THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

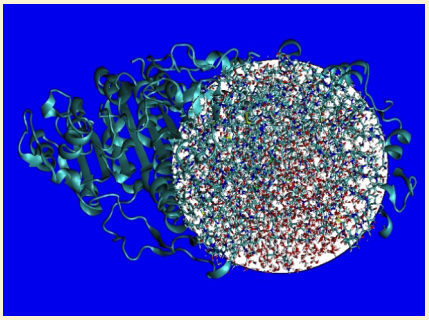
# Implementation of the Solvent Macromolecule Boundary Potential and Application to Model and Realistic Enzyme Systems

Jan Zienau and Qiang Cui*

Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, 1101 University Avenue, Madison, Wisconsin 53706, United States

Supporting Information

**ABSTRACT:** The implementation of the solvent macromolecule boundary potential (SMBP) by Benighaus and Thiel (*J. Chem. Theory Comput.* **2009**, *5*, 3114) into the program package CHARMM is presented. The SMBP allows for the efficient calculation of solvent effects for large macromolecules using irregularly shaped dielectric boundaries. In contrast to the generalized solvent boundary potential (GSBP) by Roux et al. (*J. Chem. Phys.* **2001**, *114*, 2924) from which it is derived, the SMBP is targeted for quantum mechanical/molecular mechanical (QM/MM) setups using ab initio methods for the QM part. After presenting benchmark results for simple model systems, applications of the SMBP for the calculation of geometries, reaction energy barriers, and vibrational frequencies for an alkaline phosphatase (AP) enzyme are discussed. Although the effect of the boundary potential on optimized structures (including the transition state) and vibrational frequencies is relatively small, the energetics of the phosphoryl transfer catalyzed by AP depend significantly on the boundary potential. Finally, to emphasize a unique feature of our implementation, we apply both SMBP and GSBP to the calculation of the energy barrier for a proton transfer reaction in a simple model channel, where the effect of an external transmembrane potential is studied. Due to the dipolar response of the polar environment, the effective charge displacement estimated based on the effect of the membrane potential on the proton transfer energetics deviates from the net charge that passes the membrane.



## 1. INTRODUCTION

It has been long recognized that the environment of biomolecules plays important roles in determining their structure, dynamics, and function. Reliable computer simulations of biomolecules, therefore, need to provide an efficient and robust description of the environment. At the current stage, the most effective approach is to explicitly represent all solvents, ions,[1,2] and possibly other small solutes[3−5] (or even other proteins[6]) that form the environment of the main component (e.g., an enzyme) of interest. This can be often prohibitively expensive and therefore not practical for many problems. An alternative choice is to divide the system into inner and outer regions, where the solvent/ions are represented explicitly in the inner region, and an implicit solvent model is used for the treatment of the bulk in the outer region. When the process of interest is relatively localized, such as catalysis in the active site of an enzyme, the inner region may cover only part of the macromolecule (e.g., the active site), so that the outer region consists of the rest of the macromolecule immersed in implicit solvent.[7,8] This setup leads to an irregular boundary between the macromolecule and implicit solvent in the outer region, so that simple implicit solvent models using, for example, spherical cavities,[9] cannot be applied.

An efficient way to solve this problem is provided by the generalized solvent boundary potential (GSBP),[10] which focuses on computing the nontrivial electrostatic contribution of the solvation free energy. Here, a key element is fixing the outer region of the macromolecule, so that only the degrees of freedom of the inner region need to be explicitly considered. Although this approach is not appropriate when nonlocal motions of the macromolecule are important, it greatly enhances the efficiency for the study of local effects, such as enzyme catalysis when the structure corresponds to the proper kinetic state.[11,12] For further efficiency, the inner region reaction field is expressed in terms of a Green's function, and the inner region charge distribution is projected onto a basis set. This allows precalculation of quantities that do not depend on the inner region coordinates using the Poisson−Boltzmann (PB) equations[13] and subsequent storage of these quantities in a reaction field matrix (vide infra).[10] Quantities that depend on the coordinates of the inner region can be evaluated quickly without having to repeatedly solve the PB equations; thus all minimizations or molecular dynamics can be done very efficiently once the reaction field matrix has been calculated. This makes the GSBP a particularly attractive method for MD simulations.

For geometry optimizations, however, the advantages of the precalculated reaction field matrix are less pronounced, as its calculation, depending on the number of basis functions chosen and the desired accuracy as such, usually is equally expensive as

**Received:** August 17, 2012
**Revised:** September 17, 2012
**Published:** September 18, 2012





dx.doi.org/10.1021/jp308218m | *J. Phys. Chem. B* 2012, 116, 12522−12534

solving for the PB equations several hundred times. Therefore, solving the PB equations in every step might be a faster approach. More importantly, the GSBP has not been well-integrated with combined quantum mechanical/molecular mechanical (QM/MM) approaches when ab initio methods are used as the quantum part. This is due to the fact that an accurate treatment of the interactions that involve the QM charge density requires the implementation of integrals involving the QM charge density with each electrostatic potential term and a GSBP basis function, respectively; although this can be done by taking advantage of the integral routines available in QM packages, especially when spherical harmonics are used for GSBP, the implementation has to be made deeply and separately into each QM code and is therefore not easily transferrable. In an approximate treatment, atomic charge representations for the QM atoms such as Mulliken charges or charges fitted to the electrostatic potential (ESP) could be employed in principle. For the calculation of gradients, however, it would be necessary to calculate the derivatives of these charges, which would have to be done by solving the coupled perturbed self-consistent field (CPSCF) equations[14] and would be prohibitively expensive for large systems.

To overcome these limitations, a modification of the GSBP method termed the solvent macromolecule boundary potential (SMBP) has been proposed and implemented recently by Benighaus and Thiel.[15] While the setup of the system as well as the objective of the method are identical to those of GSBP, the SMBP differs from the GSBP in two central aspects: First, the PB equations are solved in every step, avoiding basis set projection and calculation of a reaction field matrix. Second, all electrostatic interactions between the inner and outer regions are handled via a set of surface charges that are projected onto the boundary between those regions. The latter scheme allows for a simple and general interface with any QM code that can handle external point charges, and avoids the necessity to solve for the CPSCF equations in the evaluation of gradients. Most importantly, the SMBP can be used as a complement to GSBP in QM/MM calculations. GSBP MD simulations can be performed using either an all-MM setup or a semiempirical QM part such as the self-consistent charge density functional tight binding (SCC-DFTB) method,[16] and snapshots can be subsequently optimized with SMBP using a QM/MM setup with higher-level ab initio QM methods.

In the following, we present the implementation of the SMBP method into the program package CHARMM,[17] which has been interfaced with the semiempirical SCC-DFTB method[16] as well as the ab initio packages Q-Chem[19] and Gaussian.[19] Compared to the original work of Thiel and co-workers, our implementation takes advantage of the Poisson–Boltzmann module in CHARMM[20] and works for both spherical and rectangular box shaped boundaries; the latter is particularly useful for the study of transmembrane proteins and an analysis of effects due to transmembrane potential. After a short recapitulation of the basic theory for GSBP and SMBP, respectively, we give a detailed account of all of the necessary working equations for calculating the energy and forces within the SMBP in Section 2, which also includes treatment of excluded atoms in the QM/MM interface. In Section 4 we present results of benchmark tests for small model systems as well as an alkaline phosphatase (AP) family enzyme, including a vibrational analysis of the AP active site. Finally, we give an example of the rectangular box shaped boundary by studying the impact of membrane potential on a proton transfer reaction within a simple model membrane system using GSBP and SMBP.

## 2. THEORY

Both GSBP and SMBP aim for the efficient calculation of the electrostatic contribution to the solvation free energy, $\Delta W_{elec}$, providing an implicit solvent model ansatz for systems with complex (irregularly shaped) dielectric boundaries. In the general setup, the system is divided into an inner (i) and an outer (o) region (see Figure 1). Atoms of the macromolecule

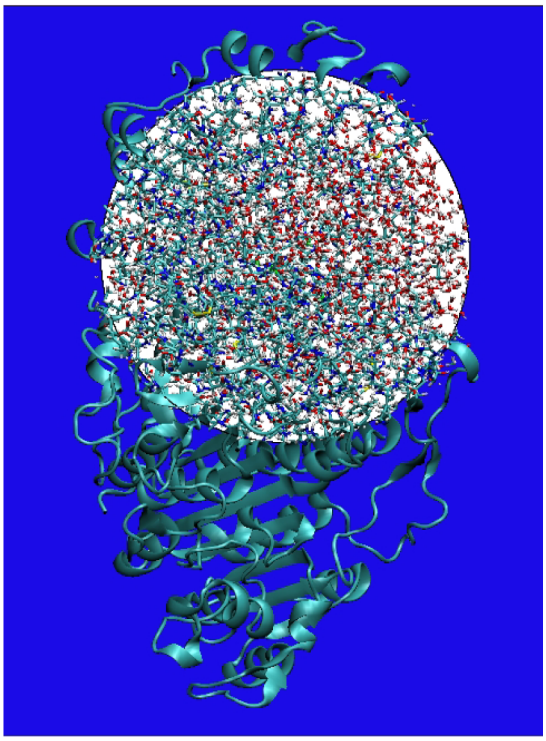


**Figure 1.** Setup as used for GSBP and SMBP with a spherical boundary. The inner region with explicit solvent is depicted with white background, the macromolecule in the outer region is shown as cyan ribbons, and the implicit solvent of the outer region is shown in blue.

are described explicitly in both regions, whereas the solvent is described explicitly only in the inner region and treated implicitly in the outer region (i.e., by setting the dielectric constant $\varepsilon$ in the respective solvent region to appropriate values). All atoms in the outer region are fixed, so that the potential of mean force (PMF) depends only on the coordinates of the inner region atoms of the macromolecule ($\mathbf{R}_i$) and the $n$ inner region solvent molecules, respectively; the precise number of solvent in the inner region should, in principle, be determined with a Grand Canonical scheme.[21] The PMF is then derived by integrating out all degrees of freedom of the outer region. As shown by Roux and co-workers,[10] it can be separated into the potential energy $U$ of the entire macromolecule and the $n$ inner solvent molecules, a configurational free energy restriction term $\Delta W_{cr}$ that arises due to the potential that prevents outer solvent molecules from entering the inner region, a nonpolar free energy interaction term $\Delta W_{np}$, as well as an electrostatic solvation free energy term $\Delta W_{elec}$, respectively:

$$W(\mathbf{R}_i, 1, ..., \mathbf{n}) = U(\mathbf{R}_i, 1, ..., \mathbf{n}) + \Delta W_{cr} + \Delta W_{np}(\mathbf{R}_i, \mathbf{1}, ..., \mathbf{n}) + \Delta W_{elec}(\mathbf{R}_i, \mathbf{1}, ..., \mathbf{n}) \quad (1)$$

(following the notation of Benighaus and Thiel[15]). The setup described above requires that the boundary of the inner region stays the same throughout the simulation/minimization. As outlined in refs 10 and 15, this requires that the atoms within the outermost layer of the inner region are kept fixed as well. Also, a constraining potential usually needs to be applied to the inner region solvent molecules to prevent them from crossing the boundary to the outer region.

The electrostatic part of the solvation free energy term, $\Delta W_{elec}$, is now separated into inner−inner (ii), inner−outer (io), and outer−outer (oo) contributions, where the latter is a constant due to the fixing of the outer region and can therefore be neglected. As shown below, it is advantageous to combine the solvation free energy term with the purely Coulombic charge−charge interaction term $U_{io}$ between the inner and the outer regions, so that the actual term calculated by the GSBP and SMBP (generally denoted as BP) is defined as:

$$\Delta W_{elec}^{BP} = U_{io} + \Delta W_{ii} + \Delta W_{io} \quad (2)$$

The free energy contributions $\Delta W_{xy}$ (with $x,y \in \{i,o\}$) can be represented as a sum over products of the atomic charges $q$ and a reaction field potential $\phi_{rf}^{y}$,

$$\Delta W_{xy} = \sum_{A \in x} q_A \phi_{rf}^{y}(\mathbf{r}_A) \quad (3)$$

where the reaction field potential is the difference between the electrostatic potentials in vacuum ($\phi_v$) and solution ($\phi_s$), respectively. Both potentials are computed by solving the linearized Poisson−Boltzmann (PB) equation:

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

($\rho(\mathbf{r})$ = charge density, $\varepsilon(\mathbf{r})$ = dielectric constant, $\bar{\kappa}$ = modified Debye−Hückel screening factor). As

$$U_{io} = \sum_{A \in inner} q_A \phi_v^{o}(\mathbf{r}_A) \quad (5)$$

both inner−outer terms can be combined to give

$$U_{io} + \Delta W_{io} = \sum_{A \in inner} q_A \phi_s^{o}(\mathbf{r}_A) \quad (6)$$

where $\phi_s^{o}$ is the static outer solvent potential, so that the final expression is

$$\Delta W_{elec}^{BP} = \sum_{A \in inner} q_A \phi_s^{o}(\mathbf{r}_A) + \Delta W_{ii} \quad (7)$$

As $\phi_s^{o}(\mathbf{r}_A)$ is constant for all configurations of the inner region,[10] it can be precomputed and reused for all subsequent simulation/minimization steps. We note here that the influence of the transmembrane potential can be included in $\phi_s^{o}(\mathbf{r}_A)$ by solving the PB-V equation[22] with the inner region charge set to zero (see Section 4.3 for an example). The remaining term $\Delta W_{ii}$, however, changes with every inner region configuration, and its treatment distinguishes the GSBP and SMBP methods (vide infra).

In the GSBP, a Green's function representation is used for the reaction field potentials within the inner−inner solvation free energy term $\Delta W_{ii}$.[10] The projection of the charge distribution of the inner region atoms onto a set of basis functions leads to[10]

$$\Delta W_{ii} = \frac{1}{2}\sum_{mn} Q_m M_{mn} Q_n \quad (8)$$

where the $Q_n$ are generalized multipoles, and $M_{mn}$ is a reaction field matrix which can be calculated from the Green's function and the basis functions, respectively.[10] Although the calculation of the reaction field matrix is rather expensive, it needs to be evaluated only once for a given simulation or minimization; only the generalized multipoles need to be updated, but in contrast to **M**, they can be obtained very quickly, as their calculation involves only a sum over products of the charge and the basis function $n$ (for $Q_n$) at the position of an inner region atom, summed over the inner region atoms. As mentioned in the introduction, this makes the GSBP a very efficient method for MD simulations.

For geometry optimizations, however, the precalculation of the reaction field matrix is less advantageous, as its calculation equals solving the PB equations several hundred times (depending on the number of basis functions used). Therefore, the SMBP avoids the use of a basis set and reaction field matrix and solves for the inner−inner energy term in the straightforward fashion

$$\Delta W_{ii} = \frac{1}{2}\sum_{A \in inner} q_A \phi_{rf}^{i}(\mathbf{r}_A) \quad (9)$$

Combining this with eq 6 and using the Green's function representation for the reaction field potentials leads to (shown for the QM/MM case)[15]

$$\Delta W_{elec}^{SMBP} = \frac{1}{2}\int[\rho^{QM}(\mathbf{r}) + \rho^{MM}(\mathbf{r})]G_{rf}(\mathbf{r},\mathbf{r}')[\rho^{QM}(\mathbf{r}') + \rho^{MM}(\mathbf{r}')]\,d\mathbf{r}\,d\mathbf{r}' \quad (10)$$

$$+\int[\rho^{QM}(\mathbf{r}) + \rho^{MM}(\mathbf{r})]\phi_s^{o}(\mathbf{r})\,d\mathbf{r} \quad (11)$$

with QM or MM inner region charge densities $\rho^{QM}$ and $\rho^{MM}$, respectively, and the Green's function $G_{rf}$. Separation of the factors and accounting for excluded atoms as used in CHARMM link-host schemes[23] yields

$$\Delta W_{elec}^{SMBP} = \frac{1}{2}\int\rho^{QM}(\mathbf{r})G_{rf}(\mathbf{r},\mathbf{r}')\rho^{QM}(\mathbf{r}')\,d\mathbf{r}\,d\mathbf{r}' + \int\rho^{QM}(\mathbf{r})G_{rf}(\mathbf{r},\mathbf{r}')(\rho^{MM}(\mathbf{r}') - \rho^{Ex}(\mathbf{r}'))\,d\mathbf{r}\,d\mathbf{r}' + \frac{1}{2}\int\rho^{MM}(\mathbf{r})G_{rf}(\mathbf{r},\mathbf{r}')\rho^{MM}(\mathbf{r}')\,d\mathbf{r}\,d\mathbf{r}' + \int\rho^{QM}(\mathbf{r})\phi_s^{o}(\mathbf{r})\,d\mathbf{r} + \int\rho^{MM}(\mathbf{r})\phi_s^{o}(\mathbf{r})\,d\mathbf{r} \quad (12)$$

(note that $G_{rf}(\mathbf{r},\mathbf{r}') = G_{rf}(\mathbf{r}',\mathbf{r})$). Switching back to the reaction field potential representation finally yields

$$\Delta W_{elec}^{SMBP} = \int\rho^{QM}(\mathbf{r})\phi_{tot}^{QM}(\mathbf{r})\,d\mathbf{r} + \int\rho^{MM}(\mathbf{r})\phi_{tot}^{MM}(\mathbf{r})\,d\mathbf{r} \quad (13)$$

where

$$\phi_{tot}^{QM}(\mathbf{r}) = \phi_s^{o}(\mathbf{r}) + \phi_{rf}^{MM}(\mathbf{r}) - \phi_{rf}^{Ex}(\mathbf{r}) + \frac{1}{2}\phi_{rf}^{QM}(\mathbf{r}) \quad (14)$$

$$\phi_{tot}^{MM}(\mathbf{r}) = \phi_s^{o}(\mathbf{r}) + \frac{1}{2}\phi_{rf}^{MM}(\mathbf{r}) \quad (15)$$


dx.doi.org/10.1021/jp308218m | J. Phys. Chem. B 2012, 116, 12522−12534

**Scheme 1. Algorithm of the Current SMBP Implementation within CHARMM**

```
Calculate φ_s^o and store on disk
For every step (e.g., minimization) do
    Calculate φ_rf^MM and φ_rf^Ex
    Assemble φ_tot^MM
    Calculate ∫ ρ^MM φ_tot^MM dr = Σ_{A∈inner} q_A φ_tot^MM
    Begin SCRF
      1. Initial guess for QM atomic charges, calculate
         initial φ_rf^QM
      2. Assemble φ_tot^QM and project onto surface charges
      3. Calculate QM wavefunction in field of surface charges
         and inner region MM charges (the latter contribute to
         U_ii)
         After convergence: Calculate QM charges
         (Mulliken or ESP)
      4. Calculate new φ_rf^QM based on new QM charges
      5. Check φ_rf^QM for convergence:
         YES: QM-Energy = U_ii^QM + ∫ ρ^QM φ_tot^QM dr
         NO: Go back to 2
    End SCRF
    Assemble total QM and MM gradient potentials
    Project total QM gradient potential onto gradient surface
    charges
    Calculate QM gradient contributions (no SCF) and add to
    total gradient
    Calculate MM gradient contributions and add to total
    gradient
Enddo
```

The reaction field potentials can be obtained by setting all charges except the ones the potential is calculated for to zero, and using the PB equation to calculate the difference between the electrostatic potentials in vacuum and solvent. In this respect, MM refers to the inner region MM atoms, QM to the (inner region) QM atoms, and Ex to excluded atoms. As all quantities except for $\phi_s^o$ change for every inner region configuration, the SMBP equations have to be solved for every step.

As mentioned above, an objective of the SMBP is to provide a simple and general interface to ab initio QM codes. This is accomplished by projecting the total QM potential $\phi_{\text{tot}}^{\text{QM}}$ onto a set of $N$ surface charges $q_i$, that are placed on the boundary dividing the inner and outer regions:[15]

$$\phi_{\text{tot}}^{\text{QM}}(\mathbf{r}_j) \approx \sum_i^N \frac{q_i}{r_{ij}} \tag{16}$$

Here, $r_{ij}$ marks the distance between the surface charge $q_i$ and atom $j$. As now the outer field and all reaction fields interacting with the QM charge density are projected onto the surface charges, all electrostatic interactions between the QM part and the rest (inner MM and outer region) are described as external charges to the QM part (which most QM codes can handle), thus avoiding having to modify the QM code used for the SMBP calculation. Furthermore, gradient calculations with a QM/MM boundary approach would normally require the solution of the CPSCF equations also in the case of SCF wave functions, as the derivatives of the QM (e.g., Mulliken) charges would have to be calculated.[15] The surface charge projection ansatz allows to avoid this in practical calculations (see below).

Because the unknown QM reaction field potential occurs in $\phi_{\text{tot}}^{\text{QM}}$, which is itself used to evaluate $\phi_{\text{rf}}^{\text{QM}}$, a self-consistent reaction field scheme (SCRF) has to be employed in solving for the QM reaction field potential, following an initial guess for the QM charges.[15] For details concerning the SCRF, see Scheme 1.

In the following, we will provide a more detailed account of the surface charge projection scheme. Abbreviating $\phi_{\text{tot}}^{\text{QM}}(\mathbf{r}_j)$ in eq 16 as $\phi_j$, the surface charges are optimized using a least-squares minimization scheme. We define a penalty function $f(\mathbf{q})$

$$2f(\mathbf{q}) = \sum_{j=1}^{N_{\text{QM}}} \left( \phi_j - \sum_i^N \frac{q_i}{r_{ij}} \right)^2 \tag{17}$$

When defining $\mathbf{A}$, $\mathbf{b}$, and $c$ as

$$c = \frac{1}{2} \sum_{j=1}^{N_{\text{QM}}} \phi_j^2 \tag{18}$$


dx.doi.org/10.1021/jp308218m | J. Phys. Chem. B 2012, 116, 12522−12534

$$\mathbf{b}^{\mathrm{T}} = \left( \sum_{j=1}^{N_{\mathrm{QM}}} \frac{\phi_j}{r_{1j}}, \cdots \sum_{j=1}^{N_{\mathrm{QM}}} \frac{\phi_j}{r_{Nj}} \right) \tag{19}$$

$$A = \begin{pmatrix} \sum_{j=1}^{N_{\mathrm{QM}}} \frac{1}{r_{1j}r_{1j}} & \cdots & \sum_{j=1}^{N_{\mathrm{QM}}} \frac{1}{r_{1j}r_{Nj}} \\ \vdots & \ddots & \\ \sum_{j=1}^{N_{\mathrm{QM}}} \frac{1}{r_{Nj}r_{1j}} & \cdots & \sum_{j=1}^{N_{\mathrm{QM}}} \frac{1}{r_{Nj}r_{Nj}} \end{pmatrix} \tag{20}$$

$f$ can be written as a quadratic form,

$$f(\mathbf{q}) = \frac{1}{2}\mathbf{q}^{\mathrm{T}}\mathbf{A}\mathbf{q} - \mathbf{b}^{\mathrm{T}}\mathbf{q} + c \tag{21}$$

For a symmetric and positive definite matrix $\mathbf{A}$, $f(\mathbf{q})$ is minimized by the solution to[24]

$$\mathbf{A}\mathbf{q} = \mathbf{b} \tag{22}$$

which in our implementation is accomplished using the conjugate gradient algorithm.[24]

While $\mathbf{A}$ is clearly symmetric, we have currently no formal proof that it is always positive definite. As a check, the maximum and mean average differences between $\mathbf{Aq}$ and $\mathbf{b}$ are printed out after solving for $\mathbf{q}$. Also, the errors between the exact total QM potential at the QM atom positions and the approximated potential are routinely evaluated and printed out.

For the calculation of gradients, one has to distinguish whether the derivative of the energy term in eq 13 is taken with respect to the coordinate of a QM or an MM atom, respectively. As shown in ref 15, taking the derivative of $\Delta W_{\mathrm{elec}}^{\mathrm{SMBP}}$ in eq 12 with respect to a QM atom coordinate yields

$$\frac{\partial}{\partial x_{\mathrm{QM}}}\Delta W_{\mathrm{elec}}^{\mathrm{SMBP}} = \int \phi_{\mathrm{tot_{QM}}}^{\mathrm{grad}}(\mathbf{r}) \cdot \frac{\partial \rho_{\mathrm{QM}}(\mathbf{r})}{\partial x_{\mathrm{QM}}} \mathrm{d}\mathbf{r} \tag{23}$$

where the total QM gradient potential is

$$\phi_{\mathrm{tot_{QM}}}^{\mathrm{grad}}(\mathbf{r}) = \phi_{\mathrm{s}}^{\mathrm{o}}(\mathbf{r}) + \phi_{\mathrm{rf}}^{\mathrm{MM}}(\mathbf{r}) - \phi_{\mathrm{rf}}^{\mathrm{Ex}}(\mathbf{r}) + \phi_{\mathrm{rf}}^{\mathrm{QM}}(\mathbf{r}) \tag{24}$$

The total QM gradient potential is now projected onto a set of $K$ gradient surface charges,

$$\phi_{\mathrm{tot_{QM}}}^{\mathrm{grad}}(\mathbf{r}_j) \approx \sum_{p}^{K} \frac{q_p}{r_{pk}} \tag{25}$$

If the QM wave function is calculated in the presence of the gradient surface charges, the gradient contribution in eq 23 is automatically added to the QM gradient.

An important caveat is that the total QM gradient potential differs from the total QM potential by the prefactor to the QM reaction field potential, which leads to a difference between the energy and gradient surface charges (compare eqs 14 and 24). This means that the QM wave function is not converged in the field of the gradient surface charges, as it has been converged in the presence of the energy surface charges. Due to the very small contribution of the QM reaction field potential to the total potential, however, the occurring off-diagonal Fock matrix elements are very small and can be neglected in practical calculations, as shown in ref 15.

Taking the derivative of $\Delta W_{\mathrm{elec}}^{\mathrm{SMBP}}$ in eq 12 with respect to an MM atom coordinate yields

$$\frac{\partial}{\partial x_{\mathrm{MM}}}\Delta W_{\mathrm{elec}}^{\mathrm{SMBP}} = \int \phi_{\mathrm{tot_{MM}}}^{\mathrm{grad}}(\mathbf{r}) \cdot \frac{\partial \rho_{\mathrm{MM}}(\mathbf{r})}{\partial x_{\mathrm{MM}}} \mathrm{d}\mathbf{r} - \int \phi_{\mathrm{rf}}^{\mathrm{QM}} \cdot \frac{\partial \rho_{\mathrm{Ex}}(\mathbf{r})}{\partial x_{\mathrm{MM}}} \mathrm{d}\mathbf{r} \tag{26}$$

where $\phi_{\mathrm{tot_{MM}}}^{\mathrm{grad}} = \phi_{\mathrm{tot_{QM}}}^{\mathrm{grad}} + \phi_{\mathrm{rf}}^{\mathrm{Ex}}$. The evaluation of the derivatives over MM atom charge densities is straightforward.[10,15]

Concerning the timing performance of the SMBP, the overall runtimes depend strongly on the size of the QM region, as well as on the mesh size for the PB grid and the size of the inner region. The latter dependence is weaker, however, if the QM calculation is time-determining. Compared to a standard QM/MM calculation with a time determining QM step and a QM region of the same size, the SMBP can be expected to be about two to four times more costly, since for every energy step the SCRF iterations have to be converged, which usually take two to three iterations if the surface charges of the previous energy step are used as the initial guess. For every SCRF iteration, meanwhile, a full QM as well as a full PB calculation have to be performed.

In Scheme 1, the current algorithm as implemented into the CHARMM c36a2 development version[23] is outlined.

## 3. MODELS AND COMPUTATIONAL DETAILS

**3.1. Benchmark Models.** To investigate the convergence and evaluate the performance of the SMBP approach, we study several systems here. First, we establish a simple model system that contains a phosphate, a serine (zwitterionic form), 10 water molecules, and 6 additional "nitrogen" cations (as outer region charges), where the latter are placed on the faces of a cube surrounding the phosphate, serine, and water molecules with a distance of 24 Å from the center. Except for the nitrogen cations, this system represents a cutout of the respective molecules from a wild type AP structure (vide infra). This model, which will be termed the "Serine Phosphate model" in the following, has highly charged components and comprises distinct inner and outer regions, yet is at the same time small (55 atoms), which makes it convenient for benchmarking boundary potentials. The more realistic AP system is also studied here, and the inner/outer partitioning follows our recent study[25] of these enzymes using SCC-DFTBPR/MM-GSBP.

Finally, we study proton transfer in a highly simplified "ion channel" embedded in an implicit membrane. The model was established in our previous work,[26] and we use it here to illustrate the effect of membrane potential on chemical reactions that occur in membrane proteins.[27]

**3.2. Computational Details.** For all calculations, the CHARMM version used is 36a2,[23] except for the proton transfer study, for which we use a locally modified 32a2. The CHARMM 22 force field[28] is employed throughout for the MM part. For the SCC-DFTB calculations, the SCC-DFTBPR parametrization[29] is employed; the exception is for the proton transfer study in a model membrane environment, for which we use the standard second-order SCC-DFTB.[16] All ab initio calculations are performed with the B3LYP functional[30] and the 6-31G* basis set,[31] using the QM packages Q-Chem 3.1[18] or Gaussian 09.[19] The DFT numerical integration grid has 75 angular and 302 radial points in the Q-Chem 3.1 calculations, whereas for Gaussian 09 the standard grid ("FineGrid") is used. For the AP frequency calculations we use Q-Chem 3.2[18] with

the SG-1 integration grid.[32] In the case of link atoms, the "div" scheme[33] is used throughout.

All Poisson−Boltzmann calculations are done using the successive over relaxation (SOR) method for the linearized PB equation[34] implemented in CHARMM,[20] employing trilinear charge interpolation. Unless stated otherwise, boundary potentials are obtained from larger grids using a focusing scheme,[35] where the grid spacings for the small/large grids are chosen as 0.4/1.2 Å. The dielectric constants for protein and implicit solvent are set to 1 and 80, respectively, with exceptions noted in the text.

For the SMBP calculations, 100 surface charges are used except where noted otherwise. The SCRF is considered converged when the norm of the difference vector between the actual and the QM reaction field potential of the previous cycle is smaller than $1 \times 10^{-3}$ Å. For the initial SCRF step, charges on the QM atoms are either set to zero for the first step of a minimization or the QM charges from the previous optimization step are used as initial guess. The maximum errors in the potential at the QM atom positions as approximated by the surface charges are generally smaller than $1 \times 10^{-3}$ Å.

## 4. RESULTS AND DISCUSSION

**4.1. Benchmark Results.** As a thorough evaluation of the general performance of the SMBP has already been conducted in refs 15 and 36, we provide here benchmark results for some aspects that have not been covered by refs 15 and 36. The discussions regarding the technical details for the determination of surface charges are included in the Supporting Information.

*4.1.1. Quality of ESP Charges.* One issue worth considering for SMBP is the performance of different approaches for evaluating QM atomic charges, since these charges directly affect the accuracy of the QM reaction field potential as calculated by the PB solver. These QM charges are fitted so as to reproduce the QM electrostatic potential (ESP) at some distance from the QM part, but since there is no unique way of doing so, many different charge fitting algorithms exist.[37] The program package Q-Chem uses a special way for the ESP charge fitting, in that it does not fit to the ESP directly, but rather to the multipole moments of the QM molecule. The respective charges are called multipole derived charges (MDC);[38] since they are fitted to a certain number of multipole moments (depending on the number of QM atoms), no grid is necessary for the charge fitting. Moreover, since the multipole moments are long-range in nature, the long-range accuracy of these charges can be expected to be superior to other ESP schemes that fit the ESP only within a certain region outside the QM part.

Figure 2 shows deviations in the ESP reproduced by different ESP charges with respect to the exact QM ESP, for a model system comprising *N*-methyl acetamide (NMA), and four water molecules that are placed in the plane of the NMA (for further details see caption of Figure 2) as probes. Results for two different average distances between the water molecules and the respective closest atom of the NMA are shown, and maximum and mean average errors in the reproduced potential are calculated at the positions of all water atoms. The Gaussian ESP charges perform best for the smaller distance, where the performance is significantly better than for Q-Chem's MDC scheme. For the larger distance, however, the MDC charges perform slightly better than Gaussian's ESP's, hinting at an improved long-range behavior. The GAMESS ESP charges, once the quadrupole moment is taken into consideration

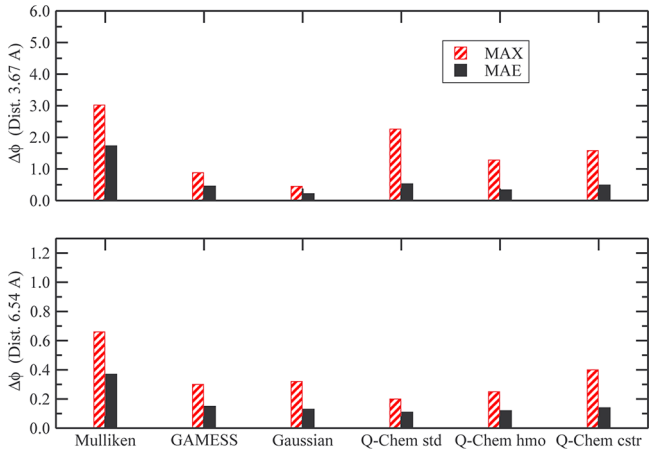



**Figure 2.** Mean average (MAE) and maximum (MAX) errors in the electrostatic potential (in [kcal/mol/unit charge]) for the potential at the atom positions of the four water molecules in the *N*-methyl acetamide (NMA) water model system, using different charge representations for the NMA atoms, with respect to the exact QM potential (all calculations done using HF/6-31G*). In the model, the water molecules are placed in the plane of the NMA, and results are shown for two average distances of the water oxygens to the respective nearest atoms of the NMA molecule. Details of the program packages and ESP charge algorithms are as follows: GAMESS v. 1 Oct 2010 (R3)[55] (ChelpG[56]), Gaussian v. G09RevB.01 12-Aug-2010[19] (MK,[46] yet the Gaussian ChelpG scheme gives results similar to MK), Q-Chem v. 3.1[18] (std: Multipole Derived Charge (MDC) scheme,[38] hmo: MDC including fitting to the next higher multipole order (see Supporting Information), cstr: same as hmo, but including constraints on the magnitude of the charges (see Supporting Information)).

during fitting, perform similarly to Gaussian ESP charges, especially at the longer distance.

The comparatively poor performance of the MDC charges at the shorter distance can be explained as follows: Since in the MDC algorithm no constraints are placed on either the magnitude or the sign of the charge for a given atom, the MDC charges can become fairly large and "unphysical",[38] leading to substantial errors at short distances, as shown in Figure 2. In fact, for large QM regions encountered in real enzyme applications, we find that the MDC charges can be very large (e.g., > 10(!)), which not only compromises the quality of the surface charges in SMBP but also hampers the convergence of the PB equations. Therefore, we have developed approaches that help limit the magnitude of the MDC charges; the details are given in the Supporting Information. As shown in Figure 2, these algorithms (labeled as "Q-Chem hmo" and "Q-Chem cstr") improve the performance of MDC charges in the short range without significantly affecting the performance in the long range.

*4.1.2. Convergence of SMBP and GSBP with Different Shapes of Inner Regions.* Next we turn to a comparison of GSBP and SMBP results using an SCC-DFTBPR/MM setup; this serves as a check for our SMBP implementation and also illustrates the convergence behavior of GSBP. Figure 3 shows differences in energies and root-mean-square (rms) gradients between GSBP and SMBP for the Serine Phosphate model system using a spherical boundary (100 surface charges), with respect to different inner region sizes and a varying number of basis functions used in the GSBP. In the model, the phosphate was chosen as the QM region; the number of atoms in the inner and outer regions is the same for all the shown size


dx.doi.org/10.1021/jp308218m | J. Phys. Chem. B 2012, 116, 12522−12534

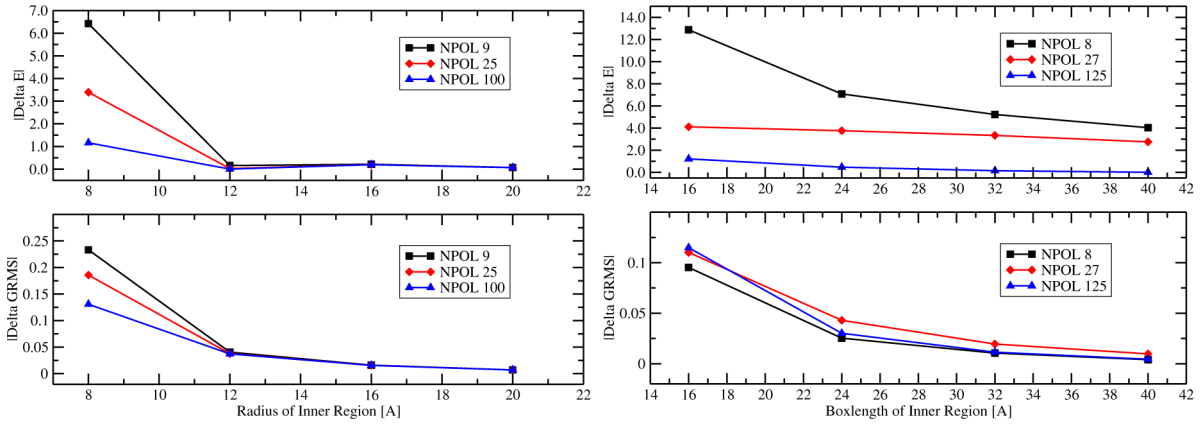



**Figure 3.** Absolute SCC-DFTBPR energy ([kcal/mol]) and rms gradient differences ([kcal/mol/Å]) between GSBP and SMBP for the Serine Phosphate model system, with respect to the size of the inner region and the number NPOL of GSBP basis functions. The left panels are for a spherical inner region and the right panels for a cubic inner region.

variations, as the serine, phosphate, and water molecules are within the 8 Å sphere and the 6 $N^+$ cations are outside the 20 Å sphere. For the largest inner region sphere, the GSBP and SMBP give practically identical results, even for the smallest number of basis functions. When going to smaller inner region sizes, however, the results obtained for both methods become more and more different, and also the influence of the number of GSBP basis functions becomes stronger, particularly for the smallest sphere. This is due to well-known boundary effects that occur when the boundary moves close to the inner region atoms.[9,10] In that case, a large number of basis functions is needed for the GSBP to converge toward the exact PB results,[10] which are given here by the SMBP in that it constitutes the large basis limit of the GSBP. A second effect not apparent in Figure 3 is the fact that also the numerical PB results will deteriorate from analytical results as given by the Kirkwood reaction field for a spherical cavity, when an atom (charge) comes very close to the boundary.[10]

Also shown in Figure 3 are corresponding results for a rectangular box setup (98 surface charges), where the chosen boxlengths correspond to the diameters of the spherical boundary of the previous test. For the largest inner region, both GSBP and SMBP yield the same result for the largest number of GSBP basis functions; however, for the energy the results are much more influenced by the basis function number, even for the largest boxlength. This is presumably due to the fact that more nonuniform basis functions are needed for the rectangular box case, as compared to the spherical case to describe the charge distribution in the inner region.[10] Interestingly, the gradient shows much less influence on the number of basis functions, where the total deviation between the SMBP and GSBP results is significantly smaller than for the spherical case. Here, the results for eight basis functions seem to give better results as the calculation with 125 functions; yet due to the small overall differences this can be attributed to numerical effects.

Overall, the GSBP and SMBP results converge for large inner regions as well as a sufficient number of basis functions. Although we have not shown results for systems containing link atoms, a validation for our excluded atoms scheme is provided in the next section, where we compare GSBP and SMBP results for AP using SCC-DFTBPR (see, e.g., Figure 5).

**4.2. Application to AP Family Enzymes.** The alkaline phosphatase (AP) family of enzymes catalyze phosphoryl and sulphuryl transfer reactions, which play a very important role in biology. They form prominent examples for bimetallic catalysis, as their active sites contain a bimetallo zinc motif.[39] Moreover, these enzymes show catalytic promiscuity,[40] which makes them an especially interesting subject for many experimental[41,42] and computational[25,43] studies on the nature of enzymatic catalysis. The objective of our current work, however, is not to provide a detailed study of AP systems, but rather to perform some fundamental benchmark calculations on a "real" and challenging system with respect to the influence of the SMBP. As a benchmark and initial application, we carry out QM/MM calculations on the structure of wildtype (wt) AP containing phosphate or hydrogenphosphate (see Supporting Information) and on the hydrolysis of methyl *p*-nitrophenyl ($MpNPP^-$) phospho-diester in the R166S AP mutant.[25] We will focus on geometrical and energetical influences of the SMBP by comparing optimized structures as well as results for a minimum energy path (MEP) calculation. Finally, we also investigate the effect of the SMBP on vibrational frequencies. The effect of grid settings as used for the PB solver is discussed in Supporting Information; based on the analysis, we chose a mesh size combination of 0.4/1.2 Å for the actual and focusing grids, respectively, in all subsequent calculations ($\mathbf{r}_{inner}$ = 27 *A*).

*4.2.1. Structure and Energetics of Phosphoryl Transfer in R166S-AP.* For the AP R166S mutant containing $MpNPP^-$ in the active site, comparisons of structures with respect to the influence of the SMBP or the chosen QM method are shown in Figure 4. The structures shown correspond to the reactant, transition state, and product of a concerted nucleophilic substitution reaction, where the serine alkoxide group attacks the phospho-diester and replaces the *p*-nitrophenyl group; these structures are obtained with adiabatic mapping that is used to calculate an approximate MEP[44] (see below).

With B3LYP/6-31G* as the QM method, no major structural differences are found for geometry optimizations with and without SMBP; the rmsd values are 0.04, 0.09, and 0.12 Å for the reactant, transition state, and product, respectively. The optimized Zn−Zn distances do not deviate by more than 0.02 to 0.03 Å for any of the structures, although the alignment of the serine and *p*-nitrophenyl groups changes from an almost perfect match in the reactant to having a significantly larger deviation in the product state. Comparing the structures optimized with the SCC-DFTB and the B3LYP/6-31G* methods (both using SMBP), the deviations are


dx.doi.org/10.1021/jp308218m | *J. Phys. Chem. B* 2012, 116, 12522−12534

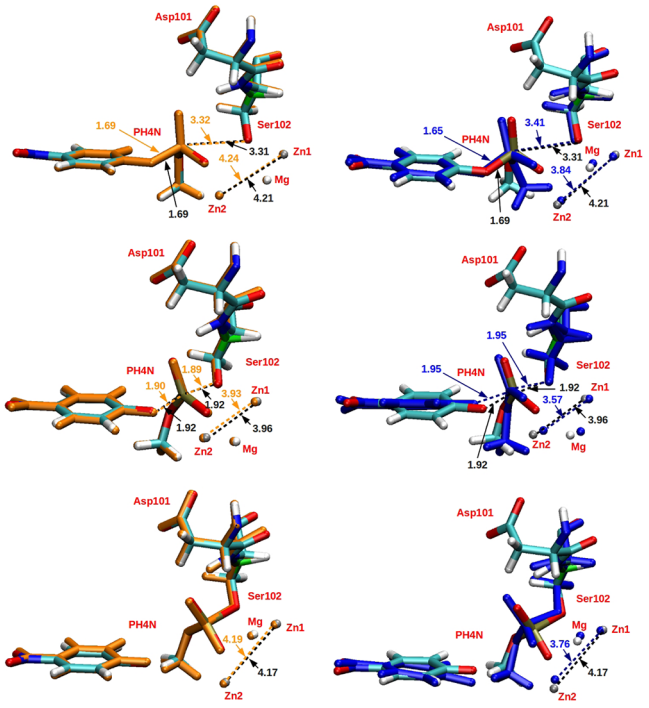



**Figure 4.** Left column: overlay of B3LYP/6-31-G*/MM optimized structure with and without the SMBP (orange) for the hydrolysis of methyl *p*-nitrophenyl ($MpNPP^-$) phospho-diester in R166S AP. Right column: overlay of SCC-DFTBPR/MM (blue) and B3LYP/6-31G*/MM structures with the same SMBP. The top, middle, and bottom rows correspond to reactant, transition state, and product, respectively. Bond lengths are shown in Å. The RMSDs between B3LYP/MM structures with and without SMBP are between 0.04 and 0.12 Å; those between SCC-DFTBPR/MM-SMBP and B3LYP/MM-SMBP are between 0.33 and 0.51. Explicit TIP3P waters are not shown.

substantially larger in all cases; whereas the reactant and product states show similar RMSDs of 0.34 and 0.33 Å, respectively, the transition state shows an even larger deviation with an rmsd of 0.51 Å. The Zn−Zn distances are smaller by ~0.4 Å for the SCC-DFTB method, although the change in the distance when going from reactant via the transition state to the product behaves quite similarly as compared to B3LYP. The impact of these differences on the reaction energetics is discussed below.

As to the energetics, Figure 5 shows the full MEP as calculated with GSBP (400 basis functions) and SMBP using an SCC-DFTBPR/MM setup. In this case, the starting structure is taken from the crystal structure (PDB code 3CMR[45]), and forward and backward scans along the reaction coordinate (distance difference of the P−O distances with respect to the attacking and leaving groups, respectively) with a step size of 0.1 Å are performed using constrained optimizations, until convergence of the profile is achieved. Both profiles correspond to the fourth forward scan, that is, going from reactant to product. For the optimizations, a convergence criterion of 0.05 kcal/(mol·Å) with respect to the rms gradient is applied. GSBP and SMBP give very similar profiles, with a maximum energy difference of roughly 1 kcal/mol, so that the GSBP calculation can be viewed as converged with respect to the number of basis functions. Although we have no accurate data concerning timings, as the calculations are carried out on different machines with different loads, our data show that in this case the GSBP is by almost 2 orders of magnitude faster than the

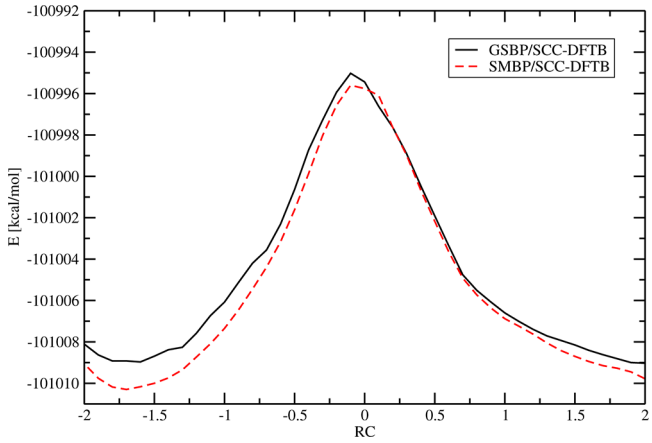



**Figure 5.** Comparison of the SCC-DFTBPR/MM-GSBP and SCC-DFTBPR/MM-SMBP energy profile for the transesterification reaction of the $MpNPP^-$ phospho-diester within the R166S AP active site. Energy differences: $\Delta E$ (ts − reac) = 14.0 kcal/mol (GSBP) vs 14.7 kcal/mol (SMBP); $\Delta E$ (ts − prod) = 14.0 kcal/mol (GSBP) vs 14.2 kcal/mol (SMBP).

SMBP calculation for the MEP, since the reaction field matrix needs to be calculated only once for the entire profile. This is not surprising, since 41 geometry optimizations (each requiring roughly 20 optimization steps) have to be carried out for each scan, yet it underlines the superiority of the GSBP ansatz using precalculated data for cases when many energy/force steps need to be calculated.

Table 1 shows corresponding energy barriers with respect to reactant and product, respectively, after further optimizations with a tighter gradient threshold (0.01 kcal/(mol·Å)). The corresponding Zn−Zn distances are also given (see also Figure.4). The SCC-DFTBPR/MM-SMBP barriers differ by a maximum of 0.6 kcal/mol from the reaction path results as shown in Figure 5; this shows that a tighter optimization does not have a large effect on the barriers. Comparison with the B3LYP data in Table 1, however, reveals that the barrier measured relative to the reactant is substantially different for the two QM methods; the reactant is destabilized by almost half the amount of the SCC-DFTBPR barrier in the B3LYP calculations. In contrast, the barrier difference with respect to the product is less significant.

As a check for the ESP charges used in our SMBP interface with Q-Chem, we repeat the calculation using the SMBP/Gaussian interface (B3LYP G09 in Table 1), which uses ESP charges using the Merz−Kollmann algorithm[46] as opposed to Q-Chem's MDC method[38] including our modifications (see Supporting Information). In addition, the DFT grids used are not exactly the same for Gaussian and Q-Chem. Nevertheless, the results are very similar for both program packages, with barrier differences within 0.4 kcal/mol and differences in the Zn−Zn distances within 0.01 Å. We thus conclude that the actual choice of the ESP charges does not exert a large influence on the SMBP results.

Finally, we note that the barriers calculated without using the SMBP are not meaningful, as the reactant is calculated to be more stable than the "transition state" (with the reaction coordinate constrained to be zero). This occurs despite the fact that the differences between structures optimized with and without the SMBP are not very large (see above). Single point energy calculations without SMBP for the SMBP optimized structures and respective calculations using the SMBP for the


dx.doi.org/10.1021/jp308218m | J. Phys. Chem. B 2012, 116, 12522−12534

Table 1. Energy Differences[a] (in kcal/mol) between Transition State (ts), Reactant (reac) and Product (prod), Respectively, for AP R166S Optimized with Different Methods (See Text for Details), and Zn−Zn Distances (Å) of the Optimized Structures

| | | | | | | Zn−Zn distance | | |
|---|---|---|---|---|---|---|---|---|
| | | ΔE(ts − reac) | | ΔE(ts − prod) | | reac | ts | prod |
| SMBP | SCC-DFTBPR | 14.9 | (14.6) | 13.6 | (13.6) | 3.8 | 3.6 | 3.8 |
| | B3LYP Q-Chem | 7.9 | (6.7) | 11.0 | (13.2) | 4.2 | 4.0 | 4.2 |
| | B3LYP G09 | 7.5 | | 10.6 | | 4.2 | 4.0 | 4.2 |
| no BP | SCC-DFTBPR | −3.8 | (−8.9) | 18.6 | (17.7) | 4.0 | 3.6 | 3.8 |
| | B3LYP Q-Chem | −10.2 | (−6.3) | 6.6 | (18.2) | 4.2 | 3.9 | 4.2 |

[a]Numbers in parentheses denote results from single point energy calculations on the same respective optimized structures, but (i) without using the SMBP for the SMBP structures and (ii) using the SMBP for the "no BP" structures.

structures optimized without boundary potential yield results (shown in parentheses) that do not change the trend significantly. Apparently, the product state is overstabilized without the boundary potential, leading to essentially a largely downhill process going from the reactant to the product state. These observations emphasize the importance of including solvent effects in the calculation of such enzymatic reaction profiles.

*4.2.2. Vibrational Frequencies of the AP Active Site.* As a final case study concerning the AP systems, we investigate the influence of the SMBP on vibrational frequencies, which are calculated using normal mode analysis, that is, using the harmonic approximation. Wildtype AP and a Ser102G mutant have been subject to an experimental study of the phosphoryl (and sulfuryl) vibrations in the active site by measuring the infrared (IR) spectra as conducted by the Herschlag group.[41] Based on their results, Herschlag et al. proposed a model that links the AP ground state stability to the presence of transferable protons on the phosphate/sulfate within the active site.[41] To aid the interpretation of the IR spectra and thus to test the validity of the model, we perform vibrational calculations on wt AP, using several starting structures with different protonation states of the phosphoryl and the adjacent serine group, respectively. Here we focus on initial results that deal with the feasibility of frequency calculations on such systems, the influence of long-range solvation effects, that is, the SMBP, as well as the effects of different starting structures on the phosphate frequencies.

As SMBP second derivatives are not part of our current implementation, all frequencies are calculated using numerical finite differences of analytical first derivatives, for which the application of the SMBP is straightforward. However, as the number of required gradient evaluations becomes large for a large number of atoms, it is necessary to restrict the number of degrees of freedoms by fixing most parts of a large system. This approach will only work for frequencies that are sufficiently localized around a respective group. Since we are focusing on vibrations of the phosphate group, which consists of P−O stretch and bending vibrations with frequencies lying generally above 1000 $cm^{-1}$, the requirement of locality should be fulfilled (see discussions in Supporting Information); for all subsequent calculations, an unfixed sphere size of 4 Å (by residue) is chosen around the phosphate. We note that, for local modes of a fairly large QM region, computing the finite difference rather than the analytical Hessian might in fact have several practical advantages. First, solving CPSCF equations[47,48] for large QM/MM calculations can be both memory- and compute-intensive, while this is avoided by doing finite difference. Second, it is straightforward to restart finite difference calculations, which is valuable for computational environments with stringent wall time limitations.

All calculations have been done using B3LYP/6-31G* for the QM part. The definitions for the boundary potential are the same as described for the AP structural study above, that is, using a spherical boundary region with an inner region radius of 27 Å. Starting structures are taken from the crystal (PDB code 3TG0[49]) but prepared in different ways (see also captions to Figure 6): Two with deprotonated Ser102 and hydrogenphosphate, where in the first structure the phosphate hydrogen is oriented away from, and in the second it is oriented toward the catalytic serine group (Ser 102). Both structures contain 15 421 atoms with 70 QM atoms. The remaining structure contains deprotonated serine and phosphate. To better relax the region around the active site, the structures are preconverged using B3LYP/MM (without SMBP) with a convergence criterion of 0.01 kcal/(mol·Å) with respect to the rms gradient, where the residues outside a sphere of 8 Å around the active side are fixed. Subsequently, B3LYP/MM optimizations with and without the SMBP are performed.

For the finite difference calculations, a step size of 0.005 Å is chosen throughout. Structures are first optimized with the rms gradient smaller than 0.05 kcal/(mol·Å). For one of the AP structures we compare the frequencies with those optimized with a tighter criterion for the rms gradient (<0.02 kcal/(mol·Å)), and no significant differences in the results are observed. Significant phosphate frequencies are identified by calculating the rms fluctuations for the respective phosphate group and choosing frequencies with a fluctuation larger than 0.001.

The final optimized geometries of the active site are shown in Figure 6, along with bond distances in Å for the SMBP and standard ("no BP") optimization (in parentheses). The top panel shows a structure ("Cov") where the phosphate is covalently bound to the serine. This is due to a spontaneous reaction that occurs during the optimization, where the serine alkoxide group attacks the hydrogen phosphate and replaces the $OH^-$ as the leaving group. The structure "$PO_4$ SerOH" is based on almost the same starting structure, except that the phosphate hydrogen is oriented toward the serine instead of away from it, so that no nucleophilic attack occurs. The phosphate proton spontaneously transfers to the serine, as postulated by Herschlag and co-workers. Finally, we also explore a structure ("$PO_4$ $SerO^-$") in which both phosphate and Ser102 are deprotonated. Here, the serine alkoxy group is clearly turning away from the phosphate, and the Zn2 ion moves somewhat between the two closest oxygens of both groups, with a short Zn-$SerO^-$ distance of 1.88 Å. Due to the high concentration of negative charge, and the fact that one proton is expected to be bound to either the serine or the

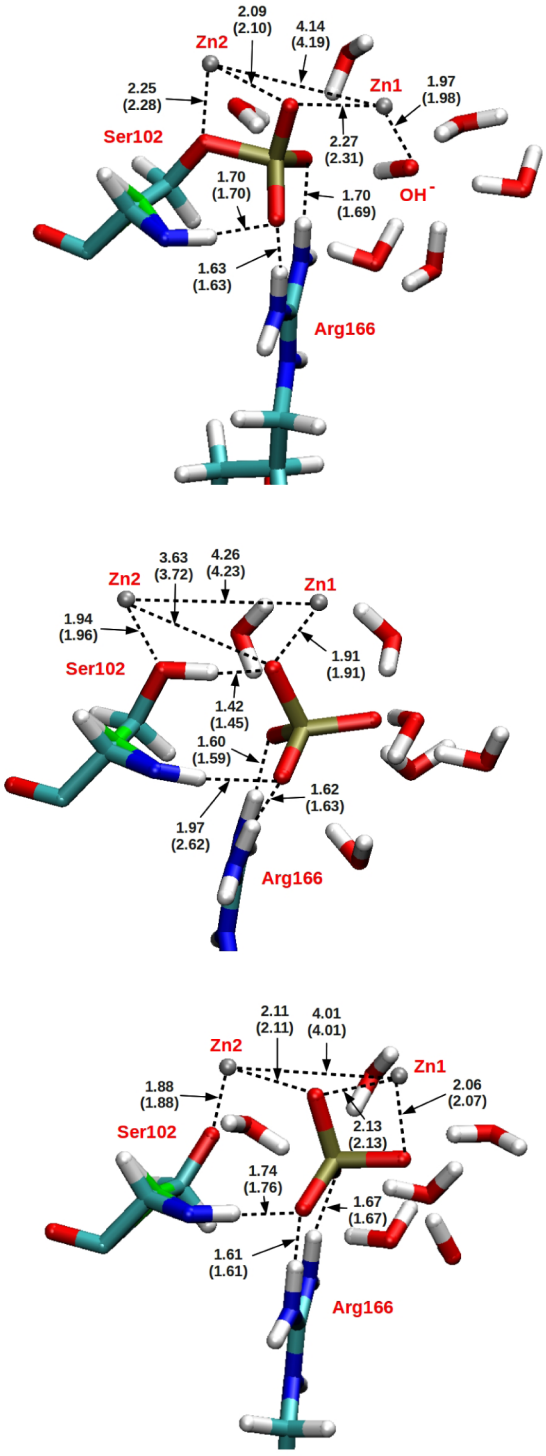



**Figure 6.** B3LYP/6-31G*/MM-SMBP optimized structures for different phosphate-AP complexes: "Cov" with the phosphate covalently bound to the serine group, "$PO_4$ SerOH" with phosphate and protonated serine, and "$PO_4$ $SerO^-$" with phosphate and deprotonated serine. Bond lengths are shown in Å, and those given in parentheses are obtained from an analogous optimization without using the SMBP. For the first two, the starting structure contained a deprotonated serine and hydrogen-phosphate, with the phosphate proton oriented either away (for "Cov") or toward (for "$PO_4$ SerOH") the serine. For "$PO_4$ $SerO^-$", the starting structure contained deprotonated serine and phosphate.

phosphate under the experimental conditions, we do not expect this structure to be a realistic model for the calculation of spectra, yet we include it to test wether the calculated frequencies for this structure can be computationally distinguished from the other optimized structures.

As already seen for the AP R166S mutant, the SMBP has generally only a small influence on all geometries shown here. The only exception is the $PO_4$ SerOH structure, where the bond length between the amino hydrogen of the serine and a phosphate oxygen differs substantially by 0.65 Å.

Table 2 shows the scaled harmonic frequencies for the three structures presented above, calculated with and without using

**Table 2. Frequencies[a] for the Phosphate Group in the wt AP Active Site, As Calculated for Three Different Optimized Structures (See Text) and As Measured by IR Spectroscopy in Solution[41]**

| Cov | | $PO_4$ SerOH | | $PO_4$ $SerO^-$ | | |
|---|---|---|---|---|---|---|
| no-BP | SMBP | no-BP | SMBP | no-BP | SMBP | exp |
| | | | | 1022.7 | 1022.7 | 1015 |
| | | 1046.1 | 1065.4 | 1057.1 | 1059.6 | |
| | | 1047.4 | 1067.8 | | 1061.5 | 1072[b] |
| 1096.4 | 1095.5 | 1082.3 | 1079.4 | | | 1085[c] |
| 1102.4 | 1108.1 | | | | | 1107[b] |
| 1111.8 | 1110.1 | | | | | |
| | | | | | | 1138[b] |

[a]The calculated QM/MM frequencies are obtained using B3LYP/6-31G* as the QM method and are scaled by a factor of 0.9614.[54] "no BP" indicates calculations without using any boundary potential. Only calculated frequencies above 1000 $cm^{-1}$ are shown. [b]Only pH 5. [c]Only pH 8.

the SMBP. Similar to the model system studied in the Supporting Information, the effect of SMBP on the calculated vibrational frequencies is rather small, especially for the structures Cov and $PO_4$ $SerO^-$, for which the corresponding influence on the structures themselves is small as well. Deviations are larger with a maximum of roughly 20 $cm^{-1}$ for $PO_4$ SerOH, likely due to the slightly larger deviation in the underlying optimized structures. The frequencies are compared to the experimental values as evaluated by measuring the difference spectra of regular and $O^{18}$ substituted phosphate in solution, respectively, for either a pH value of 5 or 8.[41] Whereas only noncovalently bound phosphate with respect to Ser102 is expected to exist at pH 8, a mixture of covalently and noncovalently bound phosphate is anticipated at lower pH, where the peak with the highest wavenumber (1138 $cm^{-1}$) was assigned to the covalent species.[41] Although an exact assignment is not possible based on our results, this trend can be seen in the calculated data, where the covalently bound structure has the highest phosphate frequencies. However, a more conclusive assignment requires sampling of different active site conformations, and at this stage it is premature to draw any firm statement regarding the protonation state of Ser102 or phosphate. This is particularly evident when comparing the calculated spectra for $PO_4$ SerOH and the rather unrealistic $PO_4$ $SerO^-$ structure, where both spectra are hardly distinguishable considering the uncertainty of calculated frequencies. Thus, a test of the validity of the model proposed by Herschlag et al.[41] requires more thorough calculations, especially more extensive sampling of the active site. The preliminary calculations performed here do highlight, never-


dx.doi.org/10.1021/jp308218m | J. Phys. Chem. B 2012, 116, 12522−12534

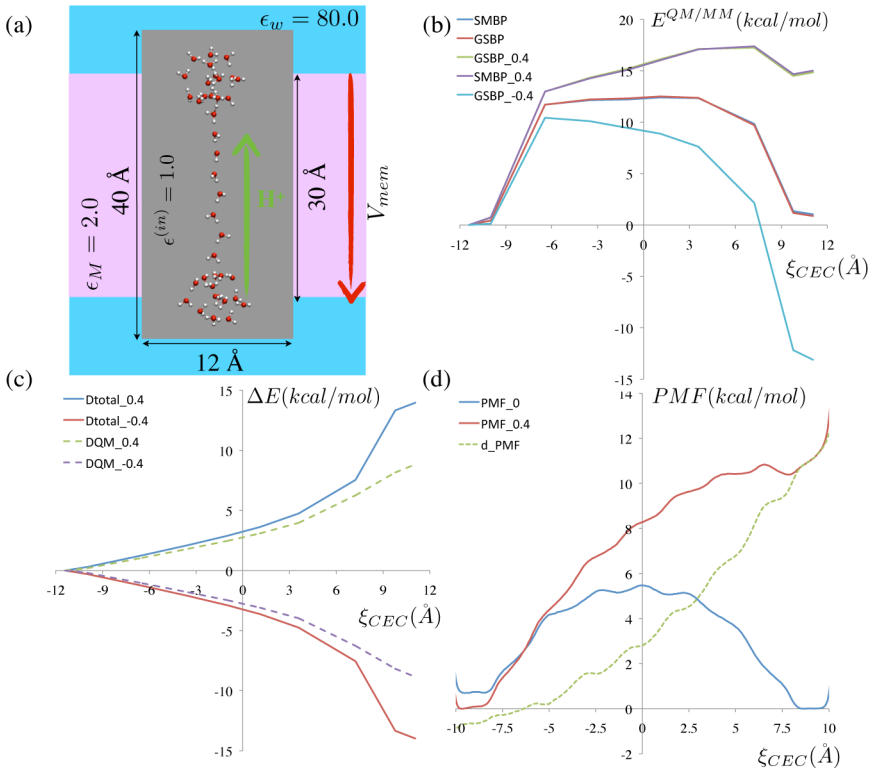



**Figure 7.** Proton transfer reaction through a model channel as described in ref 26 (parameters specified in panel a). (b) The potential energy profile along $\zeta_{CEC}$ with different magnitudes of membrane potential (0, ± 400 mV) with either GSBP or SMBP; the agreement between GSBP and SMBP is excellent, and the corresponding curves for the same membrane potential almost overlap. (c) The differences of the potential profiles relative to the SMBP results without membrane potential; the contributions from the QM region alone are plotted as dashed lines. (d) The potential of mean force (PMF) along $\zeta_{CEC}$ with 0 and 400 mV of membrane potential; the difference between them is shown as the dashed line. Note that the PMF appears less smooth than that in ref 26 because we used the standard SCC-DFTB approach here, which tends to over stabilize the Zundel form of the protonated water clusters.[57]

theless, that QM/MM calculations are required for reliable spectra assignment for enzyme active site.

**4.3. Effect of Membrane Potential on a Proton Transfer Reaction in a Model Membrane.** For the study of membrane systems, a rectangular box shaped dielectric boundary is more useful than a spherical boundary. To demonstrate the performance of the SMBP for such types of applications, we apply it to analyze how membrane potential impacts the energy barrier of a proton transfer reaction within a model membrane.

The system was established and the setup described in ref 26 where a water chain consisting of 10 water molecules with one additional proton crosses the membrane, and two small droplets of explicit water, containing 15 waters each, are situated on both end points of the chain. The membrane itself is represented by a dielectric continuum, as is the outer part of the solvent (see Figure 7a); the thickness of the membrane is taken to be 30 Å. Calculations are performed using GSBP and SMBP with SCC-DFTB as the QM method, where the QM part consists of the linear water chain (10 waters) and the additional proton; MM waters (the two droplets) are described with the TIP3P model. The proton location is controlled by the $\zeta_{CEC}$-coordinate as described in ref 26, which constrains the position of the center of excess charge along the $z$-axis (direction of membrane normal). In addition to potential of mean force (PMF), we also study the approximate MEP here by adiabatic mapping along $\zeta_{CEC}$ to ensure a consistent comparison between GSBP and SMBP. For the PB calculations, the same settings as described in ref 26 are used, except that the trilinear interpolation between the gridpoints is applied instead of the cardinal B-spline method. 194 surface charges are used for calculations on this system. To include the effect of the membrane potential, the PB-V equation[22] rather than the PB equation is solved when computing $\phi^o$; a membrane potential of either +400 mV or −400 mV is applied. For the PMF simulations, 10 windows are used with each window sampled for about 300 ps; two independent sets of simulations are carried out for each value of the membrane potential to ensure that sufficient statistics have been collected.

The approximate MEP curves along the $\zeta_{CEC}$ coordinate with GSBP and SMBP agree very well and are essentially indistinguishable in Figure 7b. In the absence of any membrane potential, the proton transfer is almost thermoneutral as expected based on the symmetry of the system; the deviation from thermoneutrality is due to the fact that local minimization is carried out. In the presence of 400 mV membrane potential, the proton transfer toward the positive $z$-direction is energetically destabilized; the reverse is observed in the presence of −400 mV membrane potential. Interestingly, the effect of the membrane potential on the proton transfer energetics appears to be larger than expected based on a simple estimate: since one proton (a unit charge) has moved across the membrane, an approximate estimate of the impact of the membrane potential on the proton transfer is ±0.4 V × 1 charge = 0.4 eV ∼ ±9.2 kcal/mol. The actually observed effects are plotted as functions of $\zeta_{CEC}$; by the end of the proton transfer, the calculated effect of the membrane potential is almost 14 kcal/mol (Figure 7c), which is ∼5 kcal/mol higher than the simple estimate.


dx.doi.org/10.1021/jp308218m | J. Phys. Chem. B 2012, 116, 12522−12534

Further analysis of the energy contributions indicate that the QM region indeed responds to the membrane potential according to the simple charge model; as shown in Figure 7c, the effect of the membrane potential on the QM energies reach about 9 kcal/mol by the end of the proton transfer. The additional contributions are from the MM water molecules, which reorient substantially during the proton transfer. For example, the *z*-component of the MM water dipole moment changes by more than 12 D (from −6 to 6 D) by the end of the proton transfer. Therefore, the simple model here illustrates that the effect of membrane potential can differ notably from the estimate based simply on the actual charge movement; the current results would suggest an effective charge movement of 14/9.2 = 1.5*e*, although only one proton gets explicitly transferred.

With thermal fluctuations included, the proton transfer PMF (Figure 7d) differs somewhat from the MEP results. First, in the absence of the membrane potential, the barrier is substantially lower than that for the MEP. In the presence of +400 mV membrane potential, the PMF is less favorable (relative to the value without membrane potential) by about 10−11 kcal/mol, which is much closer to the expected effect of 9.2 kcal/mol by considering the net movement of a single charge (proton). Apparently, the effect of MM water reorientation is substantially quenched when thermal fluctuations of these molecules are considered, although there is still a non-negligible contribution; the effective charge movement based on the voltage dependence of the PMF is ∼1.09−1.20*e*. The comparison of PMF and MEP highlights that including a proper degree of sampling is essential to the analysis of long-range proton transfers in a polar environment.[50−53]

## 5. CONCLUSION

We have implemented the SMBP[15] into the program package CHARMM and interfaced it with the semiempirical SCC-DFTB method for benchmarking purposes, as well as with the ab initio quantum mechanical program packages Q-Chem and Gaussian, respectively. We have further demonstrated the robustness of the surface charge approach to represent the electrostatic interactions with the QM part due to the boundary potential, and shown that the SMBP is also fairly robust with respect to the choice of the atomic QM (ESP) charges, where using two different ESP algorithms does not exert a significant influence on the results. The SMBP is currently the only *general* boundary potential approach that can be coupled with ab initio QM/MM methods, and we have demonstrated its applicability for large enzymatic systems such as AP, where we have calculated the influences due to the SMBP on geometries, reaction energetics, and vibrational frequencies; among those properties, the energetics tend to be most sensitive to the boundary condition while vibrational frequencies are least sensitive. Finally, a short application to proton transfer under the influence of membrane potential is provided to illustrate the unique feature of our implementation.

As the issue of timings has been covered in refs 15 and 36 already, no explicit timing information is given here. However, due to the need of converging the SCRF part, actual runtimes of the SMBP compared to a standard calculation without boundary potential are expected to be two to four times slower when the QM calculation is the time-determining step and will otherwise depend strongly on the grid settings of the Poisson−Boltzmann part as well as on the size of the inner region. With respect to the GSBP, an SMBP calculation will be faster if the total number of energy/gradient evaluations is not larger than a few hundred (depending on the number of basis functions in the GSBP calculation). However, we have observed that, for large enzymes such as AP, geometry optimizations usually take more than a few hundred steps (sometimes more than 1000 steps), so that the GSBP is observed to be already generally faster where it is applicable. For MEP calculations this effect will become even more pronounced, and it will become overwhelming for MD simulations, although the latter are still comparatively rarely applied using ab initio QM/MM setups, due to the high cost and the large number of steps needed to obtain sufficient sampling. Therefore, although the SMBP provides a robust, simple, and general boundary potential approach, it would be desirable to develop a GSBP-type implementation for ab initio QM/MM methods as well, even though such a scheme will be more involved to implement (as, for avoiding QM charge derivatives, integrals over the QM charge density and the GSBP basis functions would have to be programmed). This is under development. At the current stage, the best strategy is to combine GSBP and SMBP: GSBP MD simulations can be performed using either an all-MM setup or a semiempirical QM part, and snapshots can be subsequently optimized with SMBP using a QM/MM setup with higher-level ab initio QM methods.

## ■ ASSOCIATED CONTENT

### Supporting Information

Additional details regarding the determination of surface charges, algorithm modifications for improving the stability of MDC charge determination, grid setting in AP calculations, and a structural and vibrational frequency comparison for AP calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

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

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This research is supported in part by the National Science Foundation Grant CHE-0957285. Computational resources from the National Center for Supercomputing Applications at the University of Illinois and the Center of High Throughput Computing at UW-Madison are greatly appreciated. Computations are also supported in part by the National Science Foundation through a major instrumentation grant (CHE-0840494). We thank Prof. D. Herschlag for sending us the preprint of ref 41 before its publication. J.Z. is indebted to the "Nationale Akademie der Wissenschaften Leopoldina" for a postdoctoral fellowship (LPDS 2009-39).

## ■ REFERENCES

(1) Pegram, L. M.; Record, M. T., Jr. *J. Phys. Chem. B* **2008**, *112*, 9428−9436.
(2) Jungwirth, P.; Winter, B. *Annu. Rev. Phys. Chem.* **2008**, *59*, 343−366.
(3) Record, M. T.; Zhang, W. T.; Anderson, C. F. *Adv. Protein Chem.* **1998**, *51*, 281−353.
(4) Ma, L.; Pegram, L.; Record, M. T., Jr.; Cui, Q. *Biochemistry* **2010**, *49*, 1954−1962.

(5) Shao, Q.; Fan, Y. B.; Yang, L. J.; Gao, Y. Q. *J. Chem. Phys.* **2012**, *136*, 115101.
(6) Harada, R.; Sugita, Y.; Feig, M. *J. Am. Chem. Soc.* **2012**, *134*, 4842−4849.
(7) Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, *17*, 283.
(8) Warshel, A.; King, G. *Chem. Phys. Lett.* **1985**, *121*, 124−129.
(9) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050−9063.
(10) Im, W.; Bernèche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924.
(11) Yang, Y.; Yu, H.; Cui, Q. *J. Mol. Biol.* **2008**, *381*, 1407−1420.
(12) Gao, Y.; Karplus, M. *Curr. Opin. Struct. Biol.* **2004**, *14*, 250−259.
(13) McQuarrie, D. A. *Statistical Mechanics*: Harper and Row: New York, 1976.
(14) Yamaguchi, Y.; Goddard, J. D.; Osamura, Y.; Schaefer, H. F. *A New Dimension to Quan-tum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory*; Oxford University Press: New York, 1994.
(15) Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2009**, *5*, 3114.
(16) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.
(17) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(18) Q-Chem Program Package. http://www.q-chem.com (2007).
(19) Frisch, M. J.; et al. *Gaussian 09 Revision A.1*; Gaussian, Inc.: Wallingford, CT, 2009.
(20) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111* (1−3), 59−75.
(21) Woo, H. J.; Dinner, A. R.; Roux, B. *J. Chem. Phys.* **2004**, *121*, 6392−6400.
(22) Roux, B. *Biophys. J.* **1997**, *73*, 2980−2989.
(23) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(24) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*; Cambridge University Press: Cambridge, 1992.
(25) Hou, G.; Cui, Q. *J. Am. Chem. Soc.* **2012**, *134*, 229.
(26) König, P. H.; Ghosh, N.; Hoffmann, M.; Elstner, M.; Tajkhorshid, E.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. A* **2006**, *110*, 548.
(27) Kaila, V. R.; Verkhovsky, M. V.; Wikström, M. *Chem. Rev.* **2010**, *110*, 7062−7081.
(28) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wióorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(29) Yang, Y.; Yu, H.; York, D.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2008**, *4*, 2067.
(30) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
(31) Hariharan, P. C.; Pople, J. A. *Theoret. Chim. Acta* **1973**, *28*, 213.
(32) Gill, P. M. W.; Johnson, B. G.; Pople, J. A. *Chem. Phys. Lett.* **1993**, *209*, 506.
(33) König, P.; Hoffmann, M.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 9082.
(34) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435.
(35) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins: Struct., Funct., Genet.* **1986**, *1*, 47.
(36) Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2011**, *7*, 238.
(37) Sigfridsson, E.; Ryde, U. *J. Comput. Chem.* **1998**, *19*, 377.
(38) Simmonett, A. C.; Gilbert, A. T. B.; Gill, P. M. W. *Mol. Phys.* **2005**, *103*, 2789.
(39) Galperin, M. Y.; Hedrzejas, M. J. *Proteins: Struct., Funct., Bioinf.* **2001**, *45*, 318.
(40) O'Brien, P. J.; Herschlag, D. *Biochemistry* **2001**, *40*, 5691.
(41) Andrews, L. D.; Deng, H.; Herschlag, D. *J. Am. Chem. Soc.* **2011**, *133*, 11621.
(42) Coleman, J. E. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 441.
(43) Lopez-Canut, V.; Marti, S.; Bertran, J.; Moliner, V.; Tunon, I. *J. Phys. Chem. B* **2009**, *113*, 7816.
(44) Sheppard, D.; Terrell, R.; Henkelman, G. *J. Chem. Phys.* **2008**, *128*, 134106.
(45) O'Brien, P. J.; Lassila, J. K.; Fenn, T. D.; Zalatan, J. G.; Herschlag, D. *Biochemistry* **2008**, *47*, 7663.
(46) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.
(47) Cui, Q.; Karplus, M. *J. Chem. Phys.* **2000**, *112*, 1133−1149.
(48) Ghysels, A.; Woodcock, H. L., III; Larkin, J. D.; Miller, B. T.; Shao, Y.; Kong, J.; Van Neck, D.; Van Speybroeck, V.; Waroquier, M.; Brooks, B. R. *J. Chem. Theory Comput.* **2011**, *7*, 496.
(49) Bobyr, E.; Lassila, J. K.; Wiersma-Koch, H. I.; Fenn, T. D.; Lee, J. J.; Nikolic-Hughes, I.; Hodgson, K. O.; Rees, D. C.; Hedman, B.; Herschlag, D. *J. Mol. Biol.* **2012**, *415*, 102.
(50) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; Konig, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458−6469.
(51) Riccardi, D.; Koenig, P.; Guo, H.; Cui, Q. *Biochemistry* **2008**, *47*, 2369−2378.
(52) Kato, M.; Pisliakov, A. V.; Warshel, A. *Proteins: Struct., Funct., Bioinf.* **2006**, *64*, 829−844.
(53) Voth, G. A. *Acc. Chem. Res.* **2006**, *39*, 143−150.
(54) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502.
(55) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.
(56) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
(57) Goyal, P.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2011**, *115*, 6790−6805.