JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Solvent Boundary Potentials for Hybrid QM/MM Computations Using Classical Drude Oscillators: A Fully Polarizable Model

Eliot Boulanger and Walter Thiel*

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

ABSTRACT: Accurate quantum mechanical/molecular mechanical (QM/MM) treatments should account for MM polarization and properly include long-range electrostatic interactions. We report on a development that covers both these aspects. Our approach combines the classical Drude oscillator (DO) model for the electronic polarizability of the MM atoms with the generalized solvent boundary Potential (GSBP) and the solvated macromolecule boundary potential (SMBP). These boundary potentials (BP) are designed to capture the long-range effects of the outer region of a large system on its interior. They employ a finite difference approximation to the Poisson−Boltzmann equation for computing electrostatic interactions and take into account outer-region bulk solvent through a polarizable dielectric continuum (PDC). This approach thus leads to fully polarizable three-layer QM/MM-DO/BP methods. As the mutual responses of each of the subsystems have to be taken into account, we propose efficient schemes to converge the polarization of each layer simultaneously. For molecular dynamics (MD) simulations using GSBP, this is achieved by considering the MM polarizable model as a dynamical degree of freedom, and hence contributions from the boundary potential can be evaluated for a frozen state of polarization at every time step. For geometry optimizations using SMBP, we propose a dual self-consistent field approach for relaxing the Drude oscillators to their ideal positions and converging the QM wave function with the proper boundary potential. The chosen coupling schemes are evaluated with a test system consisting of a glycine molecule in a water ball. Both boundary potentials are capable of properly reproducing the gradients at the inner-region atoms and the Drude oscillators. We show that the effect of the Drude oscillators must be included in all terms of the boundary potentials to obtain accurate results and that the use of a high dielectric constant for the PDC does not lead to a polarization catastrophe of the DO models. Optimum values for some key parameters are discussed. We also address the efficiency of these approaches compared to standard QM/MM-DO calculations without BP. In the SMBP case, computation times can be reduced by around 40% for each step of a geometry optimization, with some variation depending on the chosen QM method. In the GSBP case, the computational advantages of using the boundary potential increase with system size and with the number of MD steps.

1. INTRODUCTION

Hybrid quantum mechanical/molecular mechanical (QM/MM) methods have become reliable tools for studying chemical reactions in large biomolecules.[1−7] Already the first such study[8] considered the embedding of a QM subsystem in a polarizable MM environment to be important for the proper description of enzymes. Since then, polarizable force fields (PFFs) have undergone much development and are now approaching maturity. They are being used increasingly in biomolecular simulations, and highly optimized PFFs for such applications are expected to be available soon.[9,10]

There are several ways to simulate the polarizability of MM atoms.[9−11] These include induced dipoles (or multipoles),[12−14] fluctuating charges,[15−19] and classical Drude oscillators (DOs).[20−26] The DO approach is also called the shell model[27−29] or charge-on-spring (COS) model.[11,30−33] In our early work, we represented MM polarization by induced dipoles at the MM atoms.[34] More recently, we have adopted the DO (COS) model, in view of its inherent simplicity and its widespread use in PFF development.[23,35−38] The model consists of a mobile charge linked to a polarizable MM atom by a spring; a charge of the same magnitude and opposite sign is added at the nucleus of this atom, so that these two virtual charges form a dipole. The mobile charge moves in response to the electrostatic interactions with the environment, thus simulating MM polarizability. We have included the DO model in the QM/MM ChemShell software using the GROMOS COS force field.[39] Other interfacing methods have been discussed for the CHARMM DO force field.[40,41]

In a QM/MM framework, the influence of MM polarization should be especially important for processes that involve charged or very polar species and significant charge relocation. In these cases, long-range electrostatic interactions are also expected to play a prominent role.[42] This calls for the development of treatments that cover both these aspects in a balanced manner.

Long-range electrostatic interactions can be taken into account in classical MM simulations using several well established techniques. At the QM/MM level, two such techniques have been implemented by different groups, namely periodic boundary conditions and boundary potentials. Periodic boundary conditions have been applied using Ewald summation.[43−45] This approach will require huge unit cells in biomolecular work (containing the large nonperiodic biomolecule and a solvent environment of sufficient size) and may thus be less practical at the QM/MM level. The alternative boundary potential approach circumvents this problem by considering only a restricted number of atoms explicitly and

Received: August 16, 2012
Published: October 12, 2012

   dx.doi.org/10.1021/ct300722e | J. Chem. Theory Comput. 2012, 8, 4527−4538

representing the distant environment by a continuum model. To be more specific, it splits the system into an explicit inner region (including the QM subsystem as well as the adjacent MM part of the macromolecule) and an implicit outer region (consisting of the distant MM atoms of the macromolecule and the bulk solvent). The long-range electrostatic effects of the outer region are captured by a boundary potential that represents the influence of the discrete MM charges in this region and of the bulk solvent treated as a polarizable dielectric continuum (PDC). This method is well suited for describing localized process such as those commonly studied with QM/MM methods.

The generalized solvent boundary potential (GSBP)[46−48] and the solvated macromolecule boundary potential (SMBP)[49,50] allow the use of irregularly shaped dielectric boundaries between the macromolecule and bulk solvent. This feature is important in QM/MM computations of enzymes as the protein and the surrounding water possess very different dielectric constants.[51] GSBP was originally designed for use in molecular dynamics (MD) simulations. At the QM/MM level, it has up to now only been interfaced with semiempirical methods. SMBP was developed for geometry optimizations with any kind of QM method and can thus be used to compute QM/MM potential energy surfaces (PESs) also with high-level QM methods. GSBP and SMBP complement each other in QM/MM free energy perturbation (FEP) calculations of free energy differences.[52,53] These can be used to estimate the entropic contribution of the environment by sampling over the MM degrees of freedom.

Polarizable force fields have already been interfaced with PDC models, with results that are promising in terms of accuracy,[54−59] and excited-state properties of small molecules have been studied by embedding the solute molecule (QM) in a few explicit polarizable solvent molecules (PFF) and a bulk solvent (PDC).[56,60] In our treatment, we combine a QM part that is intrinsically polarizable, an explicit MM region described by a PFF, and a solvent represented by a PDC, which leads to a fully polarizable three-layer model. We also note that boundary potentials have the general advantage of being quite efficient compared with full QM/MM treatments, since they consider only a small part of the system explicitly. This reduced size of the explicit MM region can be particularly beneficial when using PFFs which normally require some kind of iterative scheme to determine the proper MM polarization.

The purpose of this paper is to present combination schemes for Drude oscillators with GSBP and SMBP in order to obtain a fully polarizable three-layer QM/MM-DO/BP model at a reasonable computational cost.

2. THEORY

The system under study is separated into different spatial regions. As usual in QM/MM methods, there is a central QM region surrounded by an MM region. The latter is further partitioned into an explicit inner part (treated atomistically at the MM level) and an implicit outer part (represented by the boundary potential). All these subsystems interact with each other, and the total energy is given by the following additive expression:

$$E_{\mathrm{tot}} = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{BP}} + E_{\mathrm{QM/MM}} + E_{\mathrm{QM/BP}} + E_{\mathrm{MM/BP}} \tag{1}$$

In our implementation in the ChemShell package,[61,62] E_{QM} is the energy of the QM part obtained with any available QM method, and E_{MM} is the energy obtained from any MM force field function interfaced with the program. E_{BP} denotes the energy contribution from the boundary potential. The three last terms arise from the interactions between the subsystems. In the following, we describe the interactions that occur when using the polarizable DO force field for the MM part and the GSBP or SMBP for the boundary potential.

2.1. Drude Oscillators in a QM/MM Framework. The DO model aims at simulating the electronic polarizability at the MM level. It represents the induced dipole at every polarizable atom by two charges of the same magnitude (q) and opposite sign linked by a harmonic spring. The first charge is located at the nucleus of the atom, while the second one is mobile. Polarization arises from the competition between the forces acting on the mobile charge, which are due to the spring and the electrostatic interactions with the environment. The optimum position (d) of the mobile charge (Drude particle) is obtained by requiring that these two forces compensate each other.

$$\frac{\partial(U_{\mathrm{spring}} + U_{\mathrm{elec}})}{\partial d} = 0 \tag{2}$$

The electrostatic potential energy U_{elec} is obtained by summing over all point-charge interactions applying Coulomb's law. The potential energy U_{spring} of the harmonic spring is evaluated using a force constant (k_d) that is generally defined in terms of the polarizability (α) of the corresponding atom:

$$k_d = \frac{q^2}{\alpha} \tag{3}$$

In the CHARMM force field, k_d is always fixed to 1000 kcal mol^{-1} Å^{-2} in order to maintain a small d value and to keep the point dipole approximation valid.[26] The implementation of the DO terms within a standard force field involves only modifications in the electrostatic part of the MM potential energy function.

$$E_{\mathrm{MM}}^{\mathrm{elec}} = \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}} + \sum_i \sum_{j'} \left(\frac{q_i q_{j'}}{r_{ij'}} - \frac{q_i q_{j'}}{r_{ij}} \right) + \sum_{i'} \sum_{j'>i'} \left(\frac{q_{i'} q_{j'}}{r_{i'j'}} + \frac{q_{i'} q_{j'}}{r_{ij}} \right) - \sum_{i'} \sum_{j'} \left(\frac{q_{i'} q_{j'}}{r_{ij'}} \right) + \frac{1}{2} \sum_{i'} k_{d,i'} d_{i'}^2 \tag{4}$$

Here, indices i and j run over atoms, the prime denotes a DO term, and r is the distance between the two corresponding centers. In the CHARMM force field, several other terms have been included in an attempt to properly simulate the electronic distribution and the dipole response of a molecule.[63] The 1−2 and 1−3 interactions between Drude particles (at atoms located one or two covalent bonds away from each other) are screened by applying the Thole function:[64]

$$1 - \left(1 + \frac{T_{ij}}{2}\right) \exp(-T_{ij}) \tag{5}$$

$$T_{ij} = t\left(\frac{r_{ij}}{\sqrt[6]{\alpha_i \alpha_j}}\right) \tag{6}$$

where t is the Thole parameter. Interactions at larger distances (with three or more bonds in between) are evaluated without any screening. Also, the charge of a heteroatom (without the DO contribution) can be split and partly located at one or two nearby positions (fixed in terms of internal coordinates) to represent lone pairs. The simultaneous presence of such lone pair charges and Drude particles may allow the simulation of anisotropic polarizability.[65] Another correction has been introduced to avoid the so-called polarization catastrophe, which is caused by strong Coulombic interactions at close distances that may lead to an excessive drift of the Drude particle.[11] The associated anharmonic hyperpolarizability damping term is taken into account only if d is higher than a predefined limit (typically 0.2 Å).[24]

The implementation of the DO model within a QM/MM framework has been discussed for GROMOS (COS model)[39] and more recently also for CHARMM.[40] In both cases, the Drude oscillators are included in the QM computation by modifying the one-electron terms in the Fock matrix, in complete analogy to the classical MM point charges. The Drude particle gives rise to extra one-electron terms, while the compensating charge at the nucleus of the polarizable atom is taken into account by adjusting the corresponding atomic charge.

To determine the position of the Drude particle, and thus obtain the induced dipole, the left-hand side of eq 2 must be minimized. The required electrostatic potential is evaluated from the electric field (ϕ) at the position of the Drude particle, which is composed of QM, MM, and DO contributions. The force (F) exerted on Drude particle i' by a given component of the electric field can be written as

$$F_{d_{i'}} = \frac{\alpha_{i'}}{q_{i'}}(\varphi_{i'}^{\mathrm{MM}} + \varphi_{i'}^{\mathrm{QM}} + \varphi_{i'}^{\mathrm{DO}}) \tag{7}$$

The contributions $\varphi_{i'}^{\mathrm{MM}}$, $\varphi_{i'}^{\mathrm{QM}}$ and $\varphi_{i'}^{\mathrm{DO}}$ to the electric field are interdependent, and their computation thus calls for a dual self-consistent-field (SCF) approach[39] as indicated in Figure 1. We followed the same implementation as in our previous work[39]

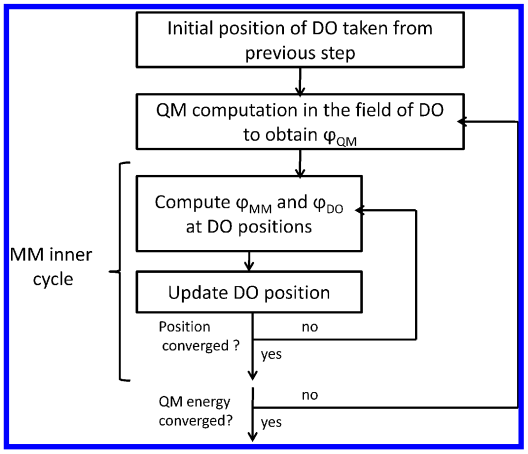

Figure 1. Dual SCF approach for determining the DO positions in a QM/MM framework. The outer SCF procedure converges the QM wave function, while the inner one converges the DO positions in the field of each other and of the MM atoms. See text for details.

but introduced an additional option: when using the CHARMM force field, the electric field is computed at the position of the Drude particle (to conform to CHARMM conventions), whereas it is calculated at the corresponding atomic position for the GROMOS COS model.[32] In our computational approach, we first evaluate $\varphi_{i'}^{\mathrm{QM}}$ for a set of fixed DO positions, which are then updated in an iterative scheme through an MM inner cycle. In this cycle, $\varphi_{i'}^{\mathrm{MM}}$ and $\varphi_{i'}^{\mathrm{DO}}$ are computed for the given geometry, and the DO positions are updated using the forces from eq 7. This inner cycle is iterated until the DO positions are converged in the field of the given QM wave function (as judged by their maximum and average changes). Thereafter, the convergence of the QM energy is checked. If this is not the case, the process is restarted at the first step by recalculating $\varphi_{i'}^{\mathrm{QM}}$ and going again through the inner cycle, until full overall convergence is achieved.

A microiterative scheme has also been proposed, with one update of the DO positions in each step of the QM SCF procedure.[40] This approach is computationally more efficient but necessitates the modification of the QM program and thus cannot be applied directly for any QM code.

The iterative relaxation of the Drude particle to its minimum energy during each step of an optimization or a sampling procedure is accurate but can become very expensive if there are a large number of polarizable atoms in the system. Therefore, this scheme may no longer be practical for long MD runs.

One key advantage of Drude oscillators is that they can be treated as dynamical degrees of freedom. Their direct inclusion in a standard MD scheme is problematic, however, since the oscillations of the Drude particles would have very high frequencies and would thus require very small integration time steps. This problem can be overcome by the use of extended Lagrangian dynamics.[20] In this scheme, the overall motion of the atoms and the relative motion within the atom−DO pairs are separated and propagated in a coupled manner as follows. A small mass (m_{D}) is taken from the polarizable atom and assigned to the Drude particle. The polarizable site is propagated in the overall MD scheme using the center of mass (R_i) of the atom−DO pair and the total mass of the atom (m_i). The relative motion in the atom−DO pair is propagated using an extended mass $m_i' = m_{\mathrm{D}}(1 - m_{\mathrm{D}}/m_i)$.[20] In the NVT ensemble, this leads to the so-called cold DO model.[20] The dynamics of the system is controlled by a thermostat of the desired temperature (T) while the relative motion within the DOs is frozen at temperature $T_* = 1$ K to avoid high-frequency oscillations. At each MD step, the DO positions are not fully converged, but it is assumed that, during the sampling process, they will oscillate around their respective minima. The equations of motions, using two Nosé−Hoover thermostats, are

$$m_i \ddot{R}_i = F_{\mathrm{R},i} - m_i \dot{R}_i \dot{\eta} \tag{8.a}$$

$$m_i' \ddot{d}_i = F_{d,i} - m_i' \dot{d}_i \dot{\eta}_* \tag{8.b}$$

$$Q\ddot{\eta} = \sum_j m_j \dot{R}_j^2 - N_{\mathrm{f}}\, k_{\mathrm{B}} T \tag{8.c}$$

$$Q_* \ddot{\eta}_* = \sum_j m_j' \dot{d}_j^2 - 3N_{\mathrm{D}}\, k_{\mathrm{B}} T_* \tag{8.d}$$

The indices i and j run over all atoms. The variables associated with the thermostats are the inertia factor Q and the friction

dx.doi.org/10.1021/ct300722e | J. Chem. Theory Comput. 2012, 8, 4527−4538

coefficients $\dot{\eta}$ that are obtained by solving eqs 8.c and 8.d. The subscript "$*$" refers to the thermostat for the relative motion of the atom−DO pair. N_f is the total number of degrees of freedom excluding constrained components and DOs, and N_D is the number of DOs. Note that if an atom is not polarizable, m_i' is zero and $R_i = r_i$. This extended system can be propagated in time using a velocity Verlet scheme.[66,67] We have implemented this approach for the NVT ensemble. Analogous MD runs can also be performed with the NPT ensemble using a barostat for the extended system,[20] but this variant was not implemented here because the use of BPs constrains the system to a constant volume (see below). We also did not consider other methods that propagate classical representations of electronic polarization using always stable estimator-corrector algorithms.[40,68,69]

The implementation of this integration scheme for QM/MM molecular dynamics necessitates the computation of the forces that act on the different particles. The contributions to these forces from the QM and MM regions are additive:

$$F_{R,i} = -\frac{\partial(U^{QM} + U^{MM})}{\partial r_i} - \frac{\partial(U^{QM} + U^{MM})}{\partial r_{i'}} \tag{9.a}$$

$$F_{d,i} = \left(\frac{m_D}{m_i}\right)\frac{\partial(U^{QM} + U^{MM})}{\partial r_i} - \left(1 - \frac{m_D}{m_i}\right) \times \frac{\partial(U^{QM} + U^{MM})}{\partial r_{i'}} \tag{9.b}$$

The QM contributions are computed only once per MD step using the fully converged QM SCF wave function for the given configuration of MM atoms and DOs. Note that forces are computed with the charges of each polarizable center reduced by the DO partial charge $q_{i'}$ (see above). QM atoms are propagated classically by treating them as dynamical degrees of freedom in the equation of motion of the general system; they are coupled to the same thermostat.

2.2. Boundary Potentials. A boundary potential simulates the electrostatic influence of an implicit infinite outer region on the explicit inner region of interest. Figure 2 illustrates this separation for a solvated protein in a QM/MM framework.

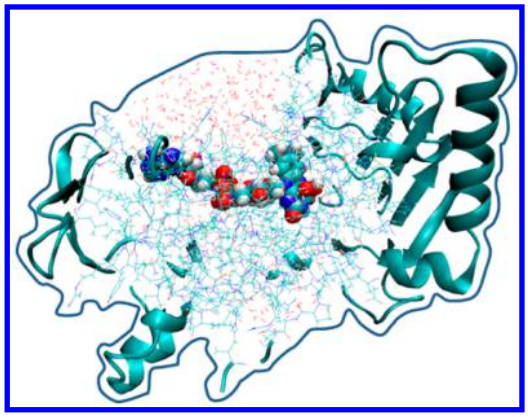

Figure 2. Schematic representation of the separation of regions in the GSBP and SMBP. The atoms in the QM region atoms are represented by their van der Waals radii. The MM atoms from the inner region are pictured explicitly by lines. The outer region of the macromolecule is symbolized by the ribbons. The region outside of the blue boundary line corresponds to the bulk solvent simulated by a PDC.

We use R to denote the generic coordinates of a macromolecule surrounded by N solvent molecules. The inner region consists of the inner part of the macromolecule (R_i) as well as n inner solvent molecules, while the outer region includes the outer part of the macromolecule (R_o) and the remaining $N - n$ solvent molecules. Statistical observations are assumed to depend only on the degrees of freedom of the inner region. They can thus be computed on the surface of its potential of mean force (PMF) by integrating out the degrees of freedom of the outer region:

$$e^{-\beta W(R_i,1,\ldots,n)} = \frac{1}{C}\int dR_0\, d(n+1)\ldots dN\, e^{-\beta U(R,1,\ldots N)} \tag{10}$$

Note that only configurations for which outer region atoms do not overlap with the inner region are considered here. By picking an appropriate normalization constant (C), Beglov and Roux demonstrated that the PMF is equivalent to the reversible work necessary to assemble the inner region inside the outer region.[70] They proposed to proceed stepwise, considering separately the different contributions to this assembly:

$$\Delta W = U + \Delta W_{cr} + \Delta W_{np} + \Delta W_{elec} \tag{11}$$

U is the potential energy of the isolated inner region, and the three following terms are the free energy contributions arising from configurational restrictions, nonpolar interactions, and electrostatic interactions, respectively. This approach is only valid if the configuration of the atoms in the outer region can be considered representative of the average of all possible configurations. It will thus be particularly suitable for studying processes localized in the center of the inner region, while its accuracy will decrease in the vicinity of the inner−outer boundary. It is commonly assumed that the configurational restrictions and nonpolar interactions will remain constant for a given system and that one may thus focus on the electrostatic contributions to the boundary potential. The following subsections describe two approaches to their determination.

2.2.1. Generalized Solvent Boundary Potential. The GSBP aims at approximating the electrostatic contribution to the PMF in a scheme suitable for MD simulations. There are two parts, arising from the direct Coulomb interaction of inner-region charges with the outer-region charges of the macromolecule and with the outer-region solvent molecules described by a PDC. The latter term can be expressed as the interaction of the inner-region point charges of the macromolecule (q_A) with the reaction field potential ϕ_{rf} at their position (r_A).

$$\Delta W_{elec}^{solv} = \frac{1}{2}\sum_A q_A \phi_{rf}(r_A) \tag{12}$$

The reaction field potential is defined as the difference of the electrostatic potentials in solution and in vacuum. It can be obtained by solving the linearized Poisson−Boltzmann (PB) equation for both situations using the corresponding dielectric constants (ε).

$$\nabla[\varepsilon(r)\nabla\phi(r)] - \bar{\kappa}^2(r)\phi(r) = -4\pi\rho(r) \tag{13}$$

Here, $\rho(r)$ is the charge density and $\bar{\kappa}(r)$ is the modified Debye−Hückel screening factor. A straightforward implementation would require solving the PB equation for every configuration, which would quickly become too expensive for typical MD runs. To overcome this problem, ΔW_{elec}^{solv} is

separated by splitting the charge distribution into an inner and outer part.

$$\Delta W_{\mathrm{elec}}^{\mathrm{solv}} = \Delta W_{\mathrm{elec}}^{\mathrm{outer-outer}} + \Delta W_{\mathrm{elec}}^{\mathrm{inner-outer}} + \Delta W_{\mathrm{elec}}^{\mathrm{inner-inner}} \tag{14}$$

The first term represents the interaction of the outer charge distribution with its self-induced reaction field. It is constant during the sampling and can thus be neglected or computed once and for all. The inner−outer contribution to the solvation free energy can be combined with the calculation of the inner-outer Coulomb interactions in an efficient scheme using the electrostatic potential of the outer region in solution ($\phi_{\mathrm{s}}^{\mathrm{outer}}(r)$).

$$\begin{aligned}\Delta W_{\mathrm{elec}}^{\mathrm{inner-outer}} + U_{\mathrm{elec}}^{\mathrm{inner-outer}} &= \sum_{A\in\mathrm{inner}} q_A \phi_{\mathrm{rf}}^{\mathrm{outer}}(r_A) + U_{\mathrm{elec}}^{\mathrm{inner-outer}} \\ &= \sum_{A\in\mathrm{inner}} q_A \phi_{\mathrm{s}}^{\mathrm{outer}}(r_A)\end{aligned} \tag{15}$$

Since the outer region is in a frozen configuration, its potential is constant during the simulation and can be calculated and stored once and for all, giving rise to a significant decrease of on-the-fly computational costs. The only terms remaining are thus the inner−inner contributions. An analytical solution for this part is provided by the Green's function (G_{rf}) that describes the inner-region reaction field potential.

$$\phi_{\mathrm{rf}}^{\mathrm{inner}}(r) = \int \mathrm{d}r' \rho_{\mathrm{i}}(r')\, G_{\mathrm{rf}}(r, r') \tag{16}$$

This formulation allows the projection of the inner charge distribution and of the Green's function onto the same set of basis functions $\{b_n\}$ with associated generalized multipole moments Q_n. The solvation free energy can be expressed as the matrix product of the reaction field matrix (M_{mn}) with these multipole moments, which yields the final expression for the GSBP.

$$\Delta W_{\mathrm{elec}}^{\mathrm{GSBP}} = \sum_{A\in\mathrm{inner}} \phi_{\mathrm{s}}^{\mathrm{outer}}(r_A) + \frac{1}{2}\sum_{mn} Q_m M_{mn} Q_n \tag{17}$$

The matrix M_{mn} can be computed once and for all at the beginning of the simulation. This requires solving the PB equation repeatedly, depending on the size of the basis set used.

In a QM/MM framework, the inner region also includes the QM part of the system. Its contribution is taken into account separately by splitting the $\Delta W_{\mathrm{elec}}^{\mathrm{GSBP}}$ expression into QM and MM parts.

$$\begin{aligned}\Delta W_{\mathrm{elec}}^{\mathrm{GSBP}} = &\sum_{A\in\mathrm{inner,MM}} \phi_{\mathrm{s}}^{\mathrm{outer}}(r_A) + \int \mathrm{d}r\, \rho^{\mathrm{QM}}(r)\, \phi_{\mathrm{s}}^{\mathrm{outer}}(r) \\ &+ \frac{1}{2}\sum_{mn} Q_m^{\mathrm{QM}} M_{mn} Q_n^{\mathrm{QM}} + \sum_{mn} Q_m^{\mathrm{QM}} M_{mn} Q_n^{\mathrm{MM}} \\ &+ \frac{1}{2}\sum_{mn} Q_m^{\mathrm{MM}} M_{mn} Q_n^{\mathrm{MM}}\end{aligned} \tag{18}$$

In previous GSBP implementations, the continuous QM charge distribution was represented by Mulliken charges.[46,48] This necessitates changes in the QM code when implementing the GSBP scheme.

2.2.2. Solvated Macromolecule Boundary Potential. The SMBP is a solvent boundary potential designed for geometry optimizations with any kind of QM method. It relies on the same approximations as the GSBP by using the same decomposition into an inner and outer region. However, as geometry optimizations require much fewer steps than MD runs, the PB equation is now solved at each step. Compared with GSBP, this saves the initial effort of computing the reaction field matrix (i.e., solving the PB equation typically 800 times for common basis sets). To allow the use of the SMBP with any QM/MM Hamiltonian, the interactions with the reaction field potential are computed separately for the QM and MM regions.

$$\Delta W_{\mathrm{elec}}^{\mathrm{SMBP}} = \int \mathrm{d}r\, \rho_{\mathrm{QM}}(r)\phi_{\mathrm{tot}}^{\mathrm{QM}}(r) + \int \mathrm{d}r\, \rho_{\mathrm{MM}}(r)\, \phi_{\mathrm{tot}}^{\mathrm{MM}}(r) \tag{19}$$

Here, $\phi_{\mathrm{tot}}^{\mathrm{QM}}(r)$ and $\phi_{\mathrm{tot}}^{\mathrm{MM}}(r)$ are the effective potentials experienced by the QM and MM regions, respectively.

$$\phi_{\mathrm{tot}}^{\mathrm{QM}}(r) = \phi_{\mathrm{s}}^{\mathrm{outer}}(r) + \phi_{\mathrm{rf}}^{\mathrm{inner-MM}}(r) + \frac{1}{2}\phi_{\mathrm{rf}}^{\mathrm{QM}}(r) \tag{20.a}$$

$$\phi_{\mathrm{tot}}^{\mathrm{MM}}(r) = \phi_{\mathrm{s}}^{\mathrm{outer}}(r) + \frac{1}{2}\phi_{\mathrm{rf}}^{\mathrm{inner-MM}}(r) \tag{20.b}$$

The reaction field potentials depend of the instantaneous configuration of the inner region and must thus be updated in every optimization step. For nonpolarizable MM point charges, the term $\phi_{\mathrm{rf}}^{\mathrm{inner-MM}}(r)$ can be computed by solving the PB equation once. On the other hand, $\phi_{\mathrm{rf}}^{\mathrm{QM}}(r)$ depends on the polarizable QM density, and hence a self-consistent reaction field procedure is needed to determine both. This involves the following steps: (1) With an initial guess for the QM charges, compute $\phi_{\mathrm{rf}}^{\mathrm{QM}}(r)$. (2) Assemble $\phi_{\mathrm{tot}}^{\mathrm{QM}}(r)$ and project it on a set of virtual charges distributed on a sphere around the inner region. (3) Evaluate the QM wave function in the field of these virtual charges and the inner-region MM point charges. (4) Determine ESP charges that represent the QM charge distribution well enough to generate a realistic electric field. (5) Loop over steps 2−4 until convergence is reached, i.e., until the QM reaction field potential changes from one iteration to the next one by less than a predefined criterion.

2.3. QM/MM-Based Combination of Polarizable Force Fields with Boundary Potentials. The use of a polarizable force field for the explicitly treated inner MM region in a QM/MM/BP setup leads to a three-layer approach which accounts for polarization effects in all layers. This introduces additional interdependences which will be described in the following.

To include Drude oscillators in the MM layer of a QM/MM/BP treatment, eq 11 needs to be supplemented with an additional term (ΔW_{pol}) that describes the free energy necessary to switch on the polarizability of the polarizable MM atoms in the field of the boundary potential. This term is expressed in different forms in the GSBP and SMBP formalisms.

2.3.1. Combination with the GSBP. Since GSBP is designed for MD simulations, the extended Lagrangian approach appears as the method of choice to treat the equations of motion. In this approach, the positions of the Drude particles are not relaxed to their energy minima at every step, and hence there is no need to apply an iterative method that would converge both the induced explicit polarization and the boundary potential simultaneously. Therefore, the DOs can be handled in the GSBP treatment just like fixed classical point charges. As usual, their contribution can be separated into inner−inner and

inner−outer terms. To account for the inner−inner terms, additional MM point charges (Q_m^{pol}) are introduced into the formalism and are projected on the same basis set as the other charges. The expression for ΔW_{elec}^{GSBP}, eq 17, is thus extended by adding the following terms.

$$\frac{1}{2}\sum_{mn} Q_m^{pol} M_{mn} Q_n^{pol} + \sum_{mn} Q_m^{pol} M_{mn} Q_n^{MM} + \sum_{mn} Q_m^{pol} M_{mn} Q_n^{QM} \quad (21)$$

The treatment of the inner-outer contribution depends of the description of the outer region. Two assumptions can be made. The first one is to describe the entire outer region by the PDC model, with different dielectric constants for the macromolecular part and the bulk solvent. In this case, the system is fully polarizable, and there is no need to include the DO model in the outer region or to make any further modification to the GSBP expression. This option differs, of course, from the standard GSBP implementation in QM/MM methods (see section 2.2). The second option is to assume that the polarization of the MM atoms of the macromolecule in the outer region remains the same during the MD run. This is clearly compatible with the basic GSBP assumption of neglecting the outer-region thermal fluctuations and hence keeping the outer-region MM atoms fixed during the MD simulation. The outer-region DO positions are thus determined by an initial single-point computation on the full system and are then kept fixed. The error arising from having a constant outer-region polarization is expected to be small. The second option allows us to solve the PB equation for the macromolecule with a dielectric constant of 1 and to use the previously introduced approximations that lead to an appreciable gain in efficiency.[46] In practice, the second option is implemented by adding Drude particles during the computation of ϕ_s^{outer} in the same manner as the outer-region MM point charges and by correcting the latter for the polarizable atoms by the DO counter charges.

To propagate the dynamical degrees of freedom, the force contributions from the GSBP have to be included in eqs 9.a and 9.b. Before solving the finite-difference PB equation, the charges describing the electrostatics of the system are projected onto a grid using B-splines. Thus, the force acting on any MM point charge will depend on its position on the grid. It is obtained by taking the first derivative of ΔW_{elec}^{GSBP}.

$$\frac{\partial \Delta W_{elec}^{GSBP}}{\partial r_i} = \frac{\partial \phi_s^{outer}}{\partial r_i} q_i + q_i \sum_{mn} \left[\frac{\partial b_n(r_i)}{\partial r_i}\right] M_{mn} \times [Q_{mn}^{QM} + Q_{mn}^{MM} + Q_{mn}^{pol}] \quad (22)$$

2.3.2. Combination with the SMBP. Inclusion of Drude oscillators into the SMBP formalism gives rise to additional contributions ($\Delta W_{elec,DO}^{SMBP}$) to the solvation free energy, accounting for the charge density $\rho_{DO}(r)$ that experiences the field $\phi_{tot}^{DO}(r)$.

$$\Delta W_{elec,DO}^{SMBP} = \int \mathrm{d}r\, \rho_{DO}(r)\, \phi_{tot}^{DO}(r) \quad (23.a)$$

In addition, the reaction field $\phi_{rf}^{DO}(r)$ will contribute to the boundary potential

$$\phi_{tot}^{DO}(r) = \phi_s^{outer}(r) + \frac{1}{2}\phi_{rf}^{DO}(r) + \phi_{rf}^{inner-MM}(r) \quad (23.b)$$

Likewise, eqs 20.a and 20.b are modified by adding $\phi_{rf}^{DO}(r)$ to take into account the DO contributions. The DO positions are updated by including the electric field contribution from the SMBP on the Drude particles in the iterative procedure. The electric field is determined at these positions by interpolation from the nearest point on the grid depending of the order of the B-spline used in the representation. The proper polarization of the inner region is obtained by solving eq 2 for every polarizable center in the field of the whole system.

$$\varphi_{tot}^{DO} = \varphi_{MM}^{DO} + \varphi_{QM}^{DO} + \varphi_{DO}^{DO} + \varphi_{SMBP}^{DO} \quad (24.a)$$

$$\varphi_{SMBP}^{DO}(r) = \phi_s^{outer}(r) + \phi_{rf}^{inner-MM}(r) + \phi_{rf}^{QM}(r) + \phi_{rf}^{DO}(r) \quad (24.b)$$

The treatment must take into account the interdependences between the different subsystems, since the QM wave function, the PFF, and the boundary potential are all polarizable and depend on each other. A sequential combination of the dual self-consistent-field procedures used to update DO positions and to determine the SMBP would lead to a very expensive approach, and therefore another technique needs to be considered.

For a QM/MM computation with a QM region of significant size modeled by an accurate method, the QM calculation is the bottleneck in terms of computational time. Therefore, in any iterative process, the number of QM calculations has to be kept as small as possible. We thus propose an iterative scheme that performs only one QM calculation for updating both the boundary potential and the DO polarization, as shown in Figure 3. Before starting the iterative procedure, the constant contribution to the boundary potential from the nonpolarizable outer-region MM part is computed. In the first step, the reaction field potential $\phi_{rf}^{QM}(r)$ is evaluated, and the total potential acting on the QM region is projected on the virtual charges. Thereafter, a QM computation is carried out to evaluate the wave function in the field of the inner MM region, the DO charges, and the virtual charges representing the SMBP. A new set of ESP charges is determined to represent the QM region in the SMBP, and the QM electric field is computed at the position of the Drude particles. Using the ESP charges, ϕ_{rf}^{QM} is computed by solving the PB equation, and the electric field arising from the boundary potential is evaluated at the DO positions. With all external contributions to the DO electric field being known at this point, an SCF procedure is performed to update the DO positions in the field of each other and of the environment. These new DO positions are then used to compute their contribution to the SMBP. The convergence criterion of this iterative procedure is the change in the QM energy from one step to another, which has been found to be the quantity that converges most slowly. The default criterion is the same as that for the QM computation itself (typically on the order of 10^{-7} atomic units). For the inner MM cycle, we have adopted the same convergence criteria as before (see section 3).

The procedure outlined above allows full convergence of the different parts. Compared with full QM/MM-DO calculations without boundary potentials, it is efficient because the inner SCF cycles for optimizing the DO positions are now restricted to a small number of inner-region DOs, and the QM computations need to include only rather few external MM point charges. Also, the overall process normally requires less QM calculations than when the full system is represented explicitly.

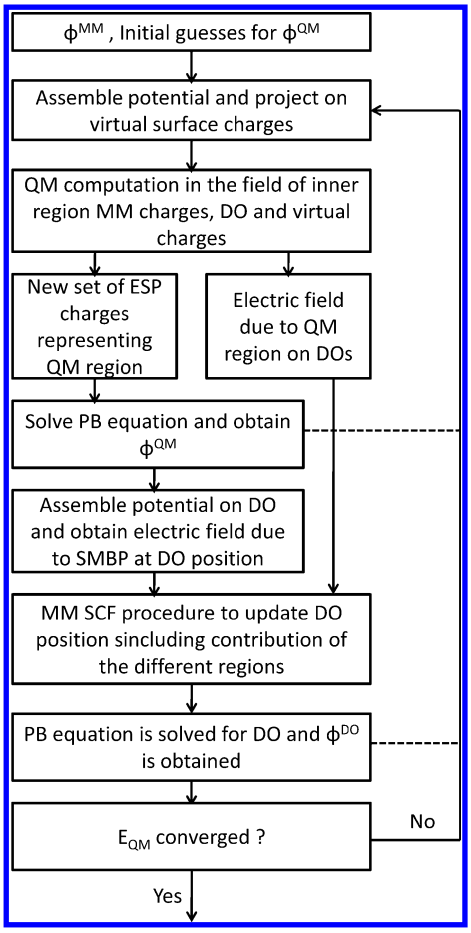

Figure 3. SCF procedure used for the update of DO positions, QM wave function, and SMBP contribution. The MM SCF procedure is the same as before (Figure 1). See text for details.

3. IMPLEMENTATION DETAILS

The methods described in section 2 have been implemented in the modular package ChemShell.[62] The code for Drude oscillators was implemented in a stand-alone module independent of the program used for MM force field evaluation. It is compatible with the CHARMM DO and GROMOS COS formalisms and can be used with any interfaced QM code. The code for lone pairs is available in a separate module and can be used together with any MM force field (polarizable or not).[23] The interaction energies involving lone pairs are computed first, and the associated gradient is assigned to the atom carrying the lone pair and its two nearest neighbors in a manner that conserves its total value and the torque (without generating any additional degrees of freedom). The propagation of cold Drude oscillators in an extended Lagrangian scheme has been implemented into the ChemShell MD module following the previously described implementation,[20] and the settle algorithm[71] was included for constraining water molecules.

The previously implemented PB equation solver[46] has been modified to handle the DO model. We have kept the approximations introduced to increase its efficiency as well as the rigid partitioning between inner and outer solvent molecules (i.e., not allowing for a dynamical and flexible separation). We use a spherical boundary both for GSBP and SMBP. When including higher-order terms such as the polarizability in the MM description, it is common practice to project the charges of the system on a discrete grid using higher-order functions.[57] For our models, fourth-order instead of third-order B-splines did not offer any improvement in terms of accuracy, and we thus adopted the latter. Similarly, the description of the dielectric boundary[57] did not improve significantly when using up to seventh-order polynomial switching functions instead of a direct approach, so we kept the latter.

In the current tests (see below), the QM calculations were performed using the MNDO2004[72] and Turbomole 6.3 programs.[73] DL_POLY was used for the additive part of the MM computations with the CHARMM force field. The HDCLOpt optimizer was employed for geometry optimizations using hybrid delocalized internal coordinates.[74]

4. ASSESSMENT

The present assessment consists of two parts. First, the accuracy of the model is discussed in terms of its ability to reproduce the gradient and the proper polarization of the inner region both for GSBP and SMBP. Second, the efficiency is evaluated by comparing computation times with and without a boundary potential.

4.1. Accuracy. Our standard test system for accuracy checks was a glycine molecule in its zwitterionic form solvated in a water ball of radius 30 Å, which has already been used in one of our previous studies.[49] Its high flexibility and polarity make it a challenging test case. The system was first thermalized using the standard nonpolarizable CHARMM force field and TIP3P water molecules. Five independent snapshots from an equilibrated MD run were investigated. These configurations were used in the QM/MM calculations without any further QM/MM-based equilibration. Since there was no significant difference between the five sets of QM/MM test results obtained, we present data only for one of the snapshots.

The glycine molecule is the QM part of the system, and the water ball is centered on its C_α carbon atom. The MM region includes 4252 water molecules, bringing the total number of atoms to 12 766. The 903 water molecules with the oxygen atom less than 18 Å away from the central carbon atom were considered as part of the inner region (together with the QM region). The 473 water molecules with their oxygen atom located in a buffer region between 14 and 18 Å from the center were frozen and represented explicitly. The other inner-region molecules were free to move. The AM1[75] semiempirical QM Hamiltonian was used for glycine, and the water molecules were represented using the SWM4-NDP PFF.[76] The relevant parameters are given in Figure 4. All inner-region DOs were considered to be active, even those assigned to a frozen explicit atom, to retain the full polarizability of the inner region.

For direct comparisons between the full Coulomb electrostatic interactions and the SMBP model, the outer-region dielectric constant was set to 1 (vacuum). Drude oscillators were included at outer-region atoms in fixed positions and were not allowed to reorganize themselves later on. Their positions were obtained by running an initial single-point computation on the full system without using a boundary potential. This way the accuracy of the SMBP could be evaluated by comparing QM/MM/SMBP results to those obtained for the entire QM/MM system. The PB equation was solved using a focusing procedure described previously[46]—first on a coarse-grained grid covering the full system and then with the use of a finer grid for the inner region, with the previously optimized spacings of 1.25 and 0.6 Å for the outer and inner grid,

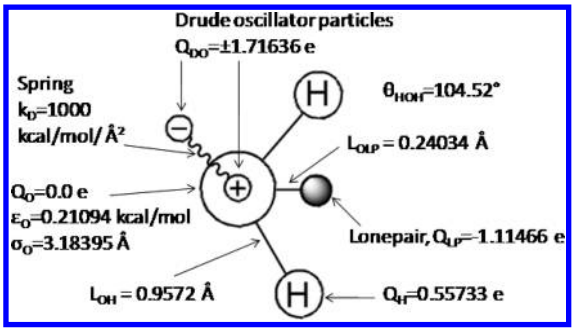

Figure 4. Schematic representation and parameters of the SWM4-NDP water model. The Drude oscillator is represented by a negative and a positive charge linked by a spring. The lone pair is shown in gray. Note that the oxygen atom has no charge by itself and is the only entity with van der Waals parameters. Geometric parameters (except for the oxygen-lone pair distance) are the same as for the TIP3P water model.

respectively. To project the SMBP potential acting on the QM region, 89 virtual surface charges were employed. Note that different definitions of the boundary potential will be used in the following tests.

Figure 5 shows (a) the mean absolute deviation (MAD) of gradient components (x,y,z) for all atoms located in the inner

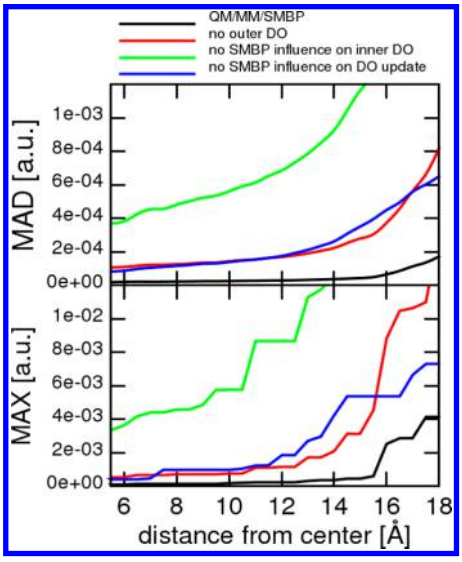

Figure 5. Deviations of the gradient components on atoms located in the inner region with a radius of 18 Å, compared with the full QM/MM results. The top panel shows the mean absolute deviation (MAD) as a function of the distance from the central carbon atom of glycine. The corresponding maximum absolute deviation (MAX) is plotted in the bottom panel. Different approximations are examined (see text).

region and (b) their maximum absolute deviation (MAX). The data are plotted as a function of the distance to the central carbon atom of glycine. The black curves correspond to the QM/MM/SMBP method as discussed in section 2. For QM/MM geometry optimizations, the convergence criteria used on gradients are typically on the order of 5×10^{-4} au. Therefore, the results obtained for this test case show that the SMBP gives a good approximation of the electrostatic interaction with the outer region. Indeed, for an active region that encompasses any molecule within 14 Å of the center, the maximum absolute deviation from QM/MM results is less than the standard convergence criterion, and the mean absolute deviation is even 1 order of magnitude smaller (i.e., only 3.6×10^{-5} au). In the buffer region between 14 and 18 Å, the gradients deviate more and more from the standard QM/MM values, which supports the convention to keep this region frozen in SMBP studies.[49] The gradients in the QM region show almost no difference, and localized processes in this region should thus be well simulated. The results in Figure 5 do not differ significantly from previously reported SMBP results obtained with identical parameters and the standard fixed-charge CHARMM force field.[49] Since the accuracy of the calculated QM/MM/SMBP gradients is essentially the same with and without DO contributions, it is not governed by the DO treatment but rather by the intrinsic errors arising from the finite-difference solution of the PB equation.[49]

How important are the DO contributions to the gradient? Do we need the full doubly iterative SCF procedure to obtain proper gradients? We address these questions by separately considering the influence of the Drude oscillators in the outer region (frozen) and in the inner region (active). We assess the necessity of including DO in the SMBP expression by considering three cases. The red curves in Figure 5 show the effects of removing the DO contributions from the outer region, which are not represented by the boundary potential. Note that for every DO, the $-q$ charge located at the atomic position is removed as well. Compared to QM/MM/SMBP, the deviations (MAD and MAX) from the full QM/MM reference gradients not only are significantly higher but also increase faster as the vicinity of the boundary is approached. This implies that the DO contributions from the outer region may have a significant influence even on localized inner-region processes, which is captured by our QM/MM/SMBP approach. The green curves show the effects of neglecting the influence of the SMBP terms on the inner-sphere Drude oscillators, both with regard to the update of the DO positions during the SCF process and their final contribution to the gradient. The deviations (MAD and MAX) are prohibitively large, so that this approximation is not to be used. The blue curves in Figure 5 show the deviations (MAD and MAX) that arise when neglecting the electric field from the SMBP during the SCF process of updating the DO positions, while including it during the final gradient computation. These deviations are smaller than those obtained upon total neglect of SMBP effects (green curves), but they are still too large to be tolerated. We conclude from these computational experiments that the full SCF procedure should be applied to ensure the needed accuracy.

We now investigate the influence of long-range electrostatic interactions on the polarization of the Drude oscillators, as indicated by the self-energy of polarization (i.e., the last term of eq 4) which is proportional to the DO polarization. Figure 6 shows the variation of this self-energy with the dielectric constant of the PDC that describes the outer-region solvent in the SMBP. The total polarization of the 903 inner-region Drude oscillators quickly increases for dielectric constants up to 20 and then levels off, with the self-energy slowly converging to a value of about 0.012 au. This corresponds for each Drude oscillator to a maximum variation of d by ca. 10^{-3} Å and a maximum increase of DO polarization by ca. 10−20%. The convergence of DO polarization in the absence of any damping terms suggests that the present SMBP treatment does not lead to the so-called "polarization catastrophe" for high dielectric constants, so that no special measures need to be taken in this regard.

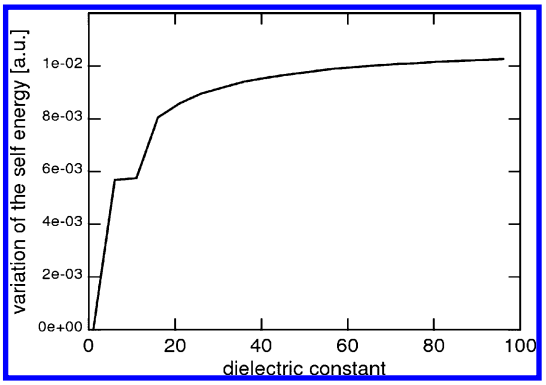

Figure 6. Total self-energy of polarization of the 903 Drude oscillators located in the inner region of the system, as a function of the dielectric constant used in the SMBP to represent the outer solvent through a PDC.

We next check in more detail how the self-energy of polarization evolves with respect to the distance from the center of the sphere (i.e., the central carbon atom of glycine). For this purpose, we again use a reference system consisting of a 30 Å sphere of water including the zwitterionic glycine molecule (12 766 atoms in total), and we represent the bulk solvent beyond this sphere in the SMBP by a PDC with a dielectric constant of 80. We compare the results obtained for this reference system with those computed for our standard system, i.e., an 18 Å sphere of water including glycine (2719 atoms in total) with bulk solvent treated as a PDC with different dielectric constants. To assess the accuracy of the results for the truncated system compared with the reference system, we consider the mean average percentage of deviation (MAD) and maximum percentage of deviation (MAX) of the self-energy of polarization of the Drude oscillators. These quantities are plotted in Figure 7 for several choices of the dielectric constant ε in the truncated system. For $\varepsilon = 80$, the MAD value is close to zero in the inner part of the active region and remains below 5% throughout the active region (for distances R up to 14 Å), while the MAX value rises to 23% at the active/frozen boundary (at $R = 14$ Å). This confirms again that the SMBP provides a reasonably accurate description of the inner active region of the truncated system. The situation is less favorable when neglecting the bulk solvent in the truncated system: for $\varepsilon = 1$ (vacuum), the MAD (MAX) value rises from ca. 12% (35%) in the inner active region to ca. 15% (80%) at the active/frozen boundary. When using truncated systems, the full SMBP approach (with a PDC treatment of bulk solvent) thus provides a clear improvement compared with a treatment that neglects the bulk solvent, as far as the polarization of the MM region is concerned. We also note that the distant-dependent self-energies converge very quickly with increasing dielectric constant (being very similar for $\varepsilon = 15$ and $\varepsilon = 80$, see Figure 7), in analogy to the fast convergence of the total self-energy (Figure 6).

We have evaluated the ability of the GSBP to properly reproduce the inner−outer electrostatic interactions in a similar way as for the SMBP. Initial positions for the Drude particles and the lone pairs were obtained from a fully converged iterative QM/MM computation on the complete system. The gradients on the Drude particles necessary for proper propagation of the dynamics were evaluated both at the full QM/MM and the QM/MM/GSBP level. Using the same computational parameters as in the SMBP case, the gradients on the mobile inner-region Drude particles (within the 14 Å active region) show a maximum absolute deviation of 2.7×10^{-4} au and a mean absolute deviation of 7.3×10^{-5} au from the full QM/MM reference data. These deviations are slightly higher than those obtained for the SMBP but in the same range as in previous GSBP validations,[46] in which this accuracy has been considered good enough for molecular dynamics simulations.

Going from a fixed-charge to a polarizable force field in the QM/MM/GSBP treatment may be expected to make the representation of the electrostatic interactions more demanding. To compute the inner region−inner region interactions, the GSBP approach employs a projection of the associated Green's function onto a basis set, which is also used to represent the corresponding multipole moments. We checked the convergence of the GSBP energy with increasing basis set size for the same test system as before. Figure 8 shows the

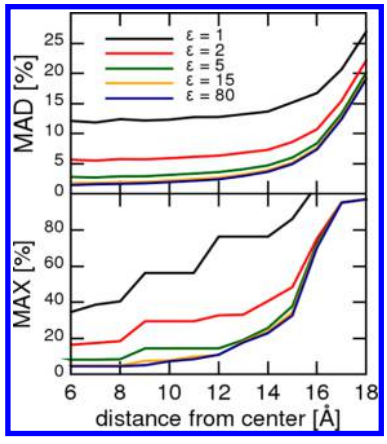

Figure 7. Mean average percentage of deviation (MAD) and maximum percentage of deviation (MAX) of the self-energy of polarization of Drude oscillators (truncated vs reference system, see text) plotted as a function of the distance from the central carbon atom of glycine. In the truncated system, different dielectric constants are used in the SMBP to represent the outer solvent through a PDC.

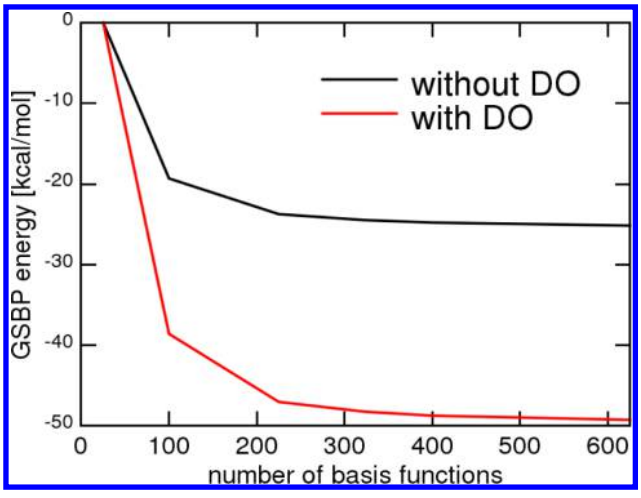

Figure 8. Variation of the GSBP energy with the size of the basis set used to project the inner−inner potential (relative to the value obtained with 25 basis functions). The dielectric constant of the outer region was fixed to 80.

dx.doi.org/10.1021/ct300722e | J. Chem. Theory Comput. 2012, 8, 4527−4538

variation of the GSBP energy for the two limiting cases, with the Drude oscillators being fully included or fully neglected in both the inner and the outer region. In the former case (red curve), the GSBP energy converges not quite as fast as in the latter case (black curve), indicating the need for a larger basis when including polarization. However, the previously recommended expansion up to l = 20 (400 functions)[46] is seen to capture the main part of the GSBP energy: an extension to l = 25 provides an additional energy gain of 0.52 kcal/mol (with DO) vs 0.37 kcal/mol (without DO). Hence, only a rather slight increase in the order of the expansion is needed in the DO case to ensure the same accuracy as before.

4.2. Efficiency. When using boundary potentials, explicit atomistic simulations are restricted to the inner region. For systems that are big enough, one may thus expect an appreciable reduction of the computational effort compared with a full QM/MM calculation.

To evaluate the efficiency of the SMBP, we use our test system (glycine in water). It consists of 10 QM atoms and 21 260 point charges in the MM region (taking into account lone pairs and Drude oscillators). The SMBP treatment employed the previously adopted parameters[49] and an 18 Å inner region containing the 10 QM atoms and a total of 4515 MM point charges. Table 1 lists the computation times and

Table 1. Average Time (s) for One Geometry Optimization Step at the Full QM/MM Level and the QM/MM/SMBP Level with the Associated Savings[a]

QM	basis	QM/MM	QM/MM/SMBP	% saved
AM1		122	66	46
BLYP	SVP	505	262	48
BLYP	TZVPP	1355	726	47
B3LYP	SVP	613	362	41
B3LYP	TZVPP	1622	1021	33

[a]Computations were run on 2.93 GHz Intel Xeon X5670 machines with 12 GB of memory. See text for further details.

associated savings obtained on average for one geometry optimization step for an active region encompassing glycine and all water molecules with their oxygen atom within 14 Å of the center. For the sake of consistency, interactions within the fixed outer region were neglected in both cases. The convergence criteria for the QM energy were 10^{-7} eV for the AM1[75] Hamiltonian and 10^{-7} au for the DFT computations using the BLYP[77,78] and B3LYP[79] exchange-correlation functionals with the SVP[80] and TZVPP[81] basis sets. The DO positions were considered converged if the maximum absolute deviation from one step to another was below 10^{-5} Bohr and the mean absolute deviation was below 2×10^{-5} Bohr. Computations were run on 2.93 GHz Intel Xeon X5670 machines with 12 GB of memory. Averages were taken over the 100 last steps of the geometry optimization (110 steps overall). The computation time for the optimizer was included and assumed to be the same in both cases. We obtain appreciable savings ranging from 33% to 48% depending on the chosen QM method. These savings arise from different contributions. In both cases, the use of the SMBP causes a strong reduction of the number of explicitly treated point charges that is even more pronounced for the polarizable force field (DO-PFF), which represents each water molecule by five point charges (rather than three without DO). We also observe that the implicit representation of the outer region in the SMBP treatment leads to faster convergence both in the overall and the inner SCF procedure.

To evaluate the efficiency of the GSBP, the glycine molecule was solvated in several water balls of different sizes. Computation times were determined for MD simulations using the extended Lagrangian approach. To ensure proper propagation, we used multiple time steps, i.e., 1 fs for the atomic motions and 1/30 fs for the thermostat.[20] An artificial mass of 0.4 au was assigned to each DO. Dynamics were run for 100 steps at 300 K to determine average times. The definitions of the active and inner regions were maintained for every system size. Figure 9 compares the average time per MD step

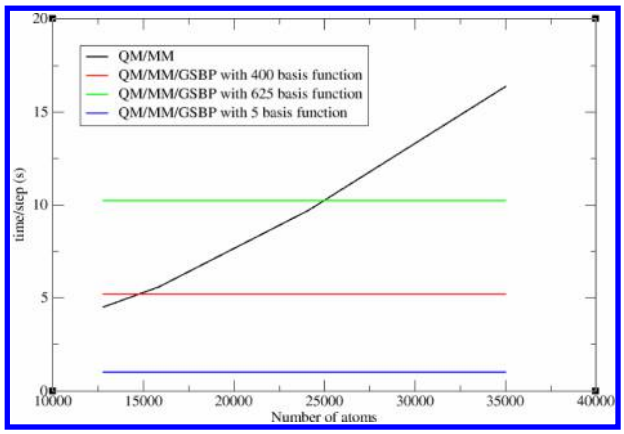

Figure 9. Computation time per MD step as a function of the total number of atoms. The black line shows the times for the full QM/MM treatment. In color: QM/MM/GSBP times for different numbers of basis functions used to project the inner−inner potential.

for QM/MM and QM/MM/GSBP MD simulations, as a function of the total number of atoms (not including the virtual DO and lone pair charges). The black line indicates the linear increase of the QM/MM computation time with system size. The colored lines show the QM/MM/GSBP computation times, which are essentially independent of system size, but depend strongly on the size of the basis set used for the required projections. The crossing points specify the system size, beyond which the QM/MM/GSBP treatment becomes more efficient than the standard QM/MM approach. For the previously recommended basis set (l = 20, 400 basis functions), the crossing occurs at a system size of about 14 500 atoms when using a polarized force field (DO-PFF), compared with about 12 500 atoms for fixed-charge force fields.[46] In QM/MM studies of enzymes, the system size is often in the range of 20 000−40 000 atoms so that the use of the GSBP provides significant savings in both cases.

5. CONCLUSION

In this article, we have combined two types of boundary potentials (SMBP and GSBP) with a QM/MM treatment, in which the MM part is described by a polarizable force field formulated in terms of Drude oscillators. For both boundary potentials, this leads to a fully polarizable three-layer QM/MM/BP model. The boundary potentials account for long-range electrostatic interactions, since they simulate the outer-region solvent by a polarizable dielectric continuum. In the case of the GSBP, the effects of the outer region are represented by a reaction field matrix, which is obtained once and for all at the beginning of a MD simulation. In the case of the SMBP, they

are computed on-the-fly during geometry optimization, in a manner that allows the use of any kind of QM method. The Drude oscillators simulate the polarization of the explicit inner MM region by two point charges of opposite sign linked by a spring, thus forming a dipole. One of the two DO charges is located at the polarizable atom, while the position of the other one is optimized in the field of the environment, with contributions from the QM and MM atoms, the boundary potential, and the other Drude oscillators. Likewise, DO terms are included in the evaluation of the QM wave function and contribute to the boundary potential arising from the inner region. Hence, the polarization effects in the three layers of our model are interdependent and coupled with each other. In the GSBP case, this coupling was treated by using an extended Lagrangian scheme, which propagates the DO positions as dynamical degrees of freedom so that they can be considered as fixed at every MD step, thus allowing us to handle the DO charges as any other classical MM point charge. In the SMBP case, the polarizations in the three layers were fully converged through an SCF procedure designed to minimize the number of QM evaluations to increase efficiency.

The accuracy of the two combination schemes was checked for a test system consisting of a glycine molecule in a water ball. Both schemes reproduce the gradients from corresponding full QM/MM calculations quite well. A proper SMBP representation of the inner−outer potential energy term requires the inclusion of frozen outer-region Drude oscillators. On the other hand, it also seems advisable to include the DO contributions during the SCF procedure to determine the SMBP. The influence of long-range electrostatic interactions on the Drude model is found to be significant, but there is no evidence for a "polarization catastrophe" when using a high dielectric constant to describe the outer-region solvent. The effects of the SMBP on the DO positions typically amount to less than 10^{-3} Å and thus to less than 10% of typical polarization effects. In the GSBP case, the results depend on the size of the basis set used to project the potential. When using Drude oscillators, maintaining the desired accuracy may require a slight extension of the multipole expansions used to describe the inner−inner interactions. QM/MM/SMBP computations become faster than the corresponding QM/MM computations beyond a certain system size, which depends on the chosen QM method. In the GSBP case, the efficiency also depends on the size of the basis set used for projection. In both cases, appreciable savings can be realized for large systems.

The presented three-layer models are particularly suitable for accurate ab initio free energy calculations of localized processes in macromolecules. Such studies require suitable polarizable force fields for proteins and other biomolecules.

AUTHOR INFORMATION

Corresponding Author

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

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are grateful to Tobias Benighaus and Yan Zhang for supporting work and helpful discussions.

REFERENCES

(1) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198.
(2) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H. *J. Phys. Chem. B* **2006**, *110*, 6458.
(3) Hu, H.; Yang, W. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573.
(4) Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, *56*, 389.
(5) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186.
(6) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185.
(7) Warshel, A. *Annu. Rev. Biophys.* **2003**, *32*, 425.
(8) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.
(9) Warshel, A.; Kato, M.; Pisliakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034.
(10) Lopes, P. E. M.; Roux, B.; MacKerell, A. D. *Theor. Chem. Acc.* **2009**, *124*, 11.
(11) Yu, H.; van Gunsteren, W. F. *Comput. Phys. Commun.* **2005**, *172*, 69.
(12) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515.
(13) Ren, P.; Ponder, J. W. *J. Phys. Chem. B* **2003**, *107*, 5933.
(14) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2004**, *108*, 621.
(15) Rappe, A. K.; Goddard, W. A., III. *J. Phys. Chem.* **1991**, *95*, 3358.
(16) Rick, S. W.; Stuart, S. J.; Bader, J. S.; Berne, B. *J. Mol. Liq.* **1995**, *65*, 31.
(17) Stuart, S. J.; Berne, B. *J. Phys. Chem.* **1996**, *100*, 11934.
(18) Patel, S.; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 1.
(19) Patel, S.; MacKerell, A. D., Jr.; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 1504.
(20) Lamoureux, G.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 3025.
(21) Lamoureux, G.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185.
(22) Vorobyov, I.; Anisimov, V. M.; Greene, S.; Venable, R. M.; Moser, A.; Pastor, R. W.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2007**, *3*, 1120.
(23) Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D., Jr; Schulten, K.; Roux, B. *J. Phys. Chem. Lett.* **2011**, *2*, 87.
(24) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D., Jr.; Roux, B. *J. Chem. Theory Comput.* **2010**, *6*, 774.
(25) Lopes, P. E. M.; Zhu, X.; Lau, A.; Roux, B.; MacKerell, A. D., Jr. *Biophys. J.* **2011**, *100*, 612.
(26) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2005**, *1*, 153.
(27) Nüsslein, V.; Schröder, U. *Phys. Status Solidi B* **1967**, *21*, 309.
(28) Schröder, U. *Solid State Commun.* **1993**, *88*, 1049.
(29) de Leeuw, N. H.; Parker, S. C. *Phys. Rev. B* **1998**, *58*, 13901.
(30) Yu, H. B.; Hansson, T.; van Gunsteren, W. *J. Chem. Phys.* **2003**, *118*, 221.
(31) Straatsma, T. P.; McCammon, J. A. *Mol. Simul.* **1990**, *5*, 181.
(32) Geerke, D. P.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2007**, *3*, 2128.
(33) Geerke, D. P.; van Gunsteren, W. F. *J. Phys. Chem. B* **2007**, *111*, 6425.
(34) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580.
(35) Baker, C. M.; Anisimov, V. M.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2011**, *115*, 580.
(36) Vosmeer, C. R.; Rustenburg, A. S.; Rice, J. E.; Horn, H. W.; Swope, W. C.; Geerke, D. P. *J. Chem. Theory Comput.* **2012**, *8*, 3839.
(37) Luo, Y.; Jiang, W.; Yu, H.; MacKerell, A. D., Jr.; Roux, B. *Faraday Discuss.* **2012**, DOI: 10.1039/C2FD20068F. http://pubs.rsc.org/en/content/articlepdf/2012/fd/c2fd20068f (accessed Sept. 5, **2012**).
(38) Kunz, A.-P. E.; Allison, J. R.; Geerke, D. P.; Horta, B. A. C.; Hünenberger, P. H.; Riniker, S.; Schmid, N.; van Gunsteren, W. F. *J. Comput. Chem.* **2012**, *33*, 340.
(39) Geerke, D. P.; Thiel, S.; Thiel, W.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2007**, *3*, 1499.
(40) Lu, Z.; Zhang, Y. *J. Chem. Theory Comput.* **2008**, *4*, 1237.

(41) Rowley, C. N.; Roux, B. *J. Chem. Theory Comput.* **2012**, *8*, 3526.
(42) Meier, K.; Thiel, W.; van Gunsteren, W. F. *J. Comput. Chem.* **2011**.
(43) Nam, K.; Gao, J.; Darrin, M. *J. Chem. Theory Comput.* **2005**, *1*, 2.
(44) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2006**, *2*, 1370.
(45) Gao, J.; Alhambra, C. *J. Chem. Phys.* **1997**, *107*, 1212.
(46) Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2008**, *4*, 1600.
(47) Im, W.; Berneche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924.
(48) Schaefer, P.; Riccardi, D.; Cui, Q. *J. Chem. Phys.* **2005**, *123*, 014905.
(49) Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2009**, *5*, 3114.
(50) Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2011**, *7*, 238−249.
(51) Gilson, M. K.; Honig, B. H. *Biopolymers* **1986**, *25*, 2097.
(52) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483.
(53) Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. *J. Chem. Theory Comput.* **2006**, *2*, 452.
(54) Jon, R.; Cao, Y.; Damm, W.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A. *J. Chem. Theory Comput.* **2005**, *1*, 694.
(55) Li, H.; Gordon, M. S. *J. Chem. Phys.* **2007**, *126*, 124112.
(56) Steindal, A. H.; Ruud, K.; Frediani, L.; Aidas, K.; Kongsted, J. *J. Phys. Chem. B* **2011**, *115*, 3027−3037.
(57) Schnieders, M. J.; Baker, N. A.; Ren, P.; Ponder, J. W. *J. Chem. Phys.* **2007**, *126*, 124114.
(58) Lipparini, F.; Barone, V. *J. Chem. Theory Comput.* **2011**, *7*, 3711.
(59) Schwabe, T.; Olsen, J. M.; Sneskov, K.; Kongsted, J.; Christiansen, O. *J. Chem. Theory Comput.* **2012**, *7*, 2209.
(60) Mennucci, B. Personal communication, July 2012.
(61) ChemShell. www.chemshell.org (accessed Aug 14, 2012).
(62) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1.
(63) Zhu, X.; Lopes, P. E. M.; MacKerell, A. D., Jr. *WIREs Comput. Mol. Sci.* **2012**, *2*, 167.
(64) Thole, B. T. *Chem. Phys.* **1981**, *59*, 341.
(65) Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D., Jr; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1587.
(66) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Mol. Phys.* **1996**, *87*, 1117.
(67) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637.
(68) Kolafa, J. *J. Comput. Chem.* **2004**, *25*, 335.
(69) Kolafa, J. *J. Chem. Phys.* **2005**, *122*, 164105.
(70) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050.
(71) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952.
(72) Thiel, W. *MNDO program*; Max-Planck-Institut für Kohlenforschung: Mülheim an der Ruhr, Germany, 2004.
(73) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
(74) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177.
(75) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
(76) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., Jr. *Chem. Phys. Lett.* **2006**, *418*, 245.
(77) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(78) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(79) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(80) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.
(81) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.

NOTE ADDED AFTER ASAP PUBLICATION

This article was published ASAP on October 24, 2012. The second line in equation 4 has been modified. The correct version was published on October 30, 2012.