without solution geometry optimization gives an error of 2 kcal/mol, while CD-COSMO gives 3 kcal/mol. For ammonium, SCC-DFTB has an error of −3 kcal/mol, while CD-COSMO gives −2 kcal/mol. For methylamine(+1), the corresponding values are −3 and −4 kcal/mol, respectively.

We note that, relatively speaking, the performance of our model for neutral molecules is less stellar. In fact, for both the training and test cases, the SM6 model consistently outperforms the SCC-DFTB(PR) solvation model; e.g., the MUE is typically smaller by ∼1 kcal/mol with SM6 (see Tables 2 and 3). This is likely because parameters in the nonpolar component, which makes a significant (relative to ions) contribution to the total solvation free energy of neutral molecules, we have not optimized in the current model. Indeed, in the work of Xie and Liu,[76] who have implemented a GBSA model with SCC-DFTB, a root mean square error (RMSE) of 1.1 kcal/mol was obtained for 60 neutral molecules containing C, H, O, N, and S when the nonpolar parameters were optimized. On the other hand, we note that for most chemical reactions of biological relevance, the nonpolar contribution likely plays a much less significant role compared to the electrostatic component. Finally, as

Table 2. Error (in kcal/mol) Analysis of Solvation Free Energies for Training Sets 1 and 2[a]

	single point[b]			optimization[c]			SM 6[d]		
	RMSE	MUE	MSE	RMSE	MUE	MSE	RMSE	MUE	MSE
neutral	2.0	1.7	0.6	2.1	1.7	0.4	0.8	0.7	0.4
ions	4	3	2	3	3	0	4	4	2
all data	3	3	1	3	2	0	3	2	1
neutral	1.6	1.3	−0.5	2.0	1.9	−1.3	1.5	0.9	0.6
ions	4	5	5	4	4	2	4	5	5
all data	4	3	2	4	3	0	4	3	2

[a] The first three rows are for the first training set (for SCC-DFTB), and the three bottom rows are for the second training set (for SCC-DFTBPR). RMSE, root-mean-square error; MUE, mean unsigned error; MSE, mean signed error. All errors measured against experimental solvation free energies, which have typical uncertainties of 0.2 and 3 kcal/mol for neutral molecules and ions, respectively. [b] With gas-phase geometries. [c] With solution-phase geometry optimizations (see Methods). [d] Results are obtained by MPW1PW91/6-31+G(d,p).

Table 3. Error Analysis (in kcal/mol) of Solvation Free Energies for Test Sets 1 and 2[a]

	single point			optimization			SM6		
	RMSE	MUE	MSE	RMSE	MUE	MSE	RMSE	MUE	MSE
neutral	2.2	1.8	0.7	2.3	1.9	0.2	1.0	0.8	−0.2
ions	5	4	2	4	3	1	6	5	2
all data	4	3	1	3	3	0	4	2	1
neutral	1.5	1.4	−1.2	2.1	2.1	−2.0	0.9	0.7	−0.1
ions	7	6	2	7	5	0	7	7	5
all data	4	3	0	4	3	−1	5	3	2

[a] See Table 2 for format.

shown in the Supporting Information, our solvation model gives rather large errors for amine and amide molecules; for example, the error for ammonia is more than 3.2 kcal/mol with or without solution geometry optimization, which is more than 70% off the experimental value. This behavior was noted in a previous analysis of implicit solvation models,[12] and it was argued that hydrogen-bonding energies are poorly correlated with classical electrostatic interaction energies, and therefore more sophisticated treatments are needed for such short-range interactions.

B. MMP Hydrolysis Reaction with Neutral Water As Nucleophile. Experimental studies of the MMP hydrolysis reaction[86−88] determined that the reaction rate peaks at a pH of 4−5 with an activation energy of 31 kcal/mol. The reaction mechanism is traditionally regarded as dissociative, though dispute still exists.[89] Here, as a benchmark calculation for the new solvation model, we investigate the first steps of both dissociative and associative pathways (see Scheme 1) and compare the results with previous theoretical studies.[55,65,67]

For the dissociative pathway, the adiabatic map in solution with our new solvation model (Figure 1a) is qualitatively consistent with previous PMF results obtained using explicit solvent SCC-DFTBPR/MM simulations.[55] The transition state region involves largely an intramolecular proton transfer from the protonated oxygen in MMP to the oxygen in the leaving group (O^{Lg}), and the $P-O^{Lg}$ bond is only slightly stretched compared to MMP. As discussed in ref 55, the $P-O^{Lg}$ bond in the transition state decreases significantly from the gas phase (∼2.1 Å) to solution (∼1.7−1.8 Å in SCC-DFTBPR/MM PMF simulations); thus our model has captured this solvation effect adequately. Following the proton transfer, a zwitterionic intermediate is formed, which is again in qualitative agreement with both SCC-DFTBPR/MM PMF calculations[55] and the previous DFT-PCM study.[65]

More quantitatively, the fully optimized structures for MMP, the transition state (dis_ts), and the zwitterionic intermediate (dis_zt) at the SCC-DFTBPR level are in decent agreement with previous calculations; the optimized structure does not depend sensitively on the grid size in the PB calculations (for a comparison of 0.2 vs 0.4 Å grid sizes, see Figure 2, which also contains an illustration for the imaginary mode in the optimized transition state, dis_ts, with a frequency of $1742i$ cm^{-1}). Compared to the work of Vigroux et al.,[65] in which the structures were optimized at the level of B3LYP-PCM with a double-ζ-quality basis set plus diffuse and polarization functions and a pseudopotential for non-hydrogen atoms, the only major difference is that their optimized $P-O^{Lg}$ distances in dis_ts and dis_zt are longer by ∼0.1 Å and 0.25 Å, respectively. The study of Florián and Warshel[67] did not examine the zwitterionic intermediate, and the $P-O^{Lg}$ distance in their transition state is substantially longer than both values from this work and from ref 66. This is likely because the geometries of Florian et al.[67] were mainly optimized in the gas phase, and the transition state in solution was only approximately located by single-point Langevin dipole calculations along the minimum energy path from gas-phase calculations.

For the energetics, the free energy barrier estimated with the current SCC-DFTBPR-based solvation model is 34.8 kcal/mol; including a single-point B3LYP/6-311++G(d,p) gas-phase correction lowers the barrier to be 31.3 kcal/mol. As shown in Table 4, these values are consistent with previous calculations[65,67] and experimental studies,[90] which range from 30.7 to 34 kcal/mol. For the zwitterionic intermediate, which was first discussed in the work of Bianciotto et al.,[65,66] the current solvation model with SCC-DFTBPR predicts a free energy of 13.7 kcal/mol above the MMP reactant; with the B3LYP correction, the value becomes 21.1 kcal/mol. The large magnitude of the gas-phase

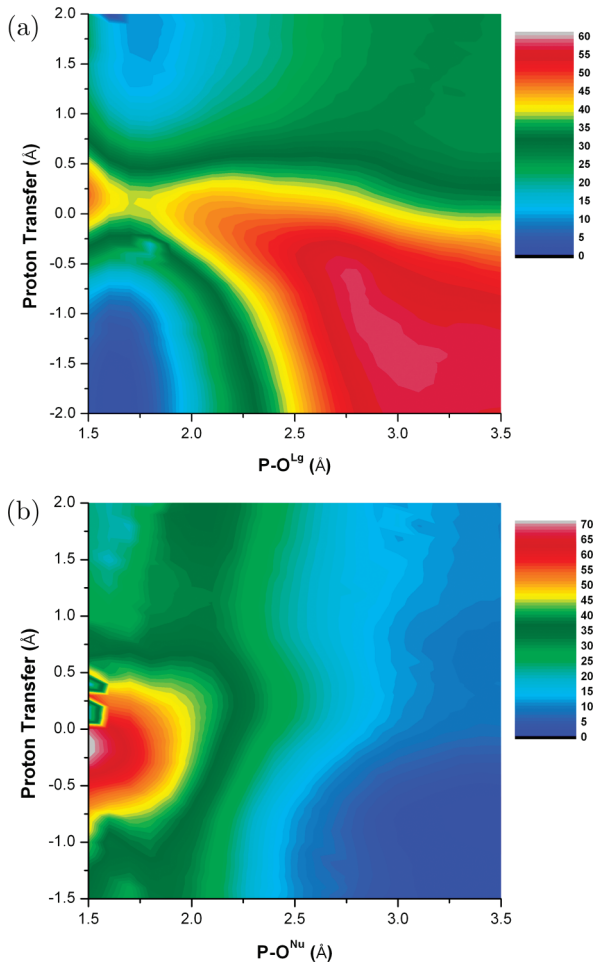


Figure 1. Adiabatic mapping results (energies in kcal/mol) for the first step of (a) the dissociative and (b) the associative pathway for the hydrolysis of monomethyl monophosphate ester (MMP). The O^{Lg} stands for the oxygen in the leaving group (see Scheme 1), which is methanol in this case; O^{Nu} stands for the oxygen in water (see Scheme 1). In a, the proton transfer coordinate is the antisymmetric stretch that describes the intramolecular proton transfer between the protonated oxygen in MMP and O^{Lg}; in b, the proton transfer coordinate is the antisymmetric stretch that describes the proton transfer between the nucleophilic water and the basic oxygen in MMP.

correction was discussed in our previous study,[55] which emphasized that the SCC-DFTBPR model was developed without any information concerning the zwitterionic region of the potential energy surface. The B3LYP corrected free energy value is in close agreement with the DFT-PCM study of Bianciotto et al.,[65] who predicted a value of 21.2 kcal/mol. Most importantly, our solvation model does not suffer from the unphysically large stabilization found in explicit solvent SCC-DFTBPR/MM simulations, which predicted that the zwitterionic intermediate is *lower* than the reactant (MMP) by ~3 kcal/mol. As discussed in ref 55, such significant overstabilization of the zwitterionic intermediate highlighted the need to improve QM/MM interactions beyond the typical form with parameters that do not reflect the electronic structure of the QM region.[91] The success of the current solvation model, on the other hand, illustrates that the charge dependence of QM/MM interactions can be effectively treated by adopting charge dependent radii in implicit solvent calculations when studying solution reactions that involve significant charge redistribution.

For the associative pathway, the adiabatic map (Figure 1b) is qualitatively similar to the PMF from explicit solvent SCC-DFTBPR/MM simulations.[55] For example, the potential energy surface is rather flat in regions with long $P-O^{Nu}$ distances but positive proton transfer coordinates, which suggests that proton transfer from the nucleophilic water to MMP can occur prior to the nucleophilic attack. Indeed, we obtained a local minimum with geometry optimization that corresponds to a molecular complex between OH^- and protonated MMP (HMMP) on the potential energy surface. Compared to the reaction complex between water and MMP (asc_pre), this complex (asc_hydro) is substantially higher in energy by ~15 kcal/mol; including the B3LYP/6-311++G(d,p) gas-phase correction further increases the value to ~26.2 − 3.6 = 22.6 kcal/mol (see Table 5). Once again, the large magnitude of the correction reflects deficiency in the current SCC-DFTBPR approach for balancing the proton affinity of phosphate and nonphosphate species, which remains an interesting challenge for future improvement.[55]

Both the adiabatic mapping and saddle point optimization point to an associative transition state in which the $P-O^{Nu}$ distance is ~2 Å and the water proton is already transferred to the phosphate oxygen (see Figure 3 for the structure of the transition state, asc_ts). Compared to the structure optimized by Florián and Warshel[67] with the Langevin dipole model, the key difference is that the proton transfer is halfway in their structure, with an $O^{Nu}-H$ distance of 1.44 Å, compared to the value of 2.18 Å in our case. Since our structure is consistent with the previous PMF results based on SCC-DFTBPR/MM simulations, we suspect that the difference is again due to the limited solution geometry optimization in the work of Florián and Warshel[67] (see the discussions above for the dis_ts). The agreement in the optimized structures for the penta-valent intermediate, asc_int, from the two sets of studies is much better, as expected (see Figure 3).

As for the energetics for the associated pathway, the SCC-DFTBPR-based solvation model gives a free energy barrier of 33.1 kcal/mol, which increases slightly to 37.5 kcal/mol when gas-phase B3LYP correction is included. These values, especially the one with B3LYP correction, are close to previous computational studies (see Table 5) but somewhat higher compared to the experimental value of 30.7 kcal/mol.[90] The pentavalent species, asc_int, is also less stable by a few kcal/mol compared to the study of Florián and Warshel.[67] We note that all calculations found that the barrier for the associative pathway is higher than that in the dissociative pathway, although the difference is fairly small (~1−2 kcal/mol) with either SCC-DFTBPR/MM or the Langevin dipole model, while the SCC-DFTBPR-based solvation model gives the largest difference (~6 kcal/mol) when B3LYP correction is included. Before more systematic analysis into the quantitative nature of B3LYP correction is performed, it remains premature to conclude that MMP hydrolysis strongly prefers a dissociative pathway.

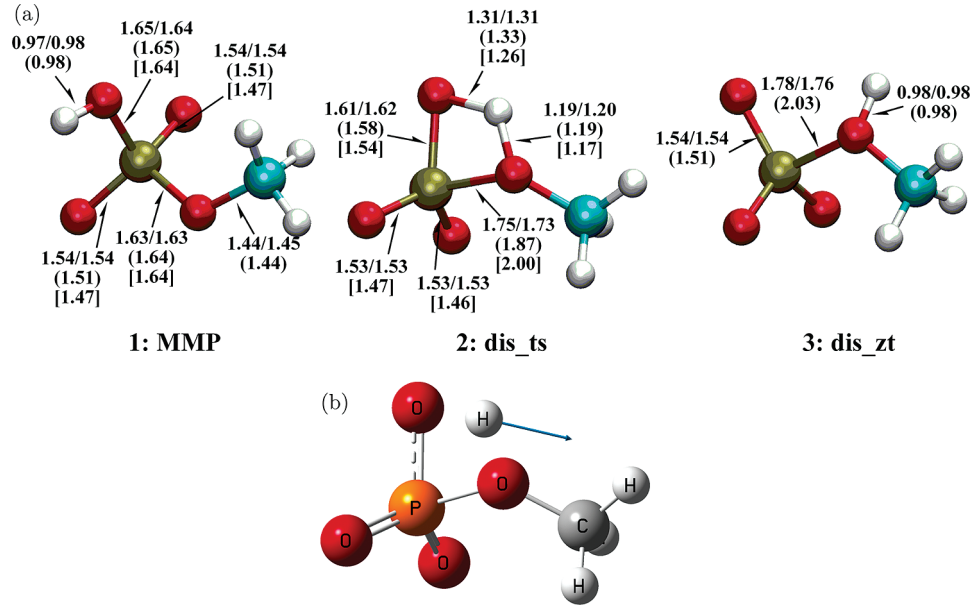


Figure 2. Geometries of the reactant, transition state, and the zwitterionic intermediate for the first step of the dissociative pathway for the hydrolysis of monomethyl monophosphate ester (MMP). (a) Values (in Å) without parentheses are from the current SCC-DFTBPR-based solvation model calculations with a grid size of 0.2/0.4 Å. Values with parentheses are from ref 65, which were obtained with B3LYP-PCM and a double-ζ-quality basis set plus diffuse and polarization functions. Values with brackets are from ref 68, which were obtained with HF/6-31G(d) in the gas phase with approximate adjustments for solvation using the Langevin dipole model. (b) An illustration of the imaginary vibrational mode in dis_ts.

Table 4. Energetics for the First Step of the Dissociative Pathway of MMP Hydrolysis from Current[a] and Previous Studies[b]

species	ΔE^c	$T\Delta S^c$	ΔZPE^c	ΔG^c	ref 55	ref 65	ref 67	exptl.[90]
MMP	*−11774.6*	*24.5*	*39.8*	*−11759.2*				
dis_ts	39.0/35.5	1.4	−2.8	34.8/31.3	32	33.5	34	30.7
dis_zt	12.6/20.0	−0.1	1.0	13.7/21.1	−3	21.2		

[a] For MMP, the total energies are given (in italics); for other species, energetics relative to MMP are given in kcal/mol. The entropic contribution ($T\Delta S$, T = 373 K in all tables, including for the experimental rate constants) and zero-point energy correction (ZPE) are calculated with the SCC-DFTBPR-based solvation model and harmonic-oscillator-rigid-rotor approximation. [b] Reference 55 employs explicit solvent SCC-DFTBPR/MM PMF simulations. Reference 65 used B3LYP-PCM and a double-ζ-quality basis set plus diffuse and polarization functions, and the pseudopotential for non-hydrogen atoms. In ref 67, geometries were obtained with HF/6-31G(d) in the gas phase with approximate adjustments for solvation using the Langevin dipole model; single-point calculations are performed at the MP2/6-31+G(d,p) level with the Langevin dipole for solvation. [c] Numbers before the slash are SCC-DFTBPR results; numbers after the slash are results after single-point gas-phase correction at the level of B3LYP/6-311++G(d,p).

Table 5. Energetics for the First Step of the Associative Pathway of MMP Hydrolysis[a]

species	ΔE	$T\Delta S$	ΔZPE	ΔG	ref 55	ref 67	exptl.[90]
MMP + H_2O	*−14345.9*	*38.3*	*53.1*	*−14331.2*			
asc_pre	−8.8/−7.0	−9.3	1.3	1.8/3.6			
asc_hydro	6.8/16.5	−9.1	0.6	16.5/26.2			
asc_ts	22.6/27.0	−9.6	0.8	33.1/37.5	34	35	30.7
asc_int	20.6/23.0	−10.4	1.4	32.5/34.9		29	

[a] Same format as in Table 4; the reference is infinitely separated MMP and H_2O.

C. HMMP and TMP Hydrolysis with OH^- as Nucleophile. A long-standing mechanistic postulate for MMP hydrolysis is that it is possible to exclude the nucleophilic attack of OH^- on the neutral phosphate. The argument was based on the high activation energy measured for the OH^- attack of trimethyl monophosphate (TMP) at high pH, which is around 25 kcal/mol (at 373K),[92] and the underlying assumption was that HMMP and TMP hydrolysis reactions have similar activation barriers. However, as pointed out by Florián and Warshel,[84] this analogy was not necessarily valid, and their calculations based on MP2 and the Langevin dipole solvation model found that the barriers for OH^- attack of HMMP and TMP differ by more than 10 kcal/mol. Moreover, the barrier of ∼12 kcal/mol found for HMMP was sufficiently low to make the OH^- attack pathway a competing mechanism of MMP hydrolysis. As an interesting benchmark of our solvation and the transferability of SCC-DFTBPR, we compare the barriers for the hydrolysis of HMMP and TMP with OH^- as the nucleophile (see Scheme 2).

As shown in Figure 4, the overall energy landscapes are quite similar for HMMP and TMP, both undergoing an associative mechanism with the new $P-O^{forming}$ bond largely formed before the $P-O^{breaking}$ broke. The transition state from the adiabatic mapping for HMMP is very consistent with the optimized saddle point asc_ts, which clearly is more

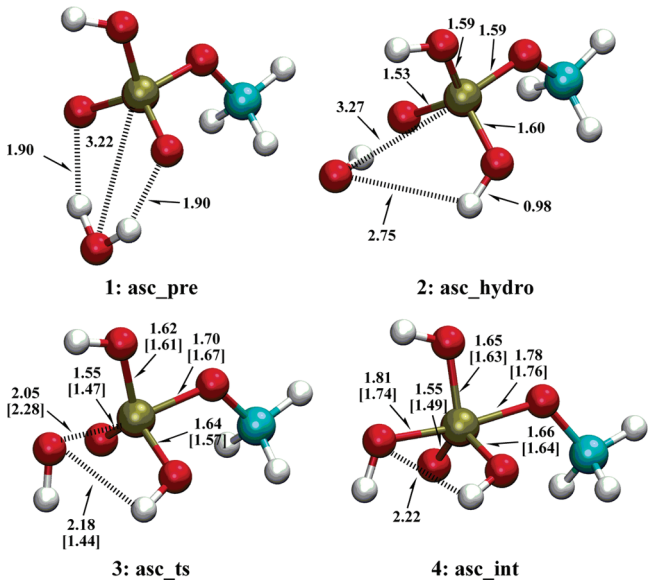


Figure 3. Similar to Figure 2, but for structures along the first step of the associative pathway for MMP hydrolysis.

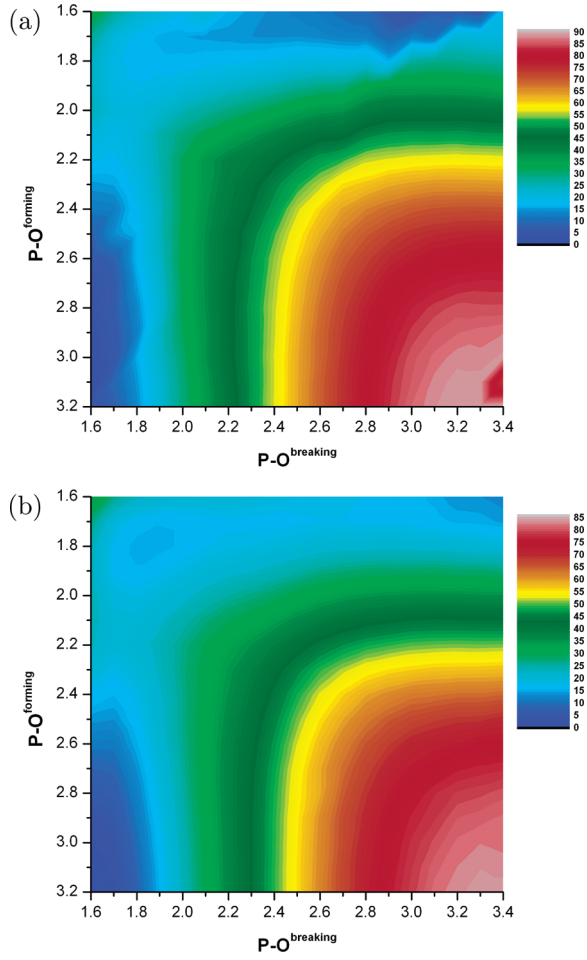


Figure 4. Adiabatic mapping results (energies in kcal/mol) for the hydrolysis of (a) hydrogen methyl monophosphate ester (HMMP) and (b) trimethyl monophosphate ester (TMP) by hydroxide. See Table 6 for the summary of the barrier heights, in which the reference is infinitely separated reactant molecules.

appropriately classified as the transition state for the OH^- attack of HMMP. According to Table 6, the corresponding energy barriers are 24.0 and 30.9 kcal/mol, with the TMP case higher by ~7 kcal/mol. Including single-point B3LYP/6-311++G(d,p) gas-phase correction further increases the gap to ~13 kcal/mol, which agrees very well with the result of Warshel and Florián.[84] This is a satisfying observation since SCC-DFTBPR was mainly parametrized on the basis of MMP and dimethyl monophosphate ester (DMP) hydrolysis; as speculated in our original work,[55] however, the parameters are likely transferrable to other phosphates that follow similar reaction mechanisms because the number of parameters is fairly small. On the absolute scale, it appears that our estimates (for both HMMP and TMP) are systematically higher, by ~4 kcal/mol, than the results of Florián and Warshel[84] and the experimental barrier for TMP.[92]

Table 6. Relative Free Energies of Key Species for the Hydrolysis of MMP and TMP along Associative Pathway with Hydroxide as the Nucleophile[a]

species	ΔE	$T\Delta S$	ΔZPE	ΔG	ref 84	Exptl.[90,92]
asc_ts	15.6/7.0	−6.9	1.4	24.0/15.4	11.7	
tmp_ts	21.9/19.4	−7.9	1.2	30.9/28.5	24.7	24.6

[a] Same format as in Table 4; the reference is infinitely separated HMMP/TMP and hydroxide.

IV. Conclusion

We report the development of an implicit solvent model for SCC-DFTB(PR) in which the solvation free energy is computed on the basis of Poisson−Boltzmann for electrostatics and a surface area term for nonpolar contributions. The unique aspect of our model is that the atomic radii that define the dielectric boundary of the solute are dependent on the solute charge distribution and are determined in a self-consistent fashion with the electronic structure of the solute. This self-consistency makes it possible to balance the solvation treatment of species with different charge distributions, such as neutral vs ionic species and structures along a chemical reaction pathway. Indeed, benchmark calculations have shown that, even for ions, our model leads to results of comparable accuracy to the much more sophisticated SM6 model; this is very encouraging since SCC-DFTB(PR) calculations are at least hundreds of times faster than the DFT calculations required in the SM6 model. We expect that the accuracy of the calculations can be further improved if more sophisticated charge schemes[9,93] are used to replace the simple Mulliken analysis.

Since our implementation has analytic first derivatives, the solvation model can be readily used to explore potential energy surfaces for solution reactions, which is the major motivation for the current work. This is demonstrated with a brief study of dissociative and associative pathways of MMP hydrolysis, as well as the hydrolysis of protonated MMP and TMP with OH^- as the nucleophile. The results (geometries and energetics) are largely in good agreement with previous computational studies using QM/MM or *ab initio*/DFT in conjunction with dielectric continuum models, as well as with available experiments. In particular, the solvation model avoids the overstabilization of the zwitterionic species along the dissociative pathway as found in explicit solvent SCC-DFTBPR/MM simulations.[55] This

highlights the complementary nature of implicit solvent model to explicit solvent approaches for studying solution reactions that involve significant charge reorganizations.

Due largely to the computational efficiency of SCC-DFTB(PR), we anticipate that the current solvation model can be effectively used in semiquantitative exploration of mechanisms for solution reactions, such as ruling out certain reaction pathways and obtaining approximate structures of key transition states and intermediates, which can be further refined with higher-level calculations. As further developments occur, it would be interesting to extend the formulation of charge-dependent radii to more approximate solvation models such as generalized Born,[75] which can be computationally more efficient than Poisson−Boltzmann; this is particularly true in molecular dynamics simulations, which can be effective for estimating the entropic contribution to reaction energetics in the framework of quasiharmonic analysis. Along this line, as extensively discussed in the literature, the first solvation shell of the solute can be treated explicitly, either at the same level of QM theory[9,12] or with a molecular mechanics model.[94,95] Since SCC-DFTB(PR) is fast, making such an extension of the molecular model for a better treatment of solvation is likely more cost-effective than with *ab initio*/DFT methods. Finally, to further improve the computational efficiency and numerical robustness of the PB aspect of the calculations, switching to a boundary element framework[96] is also of considerable interest.

Acknowledgment. This research was supported by the National Institutes of Health (R01-GM071428 and R01-GM084028) and, in part, by the National Science Foundation (CRC-CHE-0404704). Computational resources from the National Center for Supercomputing Applications at the University of Illinois and the Centre for High Throughput Computing (CHTC) at UW−Madison are greatly appreciated.

Supporting Information Available: The details of calculated and experimental solvation free energies for training and testing sets are included. This material is available free of charge via the Internet at http://pubs.acs.org.

References

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

(2) 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.

(3) Zhang, Y. K. *Theor. Chem. Acc.* **2006**, *116*, 43–50.

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

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

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

(7) Marx, D.; Hutter, J. *Ab initio molecular dynamics: Basic theory and advanced methods*; Cambridge University Press: Cambridge, U.K., 2009; pp 9−75.

(8) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161–2200.

(9) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760–768.

(10) Sato, H.; Hirata, F.; Kato, S. *J. Chem. Phys.* **1996**, *105*, 1546–1551.

(11) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; RingaldaI, M.; Goddard, W. A.; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875–11882.

(12) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A. *J. Phys. Chem.* **1996**, *100*, 11775–11788.

(13) Miertus, S.; Tomasi, *J. Chem. Phys.* **1982**, *65*, 239–245.

(14) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327–335.

(15) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210–3221.

(16) Cancès, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032–3041.

(17) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151–5158.

(18) Amovilli, C.; Mennucci, B. *J. Phys. Chem. B* **1997**, *101*, 1051–1057.

(19) Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. *Chem. Phys. Lett.* **1998**, *286*, 253–260.

(20) Barone, V.; Cossi, M.; Tomasi, J. *J. Comput. Chem.* **1998**, *19*, 404–417.

(21) Li, H.; Jensen, J. H. *J. Comput. Chem.* **2004**, *25*, 1449–1462.

(22) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Chem. Phys.* **2001**, *114*, 5691–5701.

(23) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, *117*, 43–54.

(24) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Chem. Comput.* **2003**, *24*, 669–681.

(25) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

(26) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824–19839.

(27) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.

(28) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983–10990.

(29) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *J. Chem. Phys.* **2002**, *116*, 10606–10614.

(30) Im, W. P.; Lee, M. S.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1691–1702.

(31) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133–1152.

(32) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 2011–2033.

(33) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799–805.

(34) Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224–2235.

(35) klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074–5085.

(36) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

(37) York, D. M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 11060–11079.

(38) Dolney, D. M.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2000**, *21*, 340–366.

(39) Florián, J.; Warshel, A. *J. Phys. Chem.* **1997**, *101*, 5583–5595.

(40) Wales, D. *Energy Landscapes*; Cambridge University Press: Cambridge, U.K., 2004.

(41) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210–3221.

(42) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.* **1996**, *100*, 16098–16104.

(43) Vilkas, M. J.; Zhan, C. G. *J. Chem. Phys.* **2008**, *129*, 194109.

(44) Zhan, C. G.; Chipman, D. M. *J. Chem. Phys.* **1998**, *109*, 10543–10558.

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

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

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

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

(49) Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 5614–5621.

(50) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. B* **2007**, *111*, 10861–10873.

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

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

(53) Zheng, G. S.; Witek, H. A.; Bobadova-Parvanova, P.; Irle, S.; Musaev, D. G.; Prabhakar, R.; Morokuma, K. *J. Chem. Theo. Comp.* **2007**, *3*, 1349–1367.

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

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

(56) York, D. M.; Lee, T. S.; Yang, W. T. *Chem. Phys. Lett.* **1996**, *263*, 297–304.

(57) Gogonea, V.; Merz, K. M. *J. Phys. Chem. A* **1999**, *103*, 5171–5188.

(58) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509.

(59) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144–1149.

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

(61) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111*, 59–75.

(62) Aguilar, M. A.; del Valle, F. J. O. *Chem. Phys.* **1989**, *129*, 439–450.

(63) Ginovska, B.; Camaioni, D. M.; Dupuis, M.; Schwerdtfeger, C. A.; Gil, Q. *J. Phys. Chem. A* **2008**, *112*, 10604–10613.

(64) Ginovska, B.; Camaioni, D. M.; Dupuis, M. *J. Chem. Phys.* **2008**, *129*, 014506.

(65) Bianciotto, M.; Barthelat, J. C.; Vigroux, A. *J. Am. Chem. Soc.* **2002**, *124*, 7573–7587.

(66) Bianciotto, M.; Barthelat, J. C.; Vigroux, A. *J. Phys. Chem. A* **2002**, *106*, 6521–6526.

(67) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 719–734.

(68) Roux, B.; Simonson, T. *Bio. Chem.* **1999**, *78*, 1–20.

(69) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1348–1356.

(70) Jackson, J. D. *Classical Electrodynamics*, 3rd ed.; John Wiley & Sons: New York, 2001; pp 145−174.

(71) Im, W.; Berneche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924–2937.

(72) McQuarrie, D. A. *Statistical Mechanics*; Harper & Row: New York, 1976; pp 326−357.

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

(74) Yamaguchi, Y.; Goddard, J. D.; Osamura, Y.; Schaefer, H. *A new dimension to quantum chemistry: Analytic derivative methods in Ab initio molecular electronic structure theory*; Oxford University Press: Oxford, U.K., 1994; pp 448−471.

(75) Feig, M.; Brooks, C. L. I. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217–224.

(76) Xie, L.; Liu, H. *J. Comput. Chem.* **2002**, *23*, 1404–1415.

(77) Goldberg, D. E. *Genetic algorithms in search, optimization, and machine learning*; Addison-Wesley: Reading, MA, 1989; pp 1−25.

(78) Carroll, D. L. FORTRAN Genetic algorithm (GA) driver. http://cuaerospace.com/carroll/ga.html (accessed Jun 2010).

(79) Ba-Saif, S. A.; Davis, A. M.; Williams, A. *J. Org. Chem.* **1989**, *54*, 5483–5486.

(80) Barnes, J. A.; Wilkie, J.; Williams, I. H. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 1709–1714.

(81) S., F.; Karplus, M. *Chem. Phys. Lett.* **1992**, *194*, 511–527.

(82) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *ChemPhysChem* **2009**, *10*, 1125–1134.

(83) Cui, Q.; Karplus, M. *J. Phys. Chem B* **2002**, *106*, 1768–1798.

(84) Florián, J.; Warshel, A. *J. Am. Chem. Soc.* **1997**, *119*, 5473–5474.

(85) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441–451.

(86) Barnard, P. W. C.; Bunton, C. A.; Llewellyn, D. R.; Oldham, K. *Chem. Ind. (London)* **1955**, *760*, 2420–2423.

(87) Butcher, W. W.; Wesheimer, F. H. *J. Am. Chem. Soc.* **1955**, *77*, 2420–-.

(88) Bunton, C. A.; Llewellyn, D. R.; Oldham, K. G.; Vernon, C. A. *J. Chem. Soc* **1958**, 3574-.

(89) Åqvist, J.; Kolmodin, K.; Florián, J.; Warshel, A. *Chem. Biol.* **1999**, *6*, R71–R80.

(90) Bunton, C. A.; Llewellyn, D. R.; Oldham, K. G.; Vernon, C. A. *J. Chem. Soc.* **1958**, 3574–3587.

(91) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2007**, *127*, 194101.

(92) Barnard, P. W. C.; Bunton, C. A.; Llewellyn, D. R.; Vernon, C. A.; Welch, V. A. *J. Chem. Soc.* **1961**, 2670–2676.

(93) Nicholls, A.; Wlodek, S.; Grant, J. A. *J. Phys. Chem. B* **2009**, *113*, 4521–4532.

(94) Cui, Q. *J. Chem. Phys.* **2002**, *117*, 4720–4728.

(95) Li, H.; Gordon, M. S. *J. Chem. Phys.* **2007**, *126*, 124112.

(96) Lange, A. W.; Herbert, J. M. *J. Phys. Chem. Lett.* **2010**, *1*, 556–561.

CT1001818