JCTC Journal of Chemical Theory and Computation

Efficiency and Accuracy of the Generalized Solvent Boundary Potential for Hybrid QM/MM Simulations: Implementation for Semiempirical Hamiltonians

Tobias Benighaus and Walter Thiel*

Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470, Mülheim an der Ruhr, Germany

Received May 27, 2008

Abstract: We report the implementation of the generalized solvent boundary potential (GSBP) [Im, W.; Bernèche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924] in the framework of semiempirical hybrid quantum mechanical/molecular mechanical (QM/MM) methods. Application of the GSBP is connected with a significant overhead that is dominated by numerical solutions of the Poisson−Boltzmann equation for continuous charge distributions. Three approaches are presented that accelerate computation of the values at the boundary of the simulation box and in the interior of the macromolecule and solvent. It is shown that these methods reduce the computational overhead of the GSBP significantly with only minimal loss of accuracy. The accuracy of the GSBP to represent long-range electrostatic interactions is assessed for an extensive set of its inherent parameters, and a set of optimal parameters is defined. On this basis, the overhead and the savings of the GSBP are quantified for model systems of different sizes in the range of 7000 to 40 000 atoms. We find that the savings compensate for the overhead in systems larger than 12 500 atoms. Beyond this system size, the GSBP reduces the computational cost significantly, by 70% and more for large systems (>25 000 atoms).

1. Introduction

In the past decade, hybrid quantum mechanical/molecular mechanical (QM/MM) methods have gained popularity for the simulation of biomolecules and are now frequently used for the calculation of free energy differences.[1−11] In the context of this development, the treatment of long-range electrostatic interactions in QM/MM simulations attracted significant attention. The accurate description of long-range electrostatic interactions was found to be imperative for meaningful simulations of biomolecular systems, since electrostatic interactions strongly influence their structure and function.[12−15] While the development of efficient and accurate methods to treat these interactions has been an active area of research in the field of classical simulations for a long time, these techniques are only recently adapted to QM/MM methods due to the technical difficulties introduced by the QM atoms.

For the case of periodic boundary conditions (PBC), Ewald summation is an established method to compute the electrostatic energy and forces of an infinite periodic array of systems without significant truncations.[16−18] Therefore, the applicability of the Ewald summation method has been extended to hybrid QM/MM simulations with semiempirical QM Hamiltonians.[19−21] Unfortunately, application of these methods to large nonperiodic biomolecules is affected by serious problems. The imposed periodicity may lead to significant artifacts[22−24] and even qualitatively wrong results unless the molecule is solvated in a solvent box of adequate size.[25] Thus, the number of solvent molecules is necessarily large and increases the computational costs massively, such that the Ewald summation method can only be used for small- to medium-sized biomolecules.

Very often, however, one is interested in simulating the behavior of a large biomolecule in infinite dilution, and alternative approaches were devised to facilitate these computations. For biochemical reactions that proceed in a localized region of the macromolecule, boundary potentials

* Corresponding author. E-mail: thiel@mpi-muelheim.mpg.de.

10.1021/ct800193a CCC: $40.75
Published on Web 09/19/2008

are an especially attractive approach.[26−37] Within this approach, the system is subdivided into an inner region, containing the active site and the adjacent part of the enzyme, and an outer region, containing the rest of the enzyme and the outer solvent molecules. While the inner region is simulated atomistically, the effect of the outer region onto the inner region is mimicked by the boundary potential. Ideally, the boundary potential is designed such that the statistical properties of the inner region interacting with the boundary potential are the same as those of the full solvated macromolecule. Although this may be formulated rigorously as an integration over the outer region degrees of freedom,[37] an efficient implementation necessitates the introduction of further approximations.

In the generalized solvent boundary potential (GSBP), developed by Im et al. in 2001, the outer region solvent molecules are described by a continuous polarizable dielectric and the outer region charge distribution by fixed point charges.[38] Electrostatic interactions with the outer region macromolecule and solvent molecules are separated into a solvent-shielded static field created by the outer region point charges interacting with the dielectric, and a dynamic reaction field induced by interaction of the inner region charge distribution with the dielectric. A great advantage of the GSBP is that the dynamic reaction field term can handle irregularly shaped macromolecule/solvent boundaries. Accuracy and efficiency of the GSBP in classical simulations were validated by studies on aspartyl-tRNA synthetase[38,39] and the KcsA potassium channel.[38] In 2005, Cui and co-workers adapted the GSBP method to the QM/MM framework as a means to treat long-range electrostatic interactions in QM/MM simulations accurately and to describe QM/MM and MM/MM interactions in a balanced way.[40] Here, the self-consistent-charge density-functional tight-binding (SCC-DFTB)[41] method was chosen as the QM Hamiltonian. The accuracy of the SCC-DFTB/MM/GSBP approach was evaluated by comparison to results from Ewald/PBC calculations on small model systems. The SCC-DFTB/MM/GSBP method was found to provide quantitatively very similar results at significantly lower computational costs compared to Ewald/PBC methods.[42−44] The fixation of the outer region atoms is a fundamental assumption in the GSBP that allows for a closed-form expression for the electrostatics.[38] While this assumption is valid in many cases, the use of the GSBP was found to be problematic if the macromolecule underwent major conformational changes during the course of a reaction.[44] For the investigation of localized processes in large macromolecules, the SCC-DFTB/MM/GSBP approach proved to be an efficient and accurate method and was applied subsequently to study several biological systems.[45−48]

The use of the GSBP is connected with a significant overhead. Initially, the solvent-shielded static field and the matrix representation of the reaction field Green's function have to be calculated. Computation of the reaction field matrix implies solving several hundred linearized Poisson−Boltzmann (PB) equations and is therefore rather demanding. Furthermore, the accuracy of the GSBP and the costs of its overhead strongly depend on the choice of parameters that are inherent to the GSBP and the finite-difference solution of the PB equation. To the best of our knowledge, a systematic determination of the best parameters for the GSBP has not been pursued up to date. In this study, we determine a set of parameters that provide the accuracy that is necessary to mimic the effect of the outer region at optimal computational costs. On the basis of these parameters, we quantify the overhead and the savings related to the GSBP, and estimate the minimum system size for which the GSBP is more efficient than standard approaches using nontruncated Coulombic electrostatics. Moreover, we present improved algorithms that decrease the costs for computation of the reaction field matrix significantly.

Previously, the GSBP method was adapted to the hybrid QM/MM framework exclusively in combination with the SCC-DFTB Hamiltonian for the QM region.[40] In light of the success of reaction-specific parametrizations of semiempirical methods based on the neglect of diatomic differential overlap (NDDO) approximation in QM/MM simulations[49−53] and the widespread use of NDDO-based QM/MM methods in general,[11] we found it desirable to adapt and implement the GSBP as an efficient means to treat long-range electrostatics in NDDO-based QM/MM simulations. This interest is further substantiated by recent findings that NDDO-based methods are more reliable for certain properties and systems,[54] and that the use of SCC-DFTB may be problematic for specific systems.[55] Accordingly, we adapted the GSBP for NDDO-based QM/MM approaches and present the implementation in this work.

2. Theory

In this section, we briefly review the theoretical background of the GSBP for classical MM simulations[38] and its QM/MM implementation.[40] Thereafter, the adaptation to NDDO-based QM/MM methods and strategies to accelerate computation of the reaction field matrix are presented.

2.1. GSBP for MM Methods. Consider a macromolecule R surrounded by N solvent molecules. In a boundary potential approach, the system is subdivided into an inner region that contains the inner part of the macromolecule and the n inner solvent molecules, and an outer region that contains the outer part of the macromolecule and the $N - n$ outer solvent molecules. Statistical expectation values depending only on the degrees of freedom of the inner region ($\mathbf{R}_i$, 1,..., n) can be calculated by integrating out the outer region contributions. The influence of the outer region on the inner region can be described rigorously by means of the potential of mean force (PMF) $W(\mathbf{R}_i, 1,..., n)$.

$$e^{-\beta W(\mathbf{R}_i,1,...\mathbf{n})} = \frac{1}{C}\int' d\mathbf{R}_o\, d(n+1) \ldots dN e^{-\beta U(\mathbf{R},1,...,N)} \quad (1)$$

Here, C denotes an arbitrary integration constant, and the primed integral indicates integration over the degrees of freedom of the outer region ($\mathbf{R}_o$, $n + 1$,..., N) including only those configurations with all outer region atoms outside the inner region. Beglov and Roux demonstrated that the PMF is related to the reversible thermodynamic work necessary to assemble the inner region.[37]

$$W(\mathbf{R}_i, 1, ..., n) = U(\mathbf{R}_i, 1, ..., n) + \Delta W_{cr} + \Delta W_{np}(\mathbf{R}_i, 1, ..., n) + \Delta W_{elec}(\mathbf{R}_i, 1, ..., n) \quad (2)$$

The contribution to the PMF that arise from the configurational restrictions and the nonpolar and the electrostatic

interactions are denoted ΔW_{cr}, ΔW_{np}, and ΔW_{elec}, respectively. U is the potential energy of the isolated inner region that includes bonded and nonbonded (van der Waals and electrostatic) terms.

The goal of the GSBP is to provide an efficient and accurate approximation to the electrostatic contribution to the PMF. Therefore, the outer region solvent molecules are described by a polarizable dielectric continuum (PDC) and the outer region macromolecule by fixed point charges. The electrostatic contributions to the PMF now consist of the direct Coulombic interactions of inner and outer region (U^{io}_{elec}), and the solvation free energy resulting from interaction with the PDC (ΔW^{solv}_{elec}). Representing the charge distribution of the outer macromolecule and the inner region by point charges q_A, the electrostatic solvation free energy can be calculated as

$$\Delta W^{solv}_{elec} = \frac{1}{2}\sum_A q_A \phi_{rf}(\mathbf{r}_A) \tag{3}$$

where the reaction field potential $\phi_{rf}(\mathbf{r})$ is the difference of a reference electrostatic potential computed in vacuum, $\phi_v(\mathbf{r})$, and the electrostatic potential computed in solution, $\phi_s(r)$. The electrostatic potentials are obtained by solving the linearized Poisson−Boltzmann (PB) equation

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

with the charge density of all explicit atoms $\rho(\mathbf{r})$, the space-dependent dielectric constant $\epsilon(\mathbf{r})$, and modified Debye−Hückel screening factor $\bar{\kappa}(\mathbf{r})$.[56] The solvation free energy term is problematic, since during sampling of the inner region configurations the PB equation would have to be solved for each configuration which is prohibitively expensive. To isolate the dynamic properties, the charge distribution is separated into an inner and outer part.

$$\rho(\mathbf{r}) = \rho_i(\mathbf{r}) + \rho_o(\mathbf{r}) \tag{5}$$

In consequence, the electrostatic solvation free energy splits up into three terms: outer−outer, inner−outer, and inner−inner contributions.

$$\Delta W^{solv}_{elec} = \Delta W^{oo}_{elec} + \Delta W^{io}_{elec} + \Delta W^{ii}_{elec} \tag{6}$$

The first term, ΔW^{oo}_{elec}, stems from the interaction of the outer region charge distribution with the self-induced reaction field and is constant throughout sampling. The inner−outer contribution arises from the interaction of the inner region charge distribution with the reaction field that is induced by the outer region charge distribution. Calculation of the Coulombic interaction of the inner and outer region can be combined very efficiently with the calculation of the inner−outer contribution to the solvation free energy.

$$\Delta W^{io}_{elec} + U^{io}_{elec} = \sum_{A\in inner} q_A \phi^o_{rf}(\mathbf{r}_A) + U^{io}_{elec}$$

$$= \sum_{A\in inner} q_A \phi^o_s(\mathbf{r}_A) \tag{7}$$

The outer region being fixed, the electrostatic potential of the outer region charges in solution, $\phi^o_s(\mathbf{r})$, has to be computed only once and is valid for all inner region configurations. As the interaction with all outer region point charges and solvent molecules is substituted by an interaction with a static potential, computational costs are reduced massively. However, computation of the inner−inner contribution

$$\Delta W^{ii}_{elec} = \frac{1}{2}\sum_{A\in inner} q_A \phi^i_{rf}(\mathbf{r}_A) \tag{8}$$

remains problematic because $\phi^i_{rf}(\mathbf{r})$ depends on the inner region configuration. To circumvent repeated solution of the PB equation, the inner region charge distribution is projected onto a set of basis functions $\{b_n\}$:

$$\rho_i(\mathbf{r}) = \sum_n c_n b_n(\mathbf{r}) \tag{9}$$

For a set of orthonormal basis functions, the expansion coefficients c_n are the generalized multipole moments Q_n of the charge distribution:

$$Q_n = \sum_{A\in inner} q_A b_n(\mathbf{r}_A) \tag{10}$$

Finally, the reaction field Green's function which determines the inner reaction field potential

$$\phi^i_{rf}(\mathbf{r}) = \int d\mathbf{r}'\ \rho_i(\mathbf{r}')G_{rf}(\mathbf{r},\mathbf{r}') \tag{11}$$

is projected onto the same basis set. Using M_{mn}, the matrix representation of G_{rf}, the inner−inner electrostatic contribution to the PMF can be expressed as

$$\Delta W^{ii}_{elec} = \frac{1}{2}\sum_{mn} Q_m M_{mn} Q_n \tag{12}$$

This leads to the final expression for the electrostatic contribution to the PMF

$$\Delta W_{elec} = \sum_{A\in inner} q_A \phi^o_s(\mathbf{r}_A) + \frac{1}{2}\sum_{mn} Q_m M_{mn} Q_n \tag{13}$$

In MD simulations employing the GSBP, the inner region atoms move on the PMF surface that is defined as

$$W(\mathbf{R}_i, 1, ..., n) = U(\mathbf{R}_i, 1, ..., n) + \Delta W_{cr} + \Delta W_{np} + \sum_{A\in inner} q_A \phi^o_s(\mathbf{r}_A) + \frac{1}{2}\sum_{mn} Q_m M_{mn} Q_n \tag{14}$$

2.2. GSBP Implementation for NDDO-Based QM/MM Methods. Extension of the GSBP to general QM/MM methods necessitates further subdivision of the inner region into QM and MM regions, since it is natural to assume that the QM region lies within the inner region. Consequently, the inner region charge distribution splits up into QM and MM charge distributions that interact separately with the static outer region field, ϕ^o_s, and the reaction field Green's function, G_{rf}. Equation 13 has to be modified as follows to account for these changes:

$$\Delta W_{elec} = \sum_{A\in MM} q_A \phi^o_s(\mathbf{r}_A) + \int d\mathbf{r}\rho^{QM}(\mathbf{r})\phi^o_s(\mathbf{r}) + \frac{1}{2}\sum_{mn} Q^{QM}_m M_{mn} Q^{QM}_n + \sum_{mn} Q^{QM}_m M_{mn} Q^{MM,cs}_n + \frac{1}{2}\sum_{mn} Q^{MM}_m M_{mn} Q^{MM}_n \tag{15}$$

The main issue that arises when introducing QM atoms into the GSBP framework is the representation of the QM charge distribution in the terms that describe the interaction with the outer region field and the reaction field. As NDDO-based semiempirical QM methods use only a minimum set of relatively tight basis functions, we decided to represent the QM charge distribution by a set of Mulliken charges.[57] Now, the QM-dependent terms in eq 15 can be calculated in close analogy to the MM terms

$$\int \mathrm{d}\mathbf{r}\, \rho^{\mathrm{QM}}(\mathbf{r})\phi_{\mathrm{s}}^{\mathrm{o}}(\mathbf{r}) = \sum_{A\in \mathrm{QM}} q_{\mathrm{A}}^{\mathrm{Mull}}\phi_{\mathrm{s}}^{\mathrm{o}}(\mathbf{r}_A) \tag{16}$$

and

$$Q_n^{\mathrm{QM}} = \int \mathrm{d}\mathbf{r}\, \rho^{\mathrm{QM}}(\mathbf{r})b_n(\mathbf{r}) = \sum_{A\in \mathrm{QM}} q_A^{\mathrm{Mull}} b_n(\mathbf{r}_A) \tag{17}$$

Here, q_A^{Mull} are the Mulliken charges representing the QM charge distribution, and Q_n^{QM} are the multipole moments of the QM charge distribution. Still, we are facing two technical difficulties. First, electrostatic interactions at the QM−MM boundary need to be treated with special care to avoid overpolarization of the QM electron density. Thus, the QM electron density does not interact with the full MM charge distribution but with a modified one. In this work, we implemented the GSBP for use in combination with the charge-shift scheme,[58] and therefore, the QM charge distribution interacts with the reaction field potential that is induced by the charge-shifted MM charges ($\mathrm{MM^{cs}}$) (fourth term in eq 15), with

$$Q_n^{\mathrm{MM,cs}} = \sum_{\mathrm{A}\in \mathrm{MM^{cs}}} q_{\mathrm{A}} b_n(\mathbf{r}_{\mathrm{A}}) \tag{18}$$

Second, using electronic embedding,[59] the QM wave function interacts with all MM point charges and the PDC. Hence, the GSBP contributions have to be accommodated at the level of the self-consistent field (SCF) iterations during optimization of the wave function by addition of the following terms to the Fock matrix.

$$F_{\mu\nu}^{\mathrm{GSBP}} = -\frac{1}{2}\delta_{\mu\nu}[\Omega_{\mathrm{C}} + \Omega_{\mathrm{D}}] - \frac{1}{2}\delta_{\mu\nu}\sum_{A\in \mathrm{QM}} q_A^{\mathrm{Mull}}[\Gamma_{\mathrm{CA}} + \Gamma_{\mathrm{DA}}]; \quad \mu\in \mathrm{C},\ \nu\in \mathrm{D} \tag{19}$$

Here, μ and ν denote basis functions attached to the QM atoms C and D, respectively. The atom-dependent matrices Ω_{C} and Γ_{CA} are defined as

$$\Omega_{\mathrm{C}} = \phi_{\mathrm{s}}^{\mathrm{o}}(\mathbf{r}_{\mathrm{C}}) + \sum_{mn} b_m(\mathbf{r}_{\mathrm{C}})M_{mn}Q_n^{\mathrm{MM,cs}} \tag{20}$$

and

$$\Gamma_{\mathrm{CA}} = \sum_{mn} b_m(\mathbf{r}_{\mathrm{C}})M_{mn}b_n(\mathbf{r}_{\mathrm{A}}) \tag{21}$$

Moreover, the GSBP also affects the atomic forces, and its contribution to the analytic gradient can be evaluated by taking the first derivative of the GSBP contribution to the PMF with respect to the atomic coordinates. In the case of a QM atom, the analytic derivative takes the following form

$$\frac{\partial}{\partial \mathbf{r}_{\mathrm{A}}}\Delta W_{\mathrm{elec}} = q_{\mathrm{A}}^{\mathrm{Mull}}\frac{\partial}{\partial \mathbf{r}_{\mathrm{A}}}\phi_{\mathrm{s}}^{\mathrm{o}}(\mathbf{r}_{\mathrm{A}}) + \sum_{\mathrm{B}\in \mathrm{QM}} \frac{\partial q_{\mathrm{B}}^{\mathrm{Mull}}}{\partial \mathbf{r}_{\mathrm{A}}}\phi_{\mathrm{s}}^{\mathrm{o}}(\mathbf{r}_{\mathrm{B}}) + \sum_{mn}\left[\frac{\partial}{\partial \mathbf{r}_{\mathrm{A}}}Q_m^{\mathrm{QM}}\right]M_{mn}[Q_n^{\mathrm{QM}} + Q_n^{\mathrm{MM,cs}}] \tag{22}$$

where the derivatives of the QM multipole moments are calculated as

$$\frac{\partial}{\partial \mathbf{r}_{\mathrm{A}}}Q_m^{\mathrm{QM}} = q_{\mathrm{A}}^{\mathrm{Mull}}\frac{\partial}{\partial \mathbf{r}_{\mathrm{A}}}b_m(\mathbf{r}_{\mathrm{A}}) + \sum_{\mathrm{B}\in \mathrm{QM}} \frac{\partial q_{\mathrm{B}}^{\mathrm{Mull}}}{\partial \mathbf{r}_{\mathrm{A}}}b_m(\mathbf{r}_{\mathrm{B}}) \tag{23}$$

In contrast to a previous implementation of the GSBP for hybrid QM/MM approaches,[40] we found it necessary for NDDO-based QM methods to include the contribution from coupled Mulliken charge derivatives, $\partial q_{\mathrm{A}}^{\mathrm{Mull}}/\partial \mathbf{r}_{\mathrm{B}}$, to compute accurate gradients of the QM atoms. Using only one-center Mulliken charge derivatives, $\partial q_{\mathrm{A}}^{\mathrm{Mull}}/\partial \mathbf{r}_{\mathrm{A}}$, a mean absolute deviation (MAD) of the components of the QM gradient in the range of 10^{-3} au was observed (compared with finite-difference reference values). Incorporating the contribution from the coupled Mulliken charge derivatives reduces the MAD to the order of 10^{-5} au which is sufficiently accurate. Although Mulliken charge derivatives take a very simple form in the NDDO approximation

$$\frac{\partial}{\partial \mathbf{r}_{\mathrm{B}}}q_A^{\mathrm{Mull}} = -\sum_{\alpha\in \mathrm{A}}\frac{\partial}{\partial \mathbf{r}_{\mathrm{B}}}P_{\alpha\alpha} \tag{24}$$

their computation is complicated, since the coupled-perturbed SCF (CPSCF) equations have to be solved to calculate the derivatives of the SCF density matrix.[60,61]

In the case of an MM atom, the evaluation of the derivative of the GSBP contribution to the PMF is less demanding:

$$\frac{\partial}{\partial \mathbf{r}_{\mathrm{A}}}\Delta W_{\mathrm{elec}} = q_{\mathrm{A}}\frac{\partial}{\partial \mathbf{r}_A}\phi_{\mathrm{rf}}^{\mathrm{o}}(\mathbf{r}_{\mathrm{A}}) + q_{\mathrm{A}}\sum_{mn}\left[\frac{\partial}{\partial r_{\mathrm{A}}}b_n(r_{\mathrm{A}})\right]M_{mn}[Q_n^{\mathrm{QM}} + Q_n^{\mathrm{MM}}] \tag{25}$$

2.3. Computation of the Reaction Field Matrix. Although the reaction field matrix is formally the matrix representation of the reaction field Green's function, the computation of this matrix follows a different approach that exploits the fact that its *mn*th element corresponds to the interaction of b_m with the reaction field induced by b_n.[38]

$$M_{mn} = \int \mathrm{d}\mathbf{r}\, b_n(\mathbf{r})\phi_{\mathrm{rf}}(\mathbf{r};b_m(\mathbf{r})) \tag{26}$$

To calculate $\phi_{\mathrm{rf}}(\mathbf{r};b_m(\mathbf{r}))$, it is necessary to solve the PB equation with the dielectric boundary defined by the macromolecule, and the charge distribution defined by b_m in the inner region and set to zero in the outer region, for vacuum and solvent conditions. Since a standard GSBP calculation employs about 400 basis functions,[38] computation of the reaction field matrix implies solving the PB equation about 800 times. This procedure is computationally expensive and dominates the GSBP-related overhead. Therefore, three approaches to accelerate computation of the reaction field matrix are presented in this section.

2.3.1. Coarsening of the Inner Region. In finite-difference solutions of the PB equation, the boundary values are commonly set using the Debye−Hückel expression[56]

$$\phi_i = \sum_j \frac{q_j e^{-\kappa r_{ij}}}{\epsilon r_{ij}} \tag{27}$$

that implies summation over all point charges q_j for each boundary point ϕ_i. With a continuous charge distribution in the inner region, determination of the boundary values becomes computationally expensive. In the original GSBP work, a focusing procedure[62] is used to reduce these computational costs. In this procedure, the PB equation is first solved for a rough outer grid (grid I) with large spatial extent. Subsequently, a fine inner grid (grid II) focusing on the inner region with boundary values defined by grid I is used to calculate an accurate electrostatic potential. However, even when such a focusing procedure is used, determination of the boundary values of grid I still has a significant share of the computational costs. Since the boundary points of grid I are far from the inner region and the "charge" in the outer region is zero, a less accurate representation of the basis function in the inner region is expected to be sufficient. Therefore, we introduce the "coarsening of the inner region" (CIR) approximation that utilizes a very rough grid (grid III) to represent the "charge" distribution that is only used to determine the boundary values of grid I. The mesh size of grid III is the product of the new CIR factor and the mesh size of grid I; i.e., a CIR factor of 1.0 corresponds to a standard focusing procedure.

2.3.2. Linear Interpolation. In view of the large distance between the boundary points of grid I and the inner region, it is evident that the boundary values are slowly varying. Therefore, we introduce a simple interpolation scheme that reduces the number of explicitly determined boundary values significantly. On the edges every second and on the faces every fourth boundary value is calculated using the Debye−Hückel expression. The remaining boundary values are determined by linear interpolation from the adjacent four or two boundary points. This scheme is illustrated in Figure 1. For an example grid with 100^3 points, the linear interpolation scheme reduces the number of explicitly determined boundary values from 58 416 to 14 802.

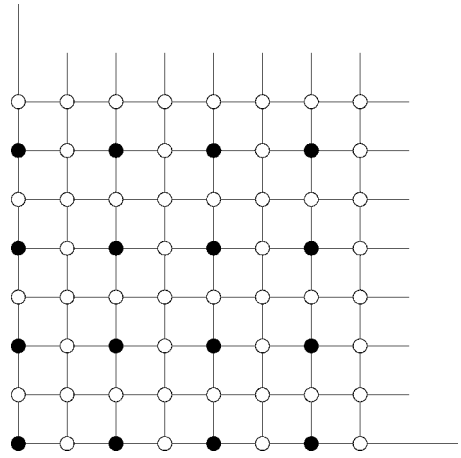

Figure 1. Interpolation scheme used to define boundary values in finite-difference solutions of the PB equation. Black circles represent boundary points that are set using the Debye−Hückel expression. White circles represent boundary points that are set by interpolation.

2.3.3. Modified Stripping. In a finite-difference solution to the PB equation with zero salt conditions, the potential at a particular grid point, ϕ_0, is updated using the potential at the six nearest neighbors, ϕ_i, the dielectric constants ϵ_i at the midpoints between ϕ_0 and ϕ_i, and the charge q_0 assigned to that grid point.

$$\phi_0 = \frac{\sum_{i=1}^{6} \epsilon_i \phi_i + 4\pi q_0/h}{\sum_{i=1}^{6} \epsilon_i} \tag{28}$$

Here, h is the distance between two grid points. This procedure implies 13 additions, 7 multiplications, and 1 division per grid point. Honig et al. demonstrated[63] that the number of mathematical operations can be reduced significantly for most grid points. For a point with zero charge that is surrounded by a uniform dielectric constant, eq 28 simplifies to

$$\phi_0 = \frac{1}{6}\sum_{i=1}^{6} \phi_i \tag{29}$$

Updating these points requires only 6 additions and 1 multiplication. This procedure is termed "stripping" because the points are updated separately.[63] As a continuous charge distribution is used in the computation of the reaction field matrix, there are no points without charge in the inner region. Therefore, we apply a "modified stripping" approach and drop the zero charge condition: for all points surrounded by a uniform dielectric constant with arbitrary charge, we simplify eq 28 as follows:

$$\phi_0 = \frac{1}{6}\sum_{i=1}^{6} \phi_i + \frac{2\pi q_0}{3h\epsilon} \tag{30}$$

Although three additional operations per grid point (one addition, one multiplication, and one division) are necessary compared to the standard stripping approach, modified stripping offers computational savings since it is applicable to a significantly larger number of grid points.

3. Computational Details

The GSBP was implemented in a developmental version of the modular program package ChemShell.[58] The energy and gradient evaluations for the QM part were performed with the MNDO2004 program that was modified locally to account for the GSBP contribution. The AM1 method was chosen as the QM Hamiltonian.[64] The SCF convergence criterion was 10^{-8} eV. For the MM part, the DL_POLY[65] code was employed to run the CHARMM22 force field.[66] The PB equation was solved with our new ChemShell PB module that uses the optimal successive over-relaxation method in combination with Gauss−Seidel relaxation to compute the electrostatic potential.[63,67] A convergence criterion of 2×10^{-5} au was employed for the maximum absolute change in every grid point. If not stated otherwise, the dielectric constants of the macromolecule, ϵ_m, and the solvent, ϵ_s, were set to 1 and 80, respectively. van der Waals radii from the CHARMM22 force field were used to define

Table 1. Average Absolute Percentage Deviation [%] of the Electrostatic Interaction between Inner and Outer Region Computed from the PB Electrostatic Potential with Different Mesh Sizes of the Inner and Outer Grid

	inner grid size (Å)			
outer grid size (Å)	0.25	0.40	0.60	0.80
1.00	0.13	0.14	0.22	0.55
1.25	0.16	0.28	0.26	0.61
1.50	0.25	0.28	0.28	0.62
1.75	0.21	0.28	0.28	0.64
2.00	0.23	0.16	0.24	0.61
2.50	0.25	0.27	0.27	0.60

the dielectric boundary. All calculations for which timings are reported were performed serially on 2.6 GHz AMD Opteron machines with 16 GB of memory.

4. Optimization of Parameters

The accuracy and the efficiency of the GSBP strongly depend on the values that are chosen for its inherent parameters. In this section, we determine a set of optimal values for the mesh sizes of the inner and outer grid, the CIR factor, and assess the accuracy of the approximations that were introduced in section 2.3.

4.1. Static Outer Region Field. The reliability of the static solvent-shielded outer region field, ϕ_s^o, to mimic the electrostatic potential in the inner region is judged by comparison to the exact Coulombic potential. In vacuum environment, i.e., with $\epsilon_m = \epsilon_s = 1$, the electrostatic interactions between the inner and outer region have to be identical when using the electrostatic potentials from solution of the PB equation or the Coulomb expression. A model system consisting of a threonine molecule solvated in a TIP3P water ball with 30 Å radius and 4175 water molecules was set up for this study. By means of classical molecular dynamics (MD) simulation, 10 different configurations of this model system were generated. For each configuration, the center of the inner region was taken to be the C_α carbon of threonine. All molecules with any atom within 18 Å from the center were assigned to the inner region. Depending on the configuration, the inner region contained between 2858 and 2978 atoms. As the electrostatic interaction energy varies with the size of the inner region, we averaged over the absolute percentage deviation in the electrostatic interaction energy.

On average, the vacuum electrostatic interaction between the inner and outer region (U_{elec}^{io}) was -3096.2 ± 243.5 kcal/mol. The average absolute percentage deviation was calculated for all combinations of outer grid mesh sizes of 1.0, 1.25, 1.5, 1.75, 2.0, and 2.5 Å, and inner grid mesh sizes of 0.25, 0.4, 0.6, and 0.8 Å. The results given in Table 1 indicate that the interaction energy calculated from the PB electrostatic potential is very accurate. All mesh size combinations provide average deviations $<0.3\%$ if the inner grid spacing is ≤ 0.6 Å, indicating that 0.6 Å is reasonable choice for the inner grid spacing. For the outer grid, no reliable correlation was found between mesh size and accuracy.

To ensure that the accuracy is not euphemized by cancelation of errors, we also assess the reliability of the electrostatic forces in the inner region. For this purpose, the MAD of the electrostatic force components of all atoms inside spherical active regions with radii of 15 and 17 Å were computed for each configuration. In Table 2, the average of the MADs is given for all mesh size combinations. For both active regions, the accuracy of the electrostatic forces seems to be rather independent of the mesh sizes. Within 15 Å of the center of the inner region, computation of the electrostatic forces based on the potential from the PB equation is quite accurate with average MADs around 4×10^{-5} to 8×10^{-5} au. The average deviation increases by a factor of 4−5 if the radius of the active region is extended to 17 Å. For each active region, there is only a very weak correlation between accuracy and mesh size. However, the accuracy strongly depends on the size of the active region. In Figure 2, the average MAD for one mesh size combination (0.6 and 1.75 Å) is plotted as a function of the radius of the active region. This figure shows that the accuracy is very high for radii of up to 16 Å, then the deviation increases

Table 2. Average Mean Absolute Deviation (10^{-4} au) of the Electrostatic Forces Computed from the PB Electrostatic Potential for Different Mesh Sizes of the Inner and Outer Grid Used for Solving the PB Equation[a]

	inner grid size (Å)			
outer grid size (Å)	0.25	0.40	0.60	0.80
	15 Å Active Region			
1.00	0.63	0.34	0.52	0.72
1.25	0.52	0.64	0.58	0.58
1.50	0.77	0.68	0.60	0.59
1.75	0.76	0.79	0.76	0.75
2.00	0.74	0.39	0.55	0.76
2.50	0.86	0.66	0.59	0.61
	17 Å Active Region			
1.00	2.13	1.98	2.36	2.85
1.25	2.02	2.27	2.42	2.72
1.50	2.28	2.31	2.43	2.73
1.75	2.27	2.42	2.58	2.87
2.00	2.26	2.03	2.39	2.88
2.50	2.37	2.28	2.43	2.75

[a] Active regions with radii of 15 and 17 Å were chosen.

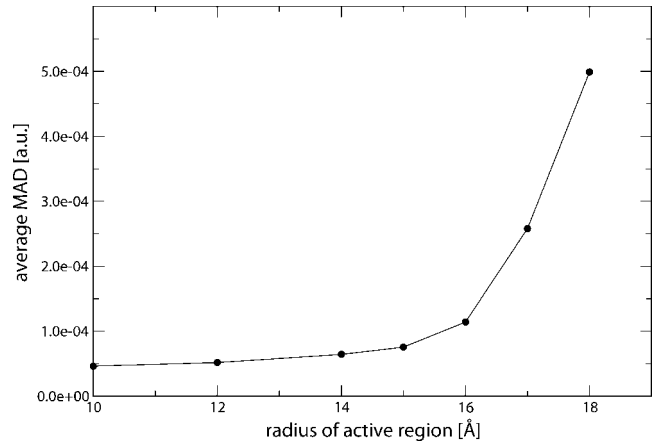

Figure 2. Average mean absolute deviation (MAD) of the electrostatic forces of all atoms inside the active region as a function of the radius of the active region. Mesh sizes of 0.6 and 1.75 Å were used for the inner and outer grid, respectively. The radius of the inner region is 18 Å (see text).

strongly. This behavior is identical for all mesh size combinations.

Keeping in mind the size of the inner region (18 Å, see above), we conclude that a grid-based PB potential is not adequate to represent the details of the electrostatic potential of the outer region in close proximity to the outer region. Although these inaccuracies are likely to have only an insignificant effect on the region of interest if the size of the inner region is adequate, we recommend to keep all atoms in the outer 2−3 Å layer of the inner region fixed, since such a frozen layer will increase the reliability of the GSBP. A mesh size of 0.6 Å for the inner grid seems to provide an ideal tradeoff between accuracy and computational costs. The results are not clearcut concerning the outer grid mesh size where the accuracy seems to be rather independent of the mesh size. This indicates that the electrostatic potential is only slowly varying at the boundary of the inner grid. To be on the safe side, we opted for an outer grid spacing of 1.75 Å.

4.2. Reaction Field Matrix. In section 2.3, three approaches to accelerate computation of the reaction field matrix were presented. While the modified stripping technique provides speed-up without loss of accuracy, the CIR and the linear interpolation approaches are approximations to define the boundary values more efficiently. Therefore, the computational savings and the associated loss of accuracy of these methods have to be analyzed.

For this assessment, one configuration of our model system with an inner region of 2978 atoms was selected. Spherical harmonics with multipole moments up to 20th order ($L = 0-19$), i.e., 400 basis functions, were used to represent the charge distribution. The previously determined best mesh size combination of 0.6 and 1.75 Å for the inner and outer grid was employed. The accuracy of the reaction field matrix was assessed by comparing the GSBP results for the solvation free energy of the inner region (ΔW^{ii}_{elec}) to the results of a finite-difference solution of the PB equation without a basis set representation, i.e., in the complete basis set limit.

The accuracy and the costs for computation of the reaction field matrix were tested for CIR factor values of 1.0, 1.5, 2.0, 2.5, and 3.0 in combination with the standard Debye−Hückel (DH) method and the DH expression with linear interpolation (DHLI). The results are given in Table 3. The combination of a CIR factor of 1.0 with DH boundary values corresponds to the standard GSBP method that reproduced the free solvation energy very well. With the selected basis set, about 97% of the free solvation energy is recovered. These results are certainly satisfying and support the finding of Im et al. that the solvation free energy is sufficiently converged with a basis set of this size.[38] If the CIR factor is increased to 1.5, 2.0, or 2.5, the deviation increases by only 0.2 kcal/mol from 3.73 to 3.93 kcal/mol. At the same time, computational costs are reduced by 54% from 8.22 h to only 3.79 h. The DHLI method proves to be similarly efficient. With a CIR factor of 1.0 and DHLI boundary values, the deviation increases by only 0.01 kcal/mol relative to DH boundary values and the computational costs are reduced by 45% to 4.53 h. Unfortunately, these two methods cannot be combined without loss of efficiency. The DHLI method in combination with a CIR factor of 2.5 yields a deviation of 3.94 kcal/mol (i.e., 0.21 kcal/mol higher relative to the standard GSBP method), but the computational costs are merely reduced from 3.79 to 3.38 h relative to the DH method with a CIR factor of 2.5. Considering the computation time for zero boundary values, it is understandable that the CIR and the DHLI method cannot be combined without loss of efficiency. Using zero boundary values, the relative computation time drops to 40.65%. Hence, in a standard reaction field matrix computation, about 60% of the computation time is used to define the boundary values. As either method, DHLI or CIR, reduces the computational costs for this step to only a fraction, combining the two methods gives only marginal extra savings. Overall, the combination of DHLI with a CIR factor of 2.5 reduces the computational costs by about 60% with minimal loss of accuracy. We also note that the computation time for this calculation increases by 10% without modified stripping (Table 3).

Table 3. Accuracy and Computational Costs of the Reaction Field Matrix Calculation Using Different Approximations To Define the Boundary Values (See Text)

boundary[a]	CIR factor	ΔW^{ii}_{elec} (kcal/mol)	deviation[b] (kcal/mol)	time (h)	rel time (%)
DH	1.0	−120.27	3.73	8.22	100.00
DH	1.5	−120.07	3.93	4.83	58.81
DH	2.0	−120.10	3.91	4.01	48.78
DH	2.5	−120.08	3.93	3.79	46.08
DH	3.0	−119.21	4.80	3.56	43.28
DHLI	1.0	−120.26	3.74	4.53	55.11
DHLI	1.5	−120.06	3.95	3.68	44.80
DHLI	2.0	−120.08	3.92	3.54	43.05
DHLI	2.5	−120.06	3.94	3.38	41.19
DHLI no MS[c]	2.5	−120.06	3.94	3.71	45.17
DHLI	3.0	−119.20	4.81	3.33	40.54
ZERO	−	−116.38	7.62	3.34	40.65

[a] DH, Debye−Hückel; DHLI, Debye−Hückel with linear interpolation; ZERO, all boundary values are set to zero. [b] Deviation = calcd − ref. [c] No modified stripping.

In summary, we have found that the GSBP yields reliable results for the electrostatic potential and the free solvation energy at moderate computational costs using a recommended parameter set with an inner grid spacing of 0.6 Å, an outer grid spacing of 1.75 Å, a CIR factor of 2.5, and DHLI boundary values.

Concluding this section, it seems worthwhile to reiterate the reference that was used to assess the performance of the GSBP. We have confirmed that the GSBP provides an accurate representation of the electrostatic potential that arises from the fixed outer region point charges and the PDC. This does not necessarily imply that a biomolecular simulation with the GSBP will be realistic in a chemical sense. Whether application of the GSBP is reasonable, and which choice is appropriate for physical parameters like the size of the inner region or the dielectric constants, is highly system-specific and beyond the scope of this study.

5. GSBP Efficiency

As the application of the GSBP is linked with a significant overhead, it is of interest to quantify the computational costs

Table 4. Computation Times Related to Nontruncated Coulombic Electrostatics and the GSBP Approach for Different System Sizes[a]

		computation time [s]							
radius (Å)	atoms	overhead	normal step	GSBP step	QM saving	MM saving	$Q_n/Q_n^x/\Omega/\Gamma$[b]	saving	steps[c]
25.0	7205	12166.1	2.0	5.4	−0.2	1.3	4.5	−3.4	−
27.5	9632	12104.2	3.8	5.4	−0.1	3.0	4.5	−1.6	−
30.0	12449	12243.9	6.2	5.4	0.0	5.3	4.5	0.8	15461
32.5	15806	12538.2	8.8	5.2	0.1	7.8	4.3	3.6	3503
35.0	19670	12399.3	13.3	5.3	0.3	12.1	4.4	8.0	1555
37.5	24110	12590.8	18.7	5.3	0.5	17.3	4.3	13.4	937
40.0	29234	12511.0	25.2	5.4	0.7	23.6	4.5	19.8	633
42.5	35042	12761.3	33.4	5.8	0.9	31.6	4.9	27.6	462
45.0	41468	12697.4	43.7	5.3	1.1	41.6	4.4	38.4	331

[a] Single-step computation times are average values from a sample of 100 steps. [b] Computation of additional terms related to the GSBP. [c] Number of steps necessary to compensate for the GSBP overhead.

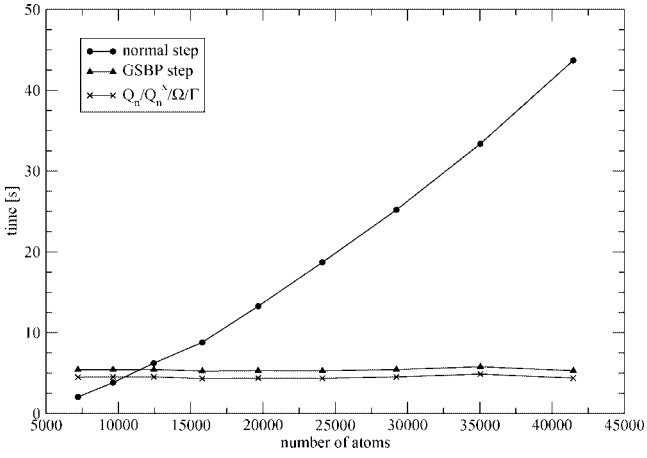

Figure 3. Computation times (s) for a single MD step using nontruncated Coulombic (normal step) or GSBP electrostatics (GSBP step) as a function of the system size. Furthermore, the computation times for the GSBP-related terms ($Q_n/Q_n^x/\Omega/\Gamma$) are plotted.

and savings related to the GSBP. In this section, the efficiency of the GSBP is documented for model systems of different sizes that were generated by solving one threonine molecule in TIP3P water balls with radii increasing from 25 to 45 Å. As in the previous calculations, the inner region was centered on the C_α carbon of threonine and contains all molecules with any atom within 18 Å of the center. While the inner region consists of 2738 atoms for all models, the overall system size increases from 7205 atoms to 41 468 atoms with increasing radius.

A detailed analysis of the computation times related to a MD simulation using either a standard approach with nontruncated Coulombic electrostatics or the GSBP is given in Table 4 and illustrated in Figure 3. This data provides interesting insights into the applicability and efficiency of the GSBP. First of all, the computation time for the GSBP overhead, i.e., calculation of the reaction field matrix and the static field, is almost constant and increases only slightly from 3.4 to 3.5 h when increasing the system size by a factor of 6. Also the computation time of a single GSBP MD step is almost constant at 5.4 s. For a standard MD step with full electrostatics in contrast, the computation time increases from 2.0 to 43.7 s. Accordingly, impressive savings per step can be achieved if the GSBP is used for extended systems.

However, the GSBP is not always more efficient than full electrostatics. For the two smallest systems, even a single MD step is computationally more expensive with the GSBP (in addition to the initial overhead). This can be attributed to two factors. First, with the GSBP several additional terms have to be computed for each step, such as the Ω and Γ matrices that allow interaction with the QM code, the multipole moments, Q_n, and their derivatives, Q_n^x (see eqs 19 and 23). Especially, the computation of all multipole moment derivatives for each degree of freedom is laborious and increases the GSBP step time by roughly 4.5 s for all system sizes. Second, evaluation of the QM energy and gradient is computationally more expensive with the GSBP, since the QM part takes 0.3 s with the standard approach and 0.5 s with the GSBP. This can be traced back to the calculation of the SCF density derivatives that is not necessary in a pure QM/MM calculation. However, these factors are dominant only for small systems. With increasing system size, evaluation of the QM energy and gradient becomes more efficient in the GSBP, due to the fact that the calculation of the numerous one-electron integrals in the standard electronic embedding procedure becomes more expensive than the solution of the CPSCF equation for large systems with 12 000 atoms and more. Moreover, we note that introduction of coupled Mulliken charge derivatives (to ensure accurate gradients) increases the computational costs of the GSBP method only marginally. The computation time for the MM part remains constant at about 1.0 s when using the GSBP, providing the main contribution to the GSBP savings.

Overall, in the chosen example, we start to see minor savings for a system with 12 500 atoms. Assuming an MD step size of 1 fs, the first 15 ps of simulation time are needed to compensate for the GSBP overhead, and afterward the computation time per step is reduced by 13%. Hence, in typical semiempirical QM/MM MD simulations, the breakeven point between the GSBP and Coulombic electrostatics without truncations appears at a system size of around 12 500 atoms. Significant savings are achieved for larger systems. In simulations of the 37.5 Å system with 24 110 atoms, only 937 steps are necessary to compensate for the GSBP overhead, and subsequently, the computation time per step decreases by more than 70% from 18.7 to 5.3 s. For larger systems even more impressive savings are observed (Table

4). Since in theoretical biochemistry one is frequently interested in QM/MM simulations of biomolecular systems with 25 000 atoms and more, the GSBP method offers an efficient approach to perform such simulations at a fraction of the computational costs compared to Coulombic electrostatics without truncation.

6. Conclusions

In this work, we have presented the implementation of the GSBP for QM/MM approaches using NDDO-based semiempirical QM methods. Moreover, three methods to accelerate computation of the reaction field matrix were introduced: coarsening of the inner region, linear interpolation of Debye−Hückel boundary values, and modified stripping. We found that a combination of these methods reduces the computational costs for assembling the reaction field matrix by 60% with only minimal loss of accuracy. Furthermore, we studied the accuracy of the GSBP as a function of its inherent parameters, and defined a set of parameter values that offer an ideal tradeoff between accuracy and computational costs. On the basis of theses values, the computational overhead and the savings of the GSBP were quantified in QM/MM MD simulations for model systems containing from around 7000 to more than 40 000 atoms. The breakeven point where the savings in comparison to nontruncated Coulombic electrostatics roughly compensate for the overhead was determined at around 12 500 atoms. For larger systems, the GSBP showed an impressive performance. Compensation for the overhead was achieved in less than 1000 MD steps, and subsequently, the computation time per step decreased by 70% and more compared to nontruncated Coulombic electrostatics.

The GSBP is thus an efficient and accurate method to perform semiempirical QM/MM MD simulations on large biomolecular systems without neglecting or truncating long-range electrostatics if the outer layer of the inner region is fixed. It is clearly desirable to achieve similar computational savings by applying the GSBP in combination with higher-level QM/MM methods. Work in this direction is currently under way in our laboratory.

Acknowledgment. This work was supported by the Max Planck Initiative on Multiscale Materials Modelling. T.B. gratefully acknowledges a Kekulé scholarship from the Fonds der Chemischen Industrie.

References

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

(2) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483–3492.

(3) Cisneros, G. A.; Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2003**, *125*, 10384–10393.

(4) Ridder, L.; Rietjens, I. M. C. M.; Vervoort, J.; Mulholland, A. J. *J. Am. Chem. Soc.* **2002**, *124*, 9926–9936.

(5) Kaminski, G. A.; Jorgensen, W. L. *J. Phys. Chem. B* **1998**, *102*, 1787–1796.

(6) Acevedo, O.; Jorgensen, W. L.; Evanseck, J. D. *J. Chem. Theory Comput.* **2007**, *3*, 132–138.

(7) Rod, T. H.; Ryde, U. *J. Chem. Theory Comput.* **2005**, *1*, 1240–1251.

(8) Strajbl, M.; Hong, G.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 13333–13343.

(9) Senn, H. M.; Thiel, S.; Thiel, W. *J. Chem. Theory Comput.* **2005**, *1*, 494–505.

(10) Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. *J. Chem. Theory Comput.* **2006**, *2*, 452–461.

(11) Senn, H. M.; Thiel, W. *Top. Curr. Chem.* **2007**, *268*, 173–290.

(12) Warshel, A.; Papazyan, A. *Curr. Opin. Struct. Biol.* **1998**, *2*, 211–217.

(13) Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Struct.* **1999**, *28*, 155–179.

(14) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509–521.

(15) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186–195.

(16) Ewald, P. *Ann. Phys.* **1921**, *369*, 253–287.

(17) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(18) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(19) Nam, K.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2–13.

(20) Gao, J.; Alhambra, C. *J. Chem. Phys.* **1997**, *107*, 1212–1217.

(21) Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019–1031.

(22) Hünenberger, P. H.; McCammon, J. A. *Biophys. Chem.* **1999**, *78*, 69–88.

(23) Hünenberger, P. H.; McCammon, J. A. *J. Chem. Phys.* **1999**, *110*, 1856–1872.

(24) Kuwajima, S.; Warshel, A. *J. Chem. Phys.* **1988**, *89*, 3751–3759.

(25) Weber, W.; Hunenberger, P. H.; McCammon, J. A. *J. Phys. Chem. B* **2000**, *104*, 3668–3675.

(26) Friedman, H. L. *Mol. Phys.* **1975**, *29*, 1533–1543.

(27) Wang, L.; Hermans, J. *J. Phys. Chem.* **1995**, *99*, 12001–12007.

(28) Berkowitz, M.; McCammon, J. A. *Chem. Phys. Lett.* **1982**, *90*, 215–217.

(29) Brooks, C. L., III; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312–6325.

(30) Brunger, A.; Brooks, C. L., III; Karplus, M. *Chem. Phys. Lett.* **1984**, *105*, 495–500.

(31) Brunger, A.; Brooks, C. L., III; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 8458–8462.

(32) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100–3107.

(33) Alper, H.; Levy, R. M. *J. Chem. Phys.* **1993**, *99*, 9847–9852.

(34) Essex, J. W.; Jorgensen, W. L. *J. Comput. Chem.* **1995**, *16*, 951–972.

(35) Warshel, A.; King, G. *Chem. Phys. Lett.* **1985**, *121*, 124–129.

(36) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451–5459.

(37) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050–9063.

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

(39) Banavali, N. K.; Im, W.; Roux, B. *J. Chem. Phys.* **2002**, *117*, 7381–7388.

(40) Schaefer, P.; Riccardi, D.; Cui, Q. *J. Chem. Phys.* **2005**, *123*, 014905/1–14.

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

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

(43) Riccardi, D.; Cui, Q. *J. Phys. Chem. A* **2007**, *111*, 5703–5711.

(44) Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 17715–17733.

(45) König, P. H.; Ghosh, N.; Hoffmann, M.; Elstner, M.; Tajkhorshid, E.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. A* **2006**, *110*, 548–563.

(46) Ma, L.; Cui, Q. *J. Am. Chem. Soc.* **2007**, *129*, 10261–10268.

(47) Zhu, X.; Jethiray, A.; Cui, Q. *J. Chem. Theory Comput.* **2007**, *3*, 1538–1549.

(48) Riccardi, D.; Konig, P.; Prat-Resina, X.; Yu, H.; Elstner, M.; Frauenheim, T.; Cui, Q. *J. Am. Chem. Soc.* **2006**, *128*, 16302–16311.

(49) Gonzalez-Lafont, A.; Truong, T. N.; Truhlar, D. G. *J. Phys. Chem.* **1991**, *95*, 4618–4627.

(50) Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231–236.

(51) Bash, P. A.; Ho, L. L.; MacKerell, A. D., Jr.; Levine, D.; Hallstrom, P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3698–3703.

(52) Lau, E. Y.; Kahn, K.; Bash, P. A.; Bruice, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9937–9942.

(53) Ridder, L.; Rietjens, I. M. C. M.; Vervoort, J.; Mulholland, A. J. *J. Am. Chem. Soc.* **2002**, *124*, 9926–9936.

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

(55) Amin, E. A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 75–85.

(56) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins* **1986**, *1*, 47–59.

(57) Mulliken, R. S. *J. Chem. Phys.* **1962**, *36*, 3428–3439.

(58) Sherwood, P.; et al. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 1–28.

(59) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

(60) Patchkovskii, S.; Thiel, W. *J. Comput. Chem.* **1996**, *17*, 1318–1327.

(61) Yamaguchi, Y.; Osamura, Y.; Goddard, J. D.; Schaefer, H. F., III *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory*; Oxford University Press: Oxford, UK, 1994; pp 128−132.

(62) Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1987**, *9*, 327–335.

(63) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435–445.

(64) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(65) Smith, W.; Forester, T. *J. Mol. Graph.* **1996**, *14*, 136–141.

(66) MacKerell, A. D., Jr.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(67) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterlig, W. T. *Numerical Recipes in C*; Cambridge University Press: London, 1988; pp 673−680.

CT800193A