Generalized solvent boundary potential for computer simulations

Wonpil Im, Simon Bernèche, and Benoît Roux[a]
Department of Biochemistry, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10021

A general approach has been developed to allow accurate simulations of a small region part of a large macromolecular system while incorporating the influence of the remaining distant atoms with an effective boundary potential. The method is called the Generalized Solvent Boundary Potential (GSBP). By representing the surrounding solvent as a continuum dielectric, both the solvent-shielded static field from the distant atoms of the macromolecule and the reaction field from the dielectric solvent acting on the atoms in the region of interest are included. The static field is calculated once, using the finite-difference Poisson–Boltzmann (PB) equation, and the result is stored on a discrete grid for efficient simulations. The solvent reaction field is developed using a basis-set expansion whose coefficients correspond to generalized electrostatic multipoles. A matrix representing the reaction field Green's function between those generalized multipoles is calculated only once using the PB equation and stored for efficient simulations. In the present work, the formalism is applied to both spherical and orthorhombic simulation regions for which orthonormal basis-sets exist based on spherical harmonics or cartesian Legendre polynomials. The GSBP method is also tested and illustrated with simple model systems and two detailed atomic systems: the active site region of aspartyl-tRNA synthetase (spherical region) and the interior of the KcsA potassium channel (orthorhombic region). Comparison with numerical finite-difference PB calculations shows that GSBP can accurately describe all long-range electrostatic interactions and remain computationally inexpensive. © *2001 American Institute of Physics.* [DOI: 10.1063/1.1336570]

I. INTRODUCTION

In studies of biomolecular systems, one is frequently interested in simulating a microscopic process taking place in a small localized region of a large macromolecule (e.g., enzymatic reactions, ligand binding, sidechain ionization, and ion permeation through membrane channels). Even though a simulation based on atomic models in which the entire macromolecular system is treated with explicit solvent certainly provides the most detailed approach to address such problems, the computational cost grows rapidly with the system size. Furthermore, practical difficulties are encountered in free energy calculations involving changes in the total charge when the long-range electrostatic interactions are truncated, or treated through Ewald summations.[1–4] Therefore, it is important to devise alternative approaches to circumvent these difficulties. One particularly attractive approximation consists in considering a reduced subset of the macromolecular system in the region of interest while incorporating the influence of the remaining atoms with an effective "boundary potential."[5–17] The general idea is illustrated schematically in Fig. 1.

Fundamentally, the proper effective boundary potential from the surrounding should be designed such that computer simulations of the finite subsystem yields statistical properties that are identical to those that would be obtained if the corresponding infinite system was simulated. This concept can be formulated rigorously.[16,17] Separating the multidimensional solute–solvent configurational integral in terms of "inner" solvent molecules nearest to an arbitrary solute, and the remaining "outer" bulk solvent molecules, it can be shown that the solvent boundary potential corresponds to the solvation free energy of an effective cluster comprising the solute and inner explicit solvent molecules embedded in a large hard sphere. The hard sphere corresponds to a configurational restriction on the outer bulk solvent molecules; its radius is variable, such that it includes the most distant inner solvent molecule. An implementation based on this formulation, called the Spherical Solvent Boundary Potential (SSBP), has been designed to simulate a spherical system embedded in bulk solvent.[16] SSBP has been used to study various biomolecular systems, e.g., electron transfer in proteins,[18,19] solvation free energy of small ions[20] and amino acids,[21] thermal stability of Chymotrypsin Inhibitor 2[22] and Human Lysozyme,[23,24] folding kinetics of proteins[25] and polypeptides,[26] and various calcium-loaded states of Calbindin D9k.[27]

Although, the theoretical formulation that led to SSBP is rigorous and general, it has been possible to achieve important simplifications only by assuming that the simulation region is roughly spherical and embedded in a uniform solvent.[16] This enabled us to use Kirkwood's classical multipolar expansion to approximate the electrostatic reaction field free energy.[28] In the general case, however, it may be advantageous to simulate a small region that is part of a much larger macromolecule (see Fig. 1). Then, all electrostatic interactions are shielded in a complex way by the ir-
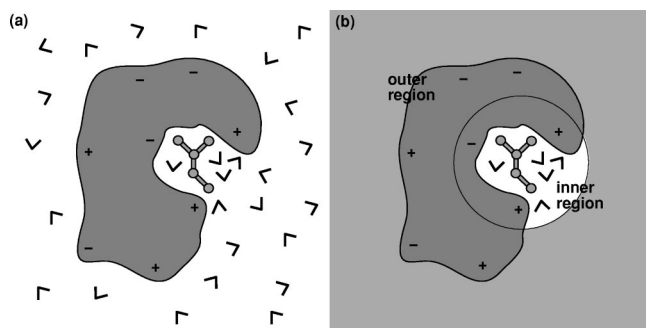
FIG. 1. Schematic representation of a large biomolecule system comprising a macromolecule, a ligand, and solvent molecules. (a) The entire system is shown with explicit solvent molecules. (b) The system is divided into an outer and an inner region. In the inner region, the ligand, the solvent molecules, and part of the macromolecule are included explicitly. In the outer region, the remaining atoms of the macromolecule are included explicitly while the solvent is represented as a continuum dielectric medium.

regular shape of the macromolecule–solvent dielectric interface outside the simulation region. The multipolar expansion, which is valid only for an arbitrary charge distribution inside a spherical cavity enclosed in a dielectric continuum, cannot be used in the general case.

In the early developments of solvent boundary potentials, such long-range electrostatic effects were neglected.[6,8,9] More recently, the problem has been revisited for calculating the free energy difference between Asp and Asn in the active site of aspartyl-tRNA synthetase.[29–31] In those applications, the effect of the solvent reaction field has been incorporated using a perturbative correction relative to vacuum,[29,30] or relative to a uniform dielectric medium,[31] and the influence of the static field from the protein atoms in the outer region was either included approximately using a distance-dependent charge scaling scheme[29,30] or ignored.[31] Although such schemes may be adequate in a specific case, there is still no general method to incorporate accurately the long-range electrostatic interactions for simulating a small region of a large solvated macromolecular system. The existing methodology needs to be extended.

The goal of the present work is to extend SSBP to macromolecular systems with arbitrary geometries. Practically, one needs to incorporate both the influence of the macromolecular charge distribution outside the simulation region as well as the complex electrostatic response from the irregular dielectric interfaces in a computationally efficient simulation method. In Sec. II, the background and theoretical developments are given in detail with the numerical implementation of the method. In Sec. III, the accuracy of the implementation is tested with a series of calculations with simple point-charge systems. In addition, calculations with aspartyl-tRNA synthetase,[29] and the KcsA potassium channel[32] are presented to test and illustrate the present developments. Molecular dynamics (MD) simulations are performed with GSBP to assess the performance of the method. The paper ends in Sec. IV with a brief summary and conclusion of the main results.

II. THEORETICAL DEVELOPMENTS

A. General considerations

We consider a macromolecule (protein or nucleic acid including any bound ligand) in a particular configuration, $\mathbf{R}$, immersed in a bulk liquid of N solvent molecules. The potential energy of the system is $U_{\text{tot}}(\mathbf{R},1,2,...,N)$. It is composed of internal energy terms from the valence force field (bonds, angles, and torsions) and nonbonded interactions (Lennard-Jones 6–12 potentials U_{LJ}, and electrostatic Coulombic interactions between partial charges, U_{elec}).[33]

It has been shown previously that the N solvent molecules can be separated in two groups: the n ''inner'' solvent molecules nearest to a region of interest, and the remaining ''outer'' $(N-n)$ solvent molecules.[16] This is represented schematically in Fig. 1. The separation of the solvent molecule in two groups makes it possible to integrate out the contribution of the outer solvent molecules such that their influence is taken into account implicitly.[16] The statistical properties of the subsystem (comprising the macromolecule and the n inner solvent molecules) can be expressed rigorously in terms of the Boltzmann average of the finite system interacting with the effective potential of mean force (PMF), $\mathcal{W}(\mathbf{R},1,...,\mathbf{n})$, defined from the restricted configurational integral,

$$e^{-\mathcal{W}(\mathbf{R},1,...,\mathbf{n})/k_B T}$$
$$= \frac{1}{C} \int d(\mathbf{n+1}) \int d(\mathbf{n+2})\cdots e^{-[U_{\text{tot}}+U_{\text{cr}}]/k_B T}, \quad (1)$$

where C is an arbitrary normalization constant, and U_{cr} is an effective hard repulsive potential acting only on the outer solvent molecules $n+1,n+2,...$, etc. The hard repulsive potential is introduced as a formal device to account for the configurational restriction (cr) on the outer solvent molecules (they must be prevented to enter in the spherical inner region to avoid multiple counting of their influence); it does not act directly on the macromolecule or the n inner solvent molecules. The radius of the restricted inner region is determined from the farthest inner solvent molecule. The radius is, thus, flexible and varies according to the instantaneous configuration of the inner solvent molecules.

According to the formulation based on Eq. (1), averages of observables as well as free energy differences involving only on the degrees of freedom of the reduced system can be expressed rigorously in terms of the PMF $\mathcal{W}$. Choosing the normalization constant as the unrestricted configurational integral of the outer solvent molecules allowed us to identify $\mathcal{W}$ as the solvation free energy of an effective system composed of one macromolecule and n solvent molecules, frozen in the $(\mathbf{R},1,...,\mathbf{n})$-configuration, and an associated hard repulsive inner region acting on the outer solvent molecules.[16] This interpretation of the influence of the outer solvent molecules in terms of a configuration-dependent solvation free energy is helpful in deriving approximations for boundary potential.

The configuration-dependent free energy due to the surrounding bulk solvent can be obtained by calculating, step by step, the reversible thermodynamical work necessary to as-

semble the effective system.[16] For the sake of notational clarity, let us summarize all the Cartesian coordinates of the atoms (macromolecules, ligand, and inner solvent molecule) as $\mathbf{X}$. The free energy is then,

$$\mathcal{W}(\mathbf{X}) = U(\mathbf{X}) + \Delta\mathcal{W}_{cr}(\mathbf{X}) + \Delta\mathcal{W}_{np}(\mathbf{X}) + \Delta\mathcal{W}_{elec}(\mathbf{X}), \quad (2)$$

where U is the microscopic potential energy of the subsystem, and $\Delta\mathcal{W}_{cr}$, $\Delta\mathcal{W}_{np}$, and $\Delta\mathcal{W}_{elec}$ represent the configurational restriction, the nonpolar, and the electrostatic free energy contributions, respectively.

To obtain a reduced boundary potential involving only the inner atoms it is necessary to integrate out all the remaining degrees of freedom corresponding to the macromolecule located in the outer region. The reduction in the number of degrees of freedom can be expressed as,

$$e^{-\overline{\mathcal{W}}(\mathbf{X}_i)/k_BT} = \frac{1}{C'} \int d\mathbf{X}_o e^{-\mathcal{W}(\mathbf{X}_i, \mathbf{X}_o)/k_BT}, \quad (3)$$

where C' is an arbitrary constant, and $\mathbf{X}_i$ and $\mathbf{X}_o$ represent the coordinates of the atoms in the inner and outer regions, respectively. Although one can formally write such a multidimensional integral, it cannot generally be resolved exactly. If the influence of atomic fluctuations in the outer region can be neglected, a reasonable approximation consists in assuming that the atoms in the outer region are fixed in a representative average configuration $\overline{\mathbf{X}}_o$, i.e., $\overline{\mathcal{W}}(\mathbf{X}_i) \approx \mathcal{W}(\mathbf{X}_i, \overline{\mathbf{X}}_o)$.[9,29–31] The key quantity of interest that needs to be characterized is thus the configuration-dependent free energy function $\mathcal{W}(\mathbf{X}_i, \mathbf{X}_o)$.

A number of approximations, based on different level of sophistication, can be used to estimate the various free energy contributions to $\mathcal{W}(\mathbf{X}_i, \mathbf{X}_o)$. It is generally desirable to use analytical functions to obtain a simple and computationally inexpensive approximation to the boundary potential. In particular, the nonpolar contribution gives rise to short-range forces which can be well described by a simple mean field approximation,[7,8,16] while the contribution to the electrostatic free energy can be approximated very well by representing the solvent in the outer region as a dielectric continuum.[16] In the implementation of SSBP,[16] Kirkwood's classical multipolar expansion[28] was used to represent the electrostatic reaction field from the surrounding solvent. This cannot work when part of the macromolecule lies outside the simulation region (see Fig. 1). In the general case, the average charge distribution of the macromolecule in the outer region gives rise to a static electrostatic potential acting on the charges in the inner region. This static field is shielded, in a nontrivial way, by the continuum dielectric surrounding the macromolecule. Complex reaction fields arise from the irregular macromolecule–solvent dielectric boundaries, effectively coupling the charges in the inner and outer regions to one another. These effects can be taken into account quantitatively by considering the electrostatic solvation free energy contribution $\Delta\mathcal{W}_{elec}(\mathbf{X})$ expressed as[34,35]

$$\Delta\mathcal{W}_{elec}(\mathbf{X}) = \frac{1}{2} \sum_\alpha q_\alpha \phi_{rf}(\mathbf{r}_\alpha), \quad (4)$$

where $\phi_{rf}(\mathbf{r}_\alpha)$ is the reaction field at the position of the atomic charge α. The electrostatic potential $\phi(\mathbf{r})$ is obtained by solving the Poisson equation[36]

$$\nabla \cdot [\epsilon(\mathbf{r})\nabla\phi_s(\mathbf{r})] = -4\pi\rho(\mathbf{r}), \quad (5)$$

where $\rho(\mathbf{r})$ is the charge density of all the explicit atoms in the system (from macromolecule and inner solvent molecules), $\rho(\mathbf{r}) = \Sigma_\alpha q_\alpha \delta(\mathbf{r} - \mathbf{r}_\alpha)$, and $\epsilon(\mathbf{r})$ is the space-dependent dielectric constant. The dielectric constant is set to ϵ_s and ϵ_m for the solvent and macromolecule in the outer region, respectively (one should note that ϵ_m could be assigned a value different from 1, see Concluding Discussion). It should be noted that the dielectric constant in the inner region is 1 since the solvent molecules are represented explicitly. The influence of ionic strength in the bulk can easily be included by replacing the Poisson Eq. (5) by the Poisson–Boltzmann (PB) equation. The reaction field potential in Eq. (4) is obtained by subtracting a reference electrostatic potential computed in vacuum (ϕ_v), from the electrostatic potential computed in the dielectric solvent environment (ϕ_s). The reference vacuum environment is obtained by setting $\epsilon(\mathbf{r})$ to 1 at all points in the solvent region in the inner and outer regions.

In principle, Eq. (5) can be solved numerically for any arbitrary shape of dielectric boundary and charge distribution using finite-difference overrelaxation methods.[36,37] However, repeated solution of Eq. (5) during a dynamical trajectory is computationally expensive and undesirable. To achieve computational efficiency, an accurate and computationally convenient representation of the electrostatic contribution to the free energy is needed. In the following section, a simple and accurate method called the Generalized Solvent Boundary Potential (GSBP) is developed to achieve this goal.

B. Generalized solvent boundary potential (GSBP)

1. Average smooth dielectric boundary

We seek an accurate closed-form expression for the electrostatic contribution to the free energy, $\Delta\mathcal{W}_{elec}(\mathbf{X})$, which is valid for any arbitrary geometry and charge distribution. One problem is caused by the dependence of the dielectric interfaces on the instantaneous configuration of the explicit atoms in the inner region of the system. While one can assumed that the atoms in the outer region will be kept fixed around some average representative position $\overline{\mathbf{X}}_o$ during a simulation, the atoms in the inner region will be allowed to move. The consequence is illustrated schematically in Fig. 2(a) with an instantaneous configuration of the explicit atoms in the system. It is observed that the details of the dielectric interface, defined by the superposition of the van der Waals envelop of the explicit atoms and the configurational restriction repulsive potential, depend on the instantaneous position of the solvent molecules in the inner region. This complex situation must be simplified to develop a computationally efficient method for simulations. A useful approximation is to assume that the dielectric interface is, on average, smooth and invariant. The nature of this approximation is illustrated schematically in Fig. 2(b). To account for the finite size of the inner solvent molecules lying near the surface of the inner region,

J. Chem. Phys., Vol. 114, No. 7, 15 February 2001

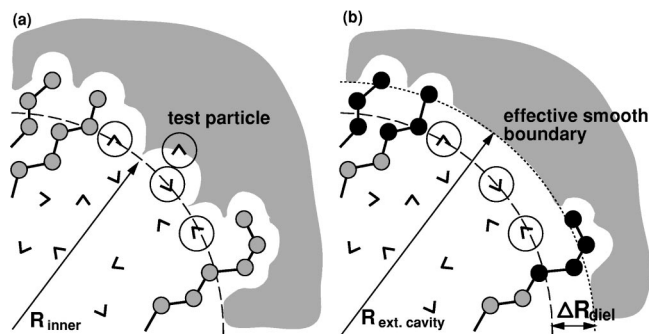Solvent boundary potential   2927

FIG. 2. Schematic description of the interface between the explicit atoms and the implicit solvent. The inner simulation region is represented by a long-dashed line and the implicit solvent region is indicated by the shaded area. (a) Instantaneous configuration of the inner simulation region of radius R_{inner}. The implicit solvent region is represented by the van der Waals envelop of the explicit atoms plus the configurational restriction of the same radius as the inner simulation region. The instantaneous shape of the van der Waals envelop of the system possesses some roughness in the boundary. (b) It is assumed that the interface is smooth and spherical on average in the solvent-exposed region for a practical computational algorithm. To account for the finite size of the inner solvent molecules lying near the surface of the inner simulation region, it is expected that the smooth dielectric boundary will be, on average, at some effective distance ΔR_{diel} from any of the explicit solvent molecules in the inner simulation region. The inner region is extended by ΔR_{diel} to define a smooth spherical dielectric cavity of $R_{\text{ext.cavity}}$. Fixed atoms in the system are represented by black-filled spheres.

it is expected that the smooth dielectric boundary will be, on average, at some effective distance ΔR_{diel} from any of the explicit solvent molecules in the inner region. The effective distance can be empirically chosen to mimic the average electrostatic effects in the solvent-exposed region.[16] It may be noted that this physically-motivated construction avoids the familiar divergence problems associated with the presence of a point charge near the dielectric interface.[5,13,16,28] Rigorously, the effective hard repulsive potential corresponding to the configurational restriction should have a variable radius (corresponding to the farthest solvent molecule). However, the fluctuations are small and can generally be neglected.[16] This considerably simplifies the treatment of the electrostatics.

2. Green's function decomposition

A separation of the contributions from the inner and outer regions in the electrostatic free energy $\Delta \mathcal{W}_{\text{elec}}(\mathbf{X})$ requires a careful analysis. Although it is represented as a sum over all atomic charges in Eq. (4), this form is deceptive. In fact, although the charge density $\rho(\mathbf{r})$ can be expressed as the sum of the charge distributions in the inner region, $\rho^{(i)}(\mathbf{r})$, and the outer region, $\rho^{(o)}(\mathbf{r})$,

$$\rho(\mathbf{r}) = \rho^{(o)}(\mathbf{r}) + \rho^{(i)}(\mathbf{r}), \tag{6}$$

the reaction field ϕ_{rf} in Eq. (4) depends on all the atomic charges. To make progress in the analysis, it is useful to express $\Delta \mathcal{W}_{\text{elec}}(\mathbf{X})$ in terms of the reaction field Green's function $G_{\text{rf}}(\mathbf{r},\mathbf{r}')$ corresponding to the reaction field potential at $\mathbf{r}$ due to a point charge at $\mathbf{r}'$,

$$\phi_{\text{rf}}(\mathbf{r}) = \int d\mathbf{r}' \, G_{\text{rf}}(\mathbf{r},\mathbf{r}') \rho(\mathbf{r}'), \tag{7}$$

such that the free energy can be expressed as,

$$\Delta \mathcal{W}_{\text{elec}} = \frac{1}{2} \int d\mathbf{r} \, d\mathbf{r}' \, \rho(\mathbf{r}) G_{\text{rf}}(\mathbf{r},\mathbf{r}') \rho(\mathbf{r}'). \tag{8}$$

Inserting the charge density Eq. (6) into Eq. (8) yields the various inner–inner, inner–outer, and outer–outer contributions,

$$\Delta \mathcal{W}_{\text{elec}} = \Delta \mathcal{W}_{\text{elec}}^{(\text{oo})} + \Delta \mathcal{W}_{\text{elec}}^{(\text{io})} + \Delta \mathcal{W}_{\text{elec}}^{(\text{ii})}. \tag{9}$$

The first term arises from the reaction field interactions of the charges in the outer region,

$$\Delta \mathcal{W}_{\text{elec}}^{(\text{oo})} = \frac{1}{2} \int d\mathbf{r} \, d\mathbf{r}' \, \rho^{(o)}(\mathbf{r}) G_{\text{rf}}(\mathbf{r},\mathbf{r}') \rho^{(o)}(\mathbf{r}')$$

$$= \frac{1}{2} \sum_{\alpha \in \text{outer}} q_\alpha \phi_{\text{rf}}^{(o)}(\mathbf{r}_\alpha). \tag{10}$$

In a fixed configuration of the atoms in the outer region, this is simply a constant term which does not affect the atoms in the inner region.

The second term arises from the coupling between the charge in the inner region and the reaction field of the charges in the outer region,

$$\Delta \mathcal{W}_{\text{elec}}^{(\text{io})} = \int d\mathbf{r} \, d\mathbf{r}' \, \rho^{(i)}(\mathbf{r}) G_{\text{rf}}(\mathbf{r},\mathbf{r}') \rho^{(o)}(\mathbf{r}')$$

$$= \sum_{\alpha \in \text{inner}} q_\alpha \phi_{\text{rf}}^{(o)}(\mathbf{r}_\alpha) \tag{11}$$

[the symmetry property of the Green's function, $G_{\text{rf}}(\mathbf{r},\mathbf{r}') = G_{\text{rf}}(\mathbf{r}',\mathbf{r})$ has been used to simplify the expression]. The outer reaction field, $\phi_{\text{rf}}^{(o)}$, is equal to the electrostatic potential computed in the dielectric solvent environment, $\phi_s^{(o)}$, minus the reference electrostatic potential computed in vacuum, $\phi_v^{(o)}$. The later corresponds to the bare Coulombic potential between the inner and outer charges of the microscopic potential function, i.e., $\sum_{\alpha \in \text{inner}} q_\alpha \phi_v^{(o)}(\mathbf{r}_\alpha) \equiv U_{\text{elec}}^{(\text{io})}$. Therefore, all the static inner–outer electrostatic interactions can be combined together as

$$U_{\text{elec}}^{(\text{io})} + \Delta \mathcal{W}_{\text{elec}}^{(\text{io})} = \sum_{\alpha \in \text{inner}} q_\alpha \phi_s^{(o)}(\mathbf{r}_\alpha), \tag{12}$$

where $\phi_s^{(o)}(\mathbf{r})$ is the total static field from the charges in the outer region acting onto the inner region taking into account the effect of the dielectric shielding of the environment. This form is computationally advantageous because it is possible to avoid the repeated sum over all pairs of inner and outer (fixed) charges in computing the Coulombic interactions. For a given configuration of the outer atoms, the total static solvent-shielded field $\phi_s^{(o)}(\mathbf{r})$, can be calculated once and stored for efficient computer simulations. This field is independent of the instantaneous configuration of the (moving) atom in the inner region because of the approximation of the smooth average dielectric interface in the solvent-exposed region (see Fig. 2).

The third and last term arises from the reaction field interactions of the charges in the inner region,

$$\Delta \mathcal{W}_{\text{elec}}^{(\text{ii})} = \frac{1}{2} \int d\mathbf{r} \, d\mathbf{r}' \, \rho^{(\text{i})}(\mathbf{r}) G_{\text{rf}}(\mathbf{r}, \mathbf{r}') \rho^{(\text{i})}(\mathbf{r}')$$

$$= \frac{1}{2} \sum_{\alpha \in \text{inner}} q_\alpha \phi_{\text{rf}}^{(\text{i})}(\mathbf{r}_\alpha). \tag{13}$$

In contrast with $\phi_{\text{rf}}^{(\text{o})}$, it is not useful to keep $\phi_{\text{rf}}^{(\text{i})}$ in memory since its value depends on the instantaneous position of the all the charges in the inner region. In principle, the reaction field Green's function is independent of the charge positions and could be stored. However, it is a function of six independent Cartesian coordinates and this would require excessive memory resources and computations. These difficulties are circumvented by expressing the atomic charge distribution in the inner region using a normalized basis set $\{b_m(\mathbf{r})\}$,

$$\rho^{(\text{i})}(\mathbf{r}) = \sum_m c_m b_m(\mathbf{r}). \tag{14}$$

The basis functions are nonzero inside a smooth extended cavity corresponding to the inner region augmented by ΔR_{diel} to account for the physical size of solvent molecules (see the discussion of the dielectric interface above). The coefficients c_m can be written as

$$c_m = \sum_n S_{mn}^{-1} Q_n, \tag{15}$$

where the elements of the overlap matrix S_{nm} and the vector Q_n are given by

$$S_{nm} = \int d\mathbf{r} \, b_n(\mathbf{r}) b_m(\mathbf{r}), \tag{16}$$

$$Q_n = \sum_{\alpha \in \text{inner}} q_\alpha b_n(\mathbf{r}_\alpha). \tag{17}$$

The quantities Q_n play the role of generalized multipole moments. In terms of Eqs. (13), (14), and (15), the basis-set representation on $\Delta \mathcal{W}_{\text{elec}}^{(\text{ii})}$ becomes

$$\Delta \mathcal{W}_{\text{elec}}^{(\text{ii})} = \frac{1}{2} \int d\mathbf{r} \, d\mathbf{r}' \left[\sum_{im} S_{im}^{-1} Q_m b_i(\mathbf{r}) \right] G_{\text{rf}}(\mathbf{r}, \mathbf{r}')$$

$$\times \left[\sum_{jn} S_{jn}^{-1} Q_n b_j(\mathbf{r}') \right]$$

$$= \frac{1}{2} \sum_{mn} \sum_{ij} S_{im}^{-1} Q_m \left[\int d\mathbf{r} \, d\mathbf{r}' \, b_i(\mathbf{r}) G_{\text{rf}}(\mathbf{r}, \mathbf{r}') b_j(\mathbf{r}') \right]$$

$$\times S_{jn}^{-1} Q_n$$

$$= \frac{1}{2} \sum_{mn} Q_m \left[\sum_{ij} S_{im}^{-1} M_{ij} S_{jn}^{-1} \right] Q_n$$

$$= \frac{1}{2} \sum_{mn} Q_m M_{mn}^* Q_n. \tag{18}$$

$\mathbf{M}^*$ is a generalized reaction field matrix for the system and represents the numerical reaction field Green's function between those generalized multipoles Q_n. $\mathbf{M}^*$ is equivalent to $\mathbf{M}$ if the basis set is made up of orthonormal functions, $S_{nm} = \delta_{nm}$. In the case of the familiar Kirkwood expansion,[16]

the corresponding matrix M^* is diagonal (see Eq. (25) below). Although the generalized multipole moments Q_n are calculated for each instantaneous configuration of the inner region during computer simulations, it should be noted that $\mathbf{M}^*$ does not depend on the instantaneous configuration of the atom in the inner region because of the smooth dielectric interface in the solvent-exposed region. Thus, it can be calculated once and stored for efficient computer simulations.

Finally, using Eqs. (9), (12), and (18), the free energy in Eq. (2) becomes

$$\mathcal{W}(\mathbf{X}) = U^{(\text{oo})} + \Delta W_{\text{elec}}^{(\text{oo})} + U^{(\text{ii})} + U_{\text{int}}^{(\text{io})} + U_{\text{LJ}}^{(\text{io})}$$

$$+ \Delta \mathcal{W}_{\text{cr}} + \Delta W_{\text{np}}$$

$$+ \sum_{\alpha \in \text{inner}} q_\alpha \phi_s^{(\text{o})}(\mathbf{r}_\alpha) + \frac{1}{2} \sum_{mn} Q_m M_{mn}^* Q_n, \tag{19}$$

where $U^{(\text{ii})}$, $U^{(\text{io})}$, and $U^{(\text{oo})}$ represent the inner–inner, inner–outer, and outer–outer potential energies, respectively. The last two terms in Eq. (19), corresponding to the total influence of the electrostatic environment on the inner region, constitute the core of the GSBP. This is the main result of the present paper.

For the sake of completeness, the energy contributions involving only the outer atoms, i.e., $U^{(\text{oo})}$ and $\Delta W_{\text{elec}}^{(\text{oo})}$, have been included in Eq. (19). However, as mentioned above, they are constant terms for a fixed configuration of the atoms in the outer region and do not affect the atoms in the inner region. It may be noted that the decomposition of Eq. (9) can be made for any arbitrary partitioning of the atoms (i.e., the atoms do not have to be associated with specific spatial regions). This fact can be exploited in the assignment of the fixed macromolecule atoms (see below). In finite-difference PB calculation the point charge q_α is projected onto the grid using a triliner function.[38] The projected grid charge q_α^{grid} thus depends on the position of atom α. The total force acting on atom α in the inner region is obtained by taking the first derivative of Eq. (19) with respect to the position of atom α,

$$\mathbf{F}_\alpha = - \frac{\partial}{\partial \mathbf{r}_\alpha} (U^{(\text{ii})} + U_{\text{int}}^{(\text{io})} + U_{\text{LJ}}^{(\text{io})}) - \frac{\partial \Delta W_{\text{np}}}{\partial \mathbf{r}_\alpha}$$

$$- \phi_s^{(\text{o})}(\mathbf{r}_\alpha) \frac{\partial q_\alpha^{\text{grid}}}{\partial \mathbf{r}_\alpha} - q_\alpha \sum_{mn} Q_m M_{mn}^* \frac{\partial b_n(\mathbf{r}_\alpha)}{\partial \mathbf{r}_\alpha}, \tag{20}$$

where we exploited the fact that the static field $\phi_s^{(\text{o})}(\mathbf{r})$ and $\mathbf{M}^*$ do not depend on the position of atom α. The spatial derivatives of the grid charge q_α^{grid} can be calculated analytically.[38,39]

C. Implementation and computational details

The GSBP method has been implemented into the PBEQ module[21,38,40] of the biomolecular simulation program CHARMM[41] for spherical and orthorhombic inner simulation regions. Orthonormal basis-sets based on spherical harmonics or cartesian Legendre polynomials for these geometries are given in the Appendix. In a standard application of GSBP, the solvent-shielded electrostatic potential $\phi_s^{(\text{o})}(\mathbf{r})$ and the generalized reaction field matrix $\mathbf{M}^*$ are first calcu-

lated and stored. Then, during a simulation, the generalized multipole moments Q_n are calculated for every instantaneous configuration of the atoms in the inner region. Of course, all the moving atoms (macromolecule, ligand, and solvent) lying in the inner region must be assigned to the inner set. However, the treatment of the fixed macromolecule atoms lying in the intermediate space located outside the inner region, but within the extended dielectric cavity with ΔR_{diel}, leaves us some choice. Those fixed macromolecule atoms can be arbitrarly assigned to the outer set or to the inner set on the basis of computational convenience. Their influence is included in $\phi_{\mathrm{s}}^{(\mathrm{o})}$ in the former, while it is included in $U_{\mathrm{elec}}^{(\mathrm{ii})} + \Delta \mathcal{W}^{(\mathrm{ii})}$ in the later. In the present implementation we chosed to assign the intermediate atoms to the inner region to have an accurate microscopic potential $U_{\mathrm{elec}}^{(\mathrm{ii})}$ and avoid the natural inaccuracy of the grid-based potential $\phi_{\mathrm{s}}^{(\mathrm{o})}$ near the source charges. This choice also facilitates the correct handling of the 1–2 and 1–3 nonbonded exclusion between covalently bonded neighbors.[41]

To calculate $\phi_{\mathrm{s}}^{(\mathrm{o})}(\mathbf{r})$, all atomic charges in the inner region are set to zero and the PB equation for the system,

$$\boldsymbol{\nabla} \cdot [\epsilon(\mathbf{r}) \boldsymbol{\nabla} \phi_{\mathrm{s}}^{(\mathrm{o})}(\mathbf{r})] - \bar{\kappa}^2(\mathbf{r}) \phi_{\mathrm{s}}^{(\mathrm{o})}(\mathbf{r}) = -4\pi \rho^{(\mathrm{o})}(\mathbf{r}), \quad (21)$$

is solved using finite-difference methods.[36,37] The result of the calculation, i.e., the electrostatic potential on each point of the discrete grid, is then stored.

The calculation of $\mathbf{M}$ exploits the fact that the nmth matrix element corresponds to the interaction between the reaction field due to the charge density supported by the basis function $b_m(\mathbf{r})$ and the charge density of $b_n(\mathbf{r})$,

$$M_{nm} = \int d\mathbf{r}\, d\mathbf{r}'\, b_n(\mathbf{r}) G_{\mathrm{rf}}(\mathbf{r}, \mathbf{r}') b_m(\mathbf{r}')$$

$$= \int d\mathbf{r}\, b_n(\mathbf{r}) \phi_{\mathrm{rf}}(\mathbf{r}; b_m(\mathbf{r})), \quad (22)$$

where $\phi_{\mathrm{rf}}(\mathbf{r}; b_m(\mathbf{r})) \equiv [\phi_{\mathrm{s}}(\mathbf{r}; b_m(\mathbf{r})) - \phi_{\mathrm{v}}(\mathbf{r}; b_m(\mathbf{r}))]$ is the reaction field due to the basis function $b_m(\mathbf{r})$. The basis functions are nonzero inside the dielectric cavity corresponding to the inner region extended by ΔR_{diel}. To calculate the matrix element, all atomic charges in the system are first set to zero. Then, the PB equation is solved with the full dielectric environment using the basis function as a continuous charge distribution in the extended dielectric cavity,

$$\boldsymbol{\nabla} \cdot [\epsilon(\mathbf{r}) \boldsymbol{\nabla} \phi_{\mathrm{s}}(\mathbf{r}; b_m(\mathbf{r}))] - \bar{\kappa}^2(\mathbf{r}) \phi_{\mathrm{s}}(\mathbf{r}; b_m(\mathbf{r})) = -4\pi b_m(\mathbf{r}).$$

$$(23)$$

The reaction field is calculated by subtracting the electrostatic potential from the reference vacuum system,

$$\boldsymbol{\nabla} \cdot [\epsilon(\mathbf{r}) \boldsymbol{\nabla} \phi_{\mathrm{v}}(\mathbf{r}; b_m(\mathbf{r}))] = -4\pi b_m(\mathbf{r}). \quad (24)$$

In the reference vacuum calculation, the dielectric constant of the solvent in the outer region is set to 1, while the dielectric constant of the macromolecule in the outer region is set to ϵ_{m} which may differ from unity.

Having chosen an appropriate basis set $\{b_m(\mathbf{r})\}$ for the shape of the inner region, the procedure for the construction of the reaction field matrix $\mathbf{M}^*$ is the following:

(i)　　Setup the finite-difference PB equation with a normalized basis function b_m. All atomic charges in the inner and outer regions are set to zero while the charge on each grid point inside the inner region extended by ΔR_{diel} is assigned by the basis function b_m, $q(i,j,k) = b_m(i,j,k) h^3$, where h is the grid spacing and $q(i,j,k)$ and $b_m(i,j,k)$ are the charge and the basis function at the grid point (x_i, y_j, z_k);

(ii)　　Solve the finite-difference PB Eqs. (23) and (24) in both environments in order to obtain the reaction field for the basis function b_m, $\phi_{\mathrm{rf}}(\mathbf{r}; b_m(\mathbf{r}))$;

(iii)　　Calculate the matrix elements (M_{nm}). The matrix elements are calculated by projecting the reaction field of the basis function b_m onto a charge distribution corresponding purely to the basis function b_n and calculating the integral Eq. (22) for $n = 1, ..., m$;

(iv)　　Repeat the previous steps for the desired number of basis functions;

(v)　　Calculate the overlap matrix $\mathbf{S}$ in Eq. (16) if the basis functions are not orthogonal. In this case, the reaction field matrix is given by $\mathbf{M}^* = \mathbf{S}^{-1} \cdot \mathbf{M} \cdot \mathbf{S}^{-1}$.

A standard procedure in finite-difference PB calculations is to set the value of the potential at the edge of the grid using the analytical solution from the Debye–Huckel theory summed over all atomic charges in the system.[36] However, this simple procedure causes dificulties when using a continuous basis function b_n as a charge density. The large number of grid-charges corresponding to the basis function increases significantly the computational cost of setting up the boundary conditions at the edge of the grid. To reduce the number of grid-charges, the PB equation is first solved for a coarse grid with large spatial extent and the result is refined subsequently by using a focused calculation.[42]

## III. NUMERICAL TESTS AND APPLICATIONS

As a first step, numerical tests were performed with very simple model systems composed of point charges inside spherical and cubic cavities to check our ability to express the reaction field $\phi_{\mathrm{rf}}^{(\mathrm{i})}(\mathbf{r})$ using a finite expansion of basis functions. As a second step, the method is tested and illustrated in the case of two realistic systems: aspartyl-tRNA synthetase[29–31] with a spherical inner region and the KcsA potassium channel[32,43] with an orthorhombic inner region.

A. Numerical tests with simple models

Two model systems were constructed of two point charges ($q_A = q_B = +1.0e$) inside a spherical cavity with a radius of 12 Å and a cubic cavity with a length of 19.6 Å. Both cavities were centered at the origin and enclosed in a isotropic dielectric continuum with $\epsilon_{\mathrm{s}} = 80$. The two simple systems were chosen because they provide important tests of the basis-set approach. All PB calculations were performed using a 65^3 cubic grid with a grid spacing of 0.8 Å (coarse grid) followed by a focussing to a grid spacing of 0.4 Å (fine grid). Both coarse and fine cubic grids were centered at the origin. 19 multipoles ($L = 18$) consisting of total of 361 basis
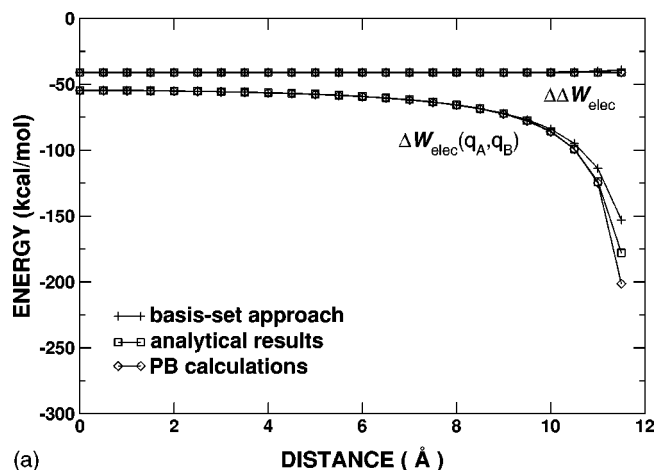
(a)
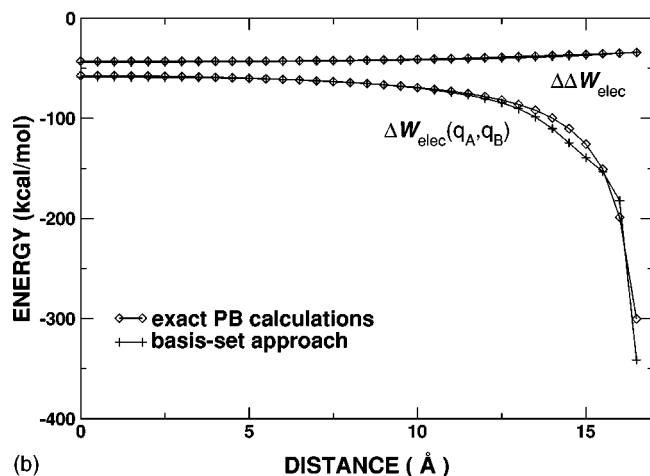
(b)

FIG. 3. (a) Total and charging electrostatic solvation energies as a function of the distance of moving point charge B in the spherical cavity; basis-set approach ($+$), analytical results ($\square$), and PB calculations ($\diamond$). (b) Total and charging electrostatic solvation energies as a function of the distance of moving point charge B in the cubic cavity; PB calculations ($\diamond$) and basis-set approach ($+$). The charging electrostatic solvation energy is defined as $\Delta\Delta\mathcal{W}_{\text{elec}} = \Delta\mathcal{W}_{\text{elec}}(q_A, q_B) - \Delta\mathcal{W}_{\text{elec}}(q_A = 0, q_B)$.

FIG. 4. Convergence of the total electrostatic solvation energies as a function of number of multipoles for different position of charge B; basis-set approach (solid line) and analytical results (dashed line).

charge distribution inside an orthorhombic cavity enclosed by a dielectric medium. In this case, the accuracy of the basis-set expansion can be examined only by comparing with numerical finite-difference PB calculation.

Figures 3(a) and 3(b) show the variations in the electrostatic free energy as a function of the position of the charge q_B in the spherical and cubic cavities, respectively, keeping the charge q_A fixed at the center of both cavities. In both cases, the results are excellent except when the moving charge q_B is approaching the dielectric boundary. Then, the reaction field exhibits the familiar divergence phenomena[28,5,13,16] and a very large number of multipoles is required to calculate the electrostatic solvation energy accurately. The convergence of the basis-set expansion for $\Delta\mathcal{W}_{\text{elec}}$ is shown in Fig. 4 for various positions of the charge q_B in the case of the spherical cavity. It is observed that the free energy does not converge very well, even with 19 multipoles, as the charge gets too close to the boundary. However, in a real system such divergences are unphysical since an intermediate region between an atomic charge and the dielectric boundary should exist (see Fig. 2).

When both point charges are located at the origin of the spherical cavity, all multipolar moments in Eq. (26) are zero and Eq. (25) reduces to the familiar Born energy,[44] $\Delta\mathcal{W}_{\text{elec}} = Q^2/2R(1/\epsilon_s - 1)$. For the spherical cavity, the Born energy is -54.65 kcal/mol (with $Q = +2.0e$, $R = 12$ Å, and $\epsilon_s = 80$). In comparison, both the GSBP basis-set expansion and the finite-difference PB calculation yield -54.84 kcal/mol. The accord with the Born energy is excellent. The slight difference between the analytical and numerical results is inevitable due to the discreteness of the grid. Interestingly, the electrostatic solvation energy arises only from the monopole even for an arbitrary charge distribution as long as it possesses spherical symmetry. According to Gauss' Law, the potential and electric field outside the charge distribution are the same, whether a single point charge is at the origin or a uniform charge density is distributed inside the sphere. Thus, although the spherical harmonic function Y_{00} does not reproduce the charge distribution corresponding to a point charge located at the origin, the electrostatic solvation energy is per-

functions were used for the spherical case and total of 343 basis functions of 7 Legendre polynomials in each direction for the cubic case.

In the case of the spherical cavity, the Kirkwood reaction field provides a closed-form expression for the electrostatic free energy of an arbitrary charge distribution,[28]

$$\Delta\mathcal{W}_{\text{elec}} = -\frac{1}{2}\sum_{lm}\frac{4\pi|Q_{lm}|^2}{(2l+1)}\frac{1}{R^{2l+1}}\left[\frac{\epsilon_s - 1}{\epsilon_s + l/(l+1)}\right], \quad (25)$$

where R is the radius of the cavity and the multipole moment Q_{lm} is given by

$$Q_{lm} = \sum_{\alpha} q_\alpha Y_{lm}^*(\theta_\alpha, \phi_\alpha)|\mathbf{r}_\alpha|^l, \quad (26)$$

where α is an index for the point charges and Y_{lm} is the spherical harmonics (see Appendix). Thus, the accuracy of the basis-set approach to the electrostatic solvation energy can be assessed by comparing with Eq. (25) and the results from finite-difference PB calculations. In contrast, there is no closed-form expression for the electrostatic free energy for a
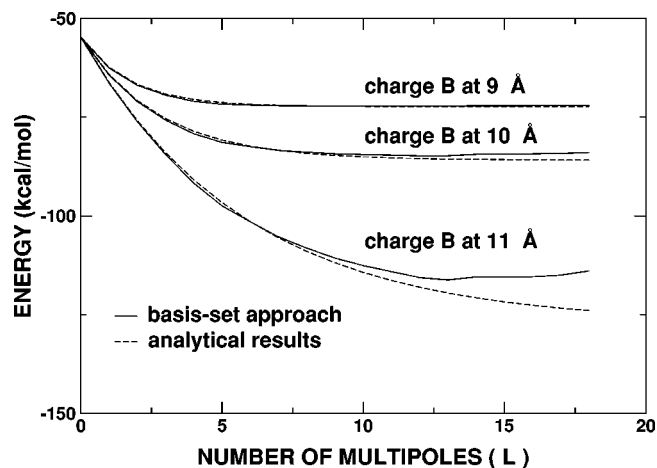
fectly reproduced. This example shows why a basis-set expansion can converge rapidly. The true test of the basis-set expansion is not how well the charge density is reproduced, but how accurate is the electrostatic free energy. Even though this argument is rigorously valid only in the case of a spherical region, it remains essentially correct, even in the case of an orthorhombic region. When both point charges are located at the origin of the cubic region, the free energy calculated with finite-diffence PB is -57.25 kcal/mol. Correspondingly, the basis-set expansion with 343 basis functions gives -58.89 kcal/mol, in good accord with the numerical calculation. The dominant contribution still arises from the uniform basis function which yielded -53.61 kcal/mol. In contrast to the case of a spherical region, however, the basis-set expansion in the orthorhombic cavity requires more than the uniform basis function, even for the centered charges.

Charging free energies play a crucial role in many free energy perturbation simulations of biomolecular systems. As shown in Figs. 3(a) and 3(b), the charging free energy of the charge q_A defined as $\Delta\Delta\mathcal{W}_{\text{elec}} = \Delta\mathcal{W}_{\text{elec}}(q_A, q_B) - \Delta\mathcal{W}_{\text{elec}}(q_A = 0, q_B)$ remains accurate, even when the charge q_B is approaching the boundary. This is an important observation which will enable us to use GSBP to perform thermodynamic free energy perturbation simulations involving changes in the charges of a subset of atoms in the inner region.

B. Tests with realistic atomic systems

The optimal shape of the inner region depends on the molecular system. A spherical inner region is well adapted for the active site region of a large protein such as aspartyl-tRNA synthetase,[29–31] whereas an orthorhombic region is better for the interior of a long narrow ion channel such as KcsA potassium channel.[32,43] Both systems were chosen to illustrate and test GSBP. In the following, all the finite-difference PB calculations were done with the PBEQ module[21,38,40] of the biomolecular simulation program CHARMM[41] using the set of optimized atomic radii for amino acids.[21]

1. Aspartyl-tRNA synthetase

The atomic model of the *Escherichia coli* aspartyl-tRNA synthetase enzyme was graciously provided by Simonson.[29–31] The model was originally constructed from one dimer of the enzyme (590 residues/monomer), one ligand molecule (aspartic acid) bound in the active site, and 384 water molecules solvating a 20 Å spherical region around the active site.[29–31] In the GSBP application, a spherical inner region was centered on the active site with a radius of 15 Å. The inner region was then extended by $\Delta R_{\text{diel}} = 2$ Å to define a smooth spherical dielectric cavity of 17 Å. For the sake of clarity and further analysis, the spherical region of 15 Å radius is called the "inner simulation region," while that of 17 Å is called the "extended cavity region." The model system is shown schematically in Fig. 5(a). All the results are summarized in Table I.

The basis functions were generated from the first 19 multipolar spherical harmonics as described in Eq. (A1) of
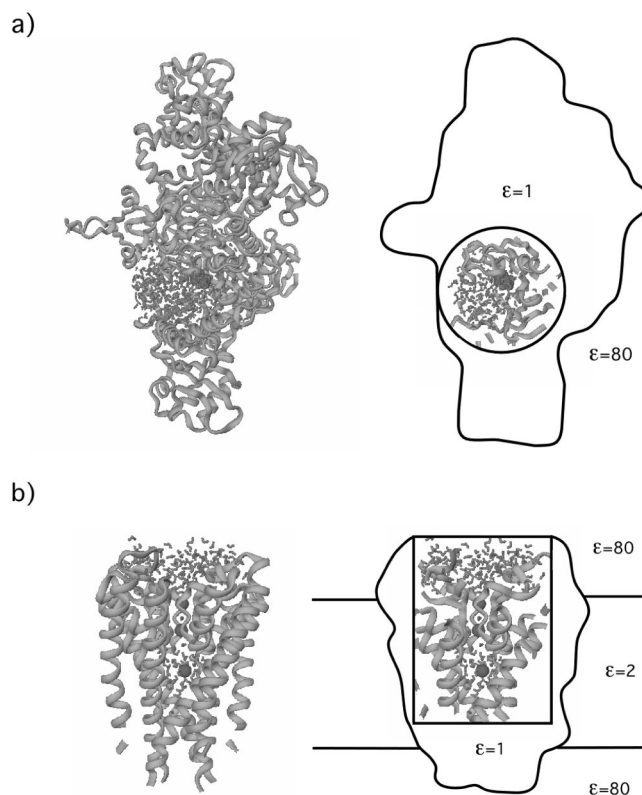

FIG. 5. Schematic representation of the simulating systems for (a) aspartyl-tRNA synthetase, and (b) the KcsA potassium channel. The full systems are shown in the left-hand side and the reduced system for GSBP with dielectric interfaces are shown in the right-hand side. The inner simulation region is defined by (a) a spherical region of a radius of 15 Å centered on the active site and (b) a 26.4 Å×26.4 Å×38.4 Å orthorhombic box centered at (0.0, 0.0, 10.0).

TABLE I. GSBP for spherical region with Aspartyl-tRNA synthetase.[a,b]

Radius of spherical inner region	15 Å
Number of atoms:	
outer region	16958
extended dielectric cavity region	2056
inner simulation region	1357
Interaction energy between the outer and extended dielectric cavity regions:	
Potential function $U_{\text{elec}}^{\text{(io)}}$	-1270.59
GSBP with potential stored on grid (vacuum environment)	-1276.12
Interaction energy between the outer and inner simulation regions:	
Potential function $U_{\text{elec}}^{\text{(io)}}$	-116.82
GSBP with potential stored on grid (vacuum environment)	-117.74
Reaction field energy for the extended dielectric cavity region:	
Finite-difference PB	-745.18
GSBP basis-set expansion	-798.09
Reaction field energy for the inner simulation region:	
Finite-difference PB	-249.35
GSBP basis-set expansion	-246.16

[a]All energies are given in kcal/mol.
[b]All PB calculations were performed using a 111^3 grid with with a grid spacing of 1.2 Å followed by a focusing to a spacing of 0.4 Å.
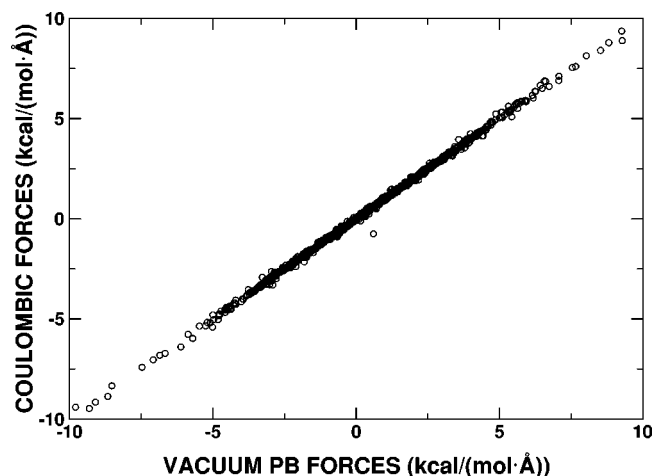
FIG. 6. Vacuum PB electrostatic forces versus the corresponding Coulombic forces in the aspartyl-tRNA synthetase system.

the Appendix. This yields a total of 361 basis functions expansion of the inner-inner reaction field. The reaction field matrix $\mathbf{M}^*$ and the static field $\phi_s^{(o)}$ were calculated using a 111^3 cubic grid with a grid spacing of 1.2 Å (coarse grid) followed by a focussing to a grid spacing of 0.4 Å (fine grid). The coarse grid was centered at the origin, whereas the fine grid was centered on the active site. The dielectric constant of the solvent outside the spherical inner region was set to 80 with no salt concentration. The dielectric constant of the protein was set to 1.

The validity of the decomposition of Eq. (9) was examined by calculating the inner–inner, inner–outer, and outer–outer contributions with PB. It was verified that the sum of the individual inner–inner, inner–outer, and outer–outer contributions was equal exactly to the total electrostatic solvation energy $\Delta \mathcal{W}_{\text{elec}}$ calculated directly with PB (data not shown). The accuracy of the solvent-shielded static field energy $\phi_s^{(o)}$ and the corresponding forces was assessed by comparing finite-difference PB calculations for a vacuum environment with the analytical Coulombic energy and forces of the microscopic potential function. As shown in Table I, the Coulombic potential energy $U_{\text{elec}}^{(\text{io})}$ (which includes explicitly all charge-charge interactions) is reproduced very well by the results from finite-difference PB calculations. Due to discreteness of the grid, the calculated potential $\phi_v^{(o)}$ becomes more accurate at a certain distance from source charges. This is clearly shown in Table I when only the inner simulation region is included in the calculation. The inner–outer forces are also very accurately reproduced. Figure 6 shows the X-, Y-, and Z-components of the vacuum PB electrostatic forces of each atom inside the inner simulation region versus the corresponding Coulombic forces (fixed protein atoms in the intermediate region and hydrogen atoms are not shown on the plot). The excellent results for the inner-outer electrostatic energy and forces for the vacuum environment suggest that the calculation of the static field $\phi_s^{(o)}$ shielded by the solvent environment should also be accurate. As also shown in Table I, some differences are observed between the PB and basis-set calculations for the reaction field energy when all the atoms inside the extended dielectric cavity are in-

cluded. According to the previous tests (Sec. III A), the inaccuracies come mostly from the charges located in the intermediate region near the dielectric surface of the cavity. In MD simulations with GSBP, the intermediate protein atoms must be fixed. Nonetheless, the finite basis set can reproduce very well the electrostatic charging free energy for the (moving) atoms in the inner region. An accurate description of the charging free energy of the moving atoms will be important to perform free energy perturbation simulations involving alchemical modification of the atomic charges using GSBP.

Generally, the electrostatic free energy may be dominated by a small subset of basis functions, typically the multipoles of lowest order. To increase the efficiency of the algorithm it is desirable to limit the total number of basis functions involved in the expansion. In Kirkwood's classical reaction field, the importance of the contributions increases with the multipolar order l. In the general case, however, the generalized reaction field matrix $\mathbf{M}^*$ is not diagonal and it is unclear that such an ordering is optimal. It is possible to make a sorted list of the basis functions according to their contribution to the reaction field energy,

$$E_m = \left| \frac{1}{2} Q_m^2 M_{mm}^* + \sum_{n \neq m} Q_n M_{nm}^* Q_m \right|. \tag{27}$$

To construct the sorted list, the matrix elements related to the basis function having the lowest contribution are removed from $\mathbf{M}$ and the sum in Eq. (27) is recalculated using the remaining basis functions. Figure 7(a) shows the variations of the GSBP reaction field energies with and without the sorting in Eq. (27) as a function of number of basis functions. The sorting significantly increases the convergence of the sum. In the sorted list, more than half of the functions contribute very little to the reaction field energy.

Finally, the stability and accuracy of the basis-set reaction field forces was examined by comparing with the analytical forces calculated from finite-difference PB (see Refs. 38 and 45 for more details about the analytical treatment of PB forces). All X-, Y-, and Z-components of the basis-set reaction field forces of each atom inside the inner simulation region vs the corresponding PB forces are shown in Fig. 8(a). [For the sake of clarity, the figures only show the forces which are between -5.0 and 5.0 kcal/(mol Å).] The basis-set reaction field forces reproduce well the PB reaction field forces. As shown in Fig. 8(b), the accuracy is slightly deteriorated when the number of basis functions is reduced from 361 to 200. Table II shows the basis-set reaction field energies and force errors as a function of number of basis functions. These results indicate that the accuracy of the forces is more sensitive to the number of basis functions than the energy.

2. KcsA potassium channel

The atomic model was originally constructed of the tetrameric KcsA channel with 3 potassium ions,[32] around 80 water molecules in the pore, 112 DPPC lipids (46 in the upper layer and 66 in the lower layer), and around 6500 water molecules to form the bulk region with 23 chloride and 12 potassium ions to represent a 150 mM ionic salt.[43] For the
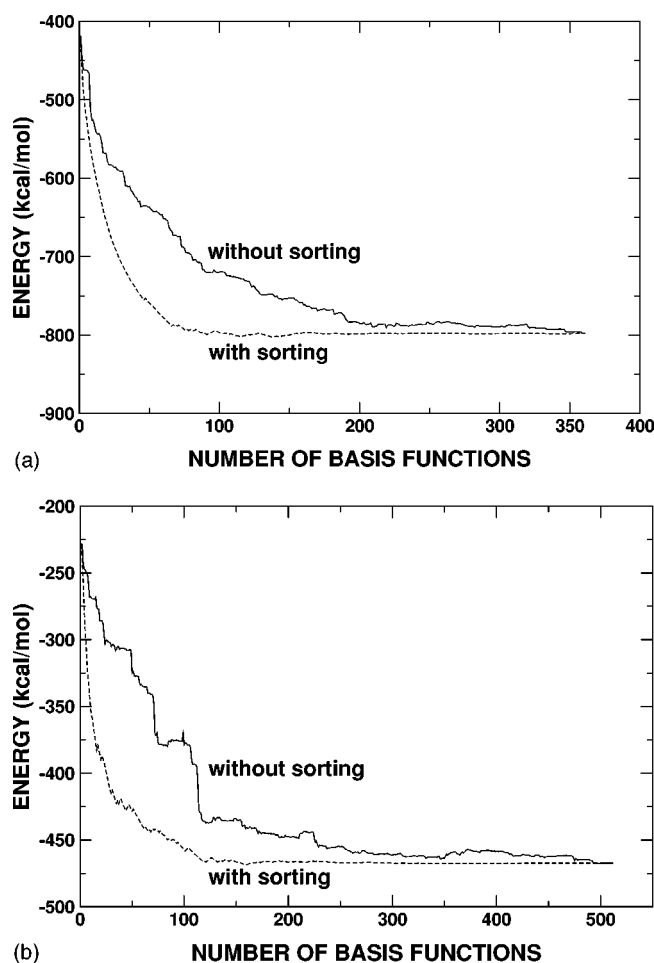
(a)

(b)

FIG. 7. Basis-set reaction field energies versus number of basis functions in (a) the aspartyl-tRNA synthetase system and (b) the KcsA potassium channel system. with (dashed line) and without (solid line) the sorting method.
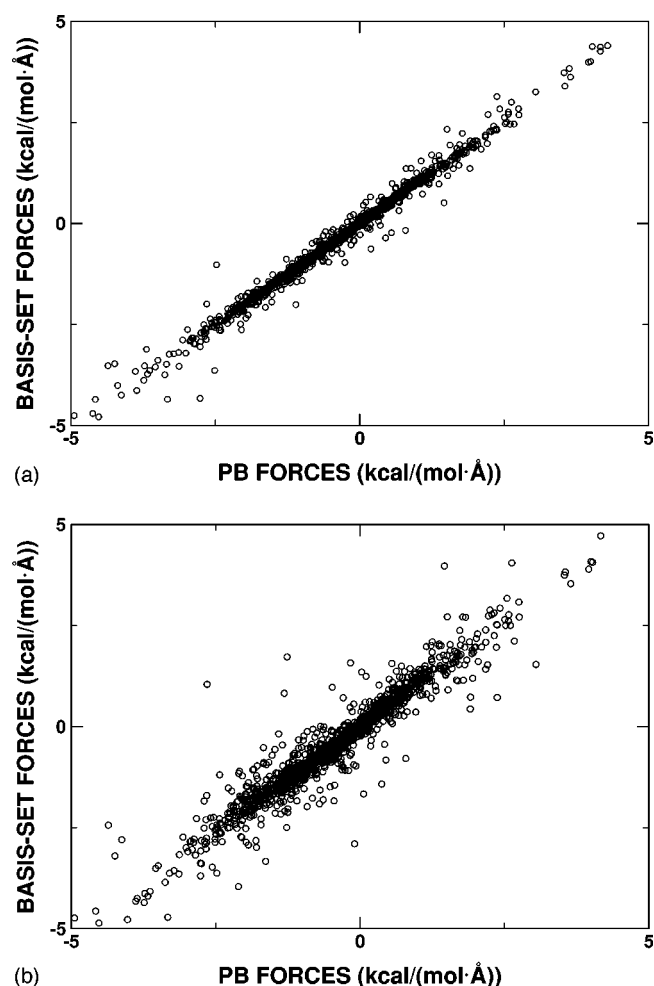

(a)

(b)

FIG. 8. Basis-set reaction field forces with (a) 361 and (b) 200 limited basis functions versus the corresponding PB forces in the aspartyl-tRNA synthetase system. For the sake of clarity, the figures only show the forces which are between -5.0 and 5.0 kcal/(mol Å). Forces which are outside the range also show good agreement (data not shown).

present test all lipid molecules and ions except three potassium ions in the pore were removed. For GSBP, the inner simulation region was defined as a 26.4 Å$\times$26.4 Å$\times$38.4 Å orthorhombic box centered at (0.0,0.0,10.0). The inner region was then extended by $\Delta R_{\mathrm{diel}} = 2$ Å in all directions to define a smooth 28.4 Å$\times$28.4 Å$\times$40.4 Å orthorhombic dielectric cavity. The model system is shown in Fig. 5(b).

The basis functions were generated from all combination of the first eight Legendre polynomials in each Cartesian direction as described in Eq. (A10) of the Appendix. This yields a total of 512 basis functions expansion of the inner-inner reaction field. The reaction field matrix $\mathbf{M}^*$ and the static field $\phi_s^{(o)}$ were calculated using a 85^3 coarse grid with a grid spacing of 1.2 Å followed by a $85\times85\times115$ fine grid with a grid spacing of 0.4 Å. The coarse grid was centered at the origin and the fine grid at (0.0,0.0,10.0). The dielectric constant of the solvent and membrane regions outside the orthorhombic inner region were set to 80 and 2, respectively. The dielectric constant of the protein was set to 1. A 150 mM ionic salt was assumed in the bulk solvent region. The thickness of the membrane was set to 29.0 Å along the Z-axis extending in the X and Y direction. The finite-difference PB equation was solved with periodic boundary conditions in

the X and Y directions. All the results about the system are summarized in Table III.

Although the results demonstrate that GSBP is valid, some differences between the numerical values of $\mathcal{W}_{\mathrm{elec,v}}^{(\mathrm{io})}$ and $U_{\mathrm{elec}}^{(\mathrm{io})}$ are observed. The values of Table III are -350.83, and -369.51, respectively. In contrast, the corresponding values from Table I (the aspartyl-tRNA synthetase system), which are -116.82 and -117.74, respectively, appear to be

TABLE II. The GSBP reaction field energies and forces in the aspartyl-tRNA synthetase system.

No. basis functions	$\Delta \mathcal{W}_{\mathrm{elec}}^{(\mathrm{ii}) \, a}$	Force error[b]			Time[c]
		Avg.	Max.	RMSD	
100	-796.81	0.24	3.87	0.12	0.10
200	-798.45	0.19	3.70	0.11	0.15
300	-798.09	0.16	2.96	0.08	0.21
361	-798.09	0.07	1.64	0.02	0.25

[a]The GSBP reaction field energy in kcal/mol.
[b]The force error is defined as $|\mathbf{F}_{\mathrm{GSBP}} - \mathbf{F}_{\mathrm{PB}}|$ in kcal/(mol Å).
[c]CPU time in second on a Pentium III (700 MHz). The sorting of the basis functions takes around 0.75 s.

TABLE III. GSBP for rthorhombic region with KcsA potassium channel.[a,b]

Dimension of orthorhombic box	28.4 Å × 28.4 Å × 40.4 Å
Number of atoms:	
outer region	3470
extended dielectric cavity region	2797
inner simulation region	1901
Interaction energy between the outer and extended dielectric cavity regions:	
Potential function $U_{elec}^{(io)}$	−3054.18
GSBP with potential stored on grid (vacuum environment)	−3252.88
Interaction energy between the outer and inner simulation regions:	
Potential function $U_{elec}^{(io)}$	−350.83
GSBP with potential stored on grid (vacuum environment)	−369.51
Reaction field energy for the extended dielectric cavity region:	
Finite-difference PB	−455.67
GSBP basis-set expansion	−467.35
Reaction field energy for the inner simulation region:	
Finite-difference PB	−68.72
GSBP basis-set expansion	−66.86

[a]All energies are given in kcal/mol.
[b]All PB calculations were performed using a 85^3 grid with a grid spacing of 1.2 Å followed by a 85×85×115 grid with a spacing of 0.4 Å.

much closer. These differences arise from contributions from the neighboring systems due to the *XY* periodicity used in the PB calculations with a membrane; $\mathcal{W}_{elec,v}^{(io)}$ includes the contributions from infinite image charges in the *XY* directions, but the biomolecular potential function $U_{elec}^{(io)}$ does not. Calculations with an increasingly large grid can be used to demonstrate the accuracy of $\phi_s^{(o)}$ in such an *XY* periodic system (results not shown). In a realistic situation with a high solvent dielectric constant and ionic screening, the influence from the periodic neighbors is negligible.

Table III also shows the results for the PB and basis-set reaction field energies. The reaction field energy of the atoms in the inner simulation region is very well reproduced by the finite basis set. Figure 7(b) shows the variations of the basis-set reaction field energies with and without the sorting in Eq. (27) as a function of number of basis functions. An efficient convergence is achieved after sorting. Figure 9(a) shows all *X*-, *Y*-, and *Z*-components of the basis-set reaction field forces of each atom inside the inner simulation region vs the exact anlytical PB forces (the intermediate atoms and hydrogen atoms are not shown). Calculation with only 300 basis functions (with the sorting method) is shown in Fig. 9(b). [For the sake of clarity, the figures only show the forces which are between −5.0 and 5.0 kcal/(mol·Å).] As observed in the previous test, the forces are more sensitive to the number of basis functions than the energy. This is also clearly shown in Table IV, where the errors are given as a function of number of basis functions. Nevertheless, the basis-set calculation with 300 basis functions reproduces well the reaction field energy and forces from PB, as shown in Fig. 9(b) and Table IV.

3. Dynamical simulations

As a final illustration of the current implementation of GSBP in the program CHARMM,[41] we calculated several tra-
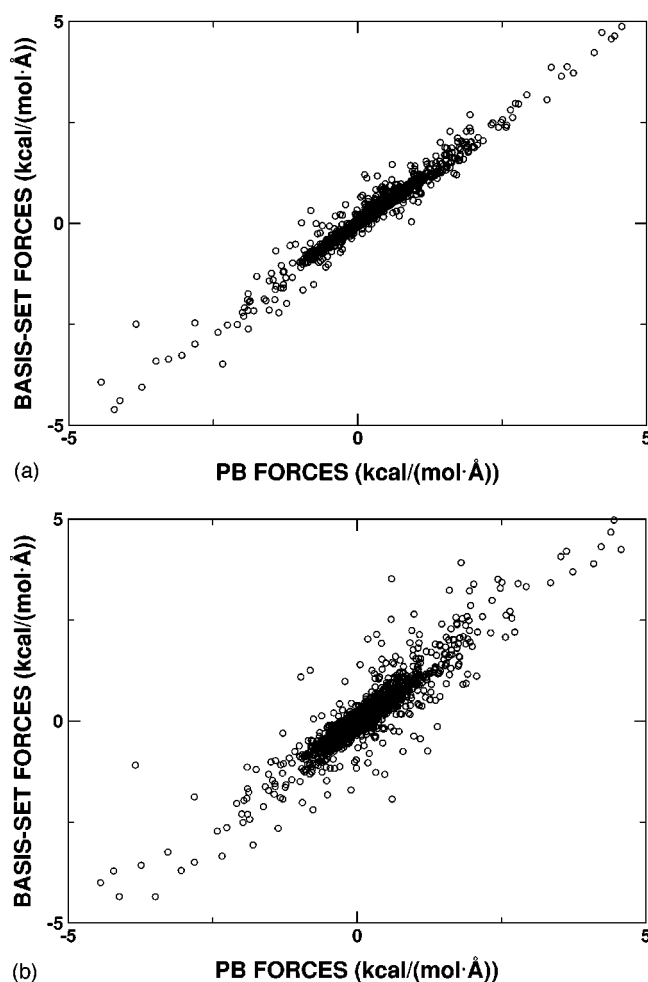

FIG. 9. Basis-set reaction field forces with (a) 512 and (b) 300 limited basis functions versus the corresponding PB forces in the KcsA potassium channel system. For the sake of clarity, the figures only show the forces which are between −5.0 and 5.0 kcal/(mol Å). Some forces which are outside the range also show good agreement (data not shown).

jectories for the aspartyl-tRNA synthetase system. The Verlet algorithm was used with a time step of 0.001 ps and the temperature was 300 K. All the 361 basis functions were used for the reaction field calculations without the sorting. All groups of protein atoms belonging to 1–3 pairs from atoms in the outer region were also fixed for the consistent

TABLE IV. The basis-set reaction field energies and force errors in the KcsA potassium channel system.

No. basis functions	$\Delta \mathcal{W}_{elec}^{(ii)}$ [a]	Force error[b]			Time[c]
		Avg.	Max.	RMSD	
100	−457.34	0.20	3.37	0.09	0.03
200	−466.61	0.15	2.93	0.06	0.05
300	−467.34	0.10	3.64	0.05	0.07
400	−467.51	0.09	2.42	0.03	0.10
512	−467.35	0.06	1.34	0.01	0.13

[a]The basis-set reaction field energy in kcal/mol.
[b]The force error is defined as $|\mathbf{F}_{GSBP} - \mathbf{F}_{PB}|$ in kcal/(mol Å).
[c]CPU time in second on a Pentium III (700 MHz). The sorting of the basis functions takes around 3.50 s.

treatment of electrostatic interactions coming from the static field $\phi_s^{(o)}$ with inner–outer Coulombic interactions $U_{elec}^{(io)}$.

Three sets of MD trajectories were generated both with and without GSBP. For the first and second sets the Coulombic potential was used with no cutoff or a 12 Å cutoff distance, respectively. For the last one, the extended electrostatic treatment using a multipole approximation for groups more than 12 Å apart was used.[46] The total energy was very well conserved in the course of trajectories with GSBP, showing that the GSBP forces are calculated accurately. The cost of GSBP depends on the calculation method for the nonbonding interactions. The simulations without GSBP took 6.7 min/ps (with cutoff), 7.8 min/ps (with extended electrostatics), and 39.8 min/ps (with no cutoff), whereas the simulations with GSBP took 9.4 min/ps (with cutoff), 9.9 min/ps (with extended electrostatics), and 46.8 min/ps (with no cutoff) on a Pentium III (700 MHz). It is quite remarkable that incorporating GSBP requires only 27% more computational time than a simple vacuum simulation with the extended electrostatic treatment. For comparison, a fully solvated aspartyl-tRNA synthetase system would contain more than 70 000 atom and would take on the order of 7 h/ps with the Particle Mesh Ewald (PME). Since the relative cost strongly depends on the number of atoms in the inner region, it may be most useful to use GSBP for systems with few hundreds to few thousands atoms in the inner region. Furthermore, a judicious choice of the frequency of the sorting with a limited number of basis functions can reduce the cost of GSBP (see Table II). For example, the simulation with a cutoff took 8.3 min/ps with limited 200 basis functions and update frequency of 100 step for the sorted list.

IV. CONCLUDING DISCUSSION

A general approach has been developed to allow accurate simulations of a small region of interest (the inner atoms) from a large macromolecular system and incorporate the influence of the remaining atoms (the outer atoms) with an effective boundary potential. Long-range electrostatic effects are included by treating the surrounding solvent as a continuum dielectric. Both the solvent-shielded static field from the outer atoms of the macromolecule and the reaction field from the dielectric solvent acting on the inner atoms are included. The solvent-shielded static field from the outer atoms is calculated once from the finite-difference PB equation and the result is stored for efficient simulations. The solvent reaction field is expressed in terms of a basis-set expansion. The basis-set coefficients correspond to generalized electrostatic multipoles. A matrix representing the reaction field Green's function between those generalized multipoles is calculated only once using the PB equation and stored for efficient simulations. The generalized multipoles are calculated ''on the fly'' during a simulation. Interestingly, a basis-set expansion of the electrostatic potential has been used previously to solve the Poisson equation for small molecules immersed in solvent.[47] Despite some similarities, there are significant differences with the present method. In particular, we chose to expand the charge density in the inner region instead of the potential. This choice enabled us to calculate

directly a generalized Green's function reaction field matrix **M*** for dielectric interfaces of arbitrary shape.

The GSBP method was implemented in the biomolecular simulation program CHARMM[41] for spherical and orthorhombic inner simulation regions. The accuracy of GSBP was examined and tested for simple systems, and two detailed atomic systems: the active site region of a large aspartyl-tRNA synthetase with a spherical inner region, and the interior of the KcsA potassium channel with an orthorhombic inner region. Orthonormal basis-sets exists for those two regular shapes based on spherical harmonics or cartesian Legendre polynomials. More generally, the formalism could also support nonorthogonal for simulating an inner region of arbitrary shape. Comparison with numerical finite-difference PB calculations demonstrates that GSBP can accurately describe all long-range electrostatic interactions.

Several extensions to the current implementation of GSBP may be considered. For example, the calculation of the static field could easily be extended to incorporate the influence of an external transmembrane voltage in simulation of ion channels by using a modified PB equation developed recently.[40] Furthermore, we have assumed a frozen configuration of the atoms in the outer region in the current implementation. Nonetheless, it would be desirable to incorporate the influence of dynamical fluctuations of the outer atoms. One possible approximation would be to assign a dielectric constant greater than 1 to the interior of the macromolecule in the PB calculations. The boundary potential would then be $\overline{\mathcal{W}}(\mathbf{X}_i) \approx \mathcal{W}(\mathbf{X}_i, \overline{\mathbf{X}}_o; \epsilon_m)$, where $\overline{\mathbf{X}}_o$ is the average equilibrium position of the outer atoms of the macromolecule, and ϵ_m is an appropriately chosen value for the dielectric constant of the macromolecule. More sophisticated approaches could also be obtained from Eq. (3) using a cumulant expansion with a combination of normal modes or gaussian quasiharmonic fluctuation analysis.[48,49] Lastly, it is also our hope that GSBP will be useful for incorporating the influence of the surrounding environment on approaches combining quantum mecanical and molecular mechanical description (QM/MM) of biomolecular reactions.[50] Future applications will explore these avenues and test GSBP in the context of free energy calculations.

APPENDIX

1. Basis set in a spherical inner region

The orthonormal basis functions for the components of the lth multipole charge distribution inside a spherical region of radius R of interest are defined as[51]

$$b_{lm}(\mathbf{r}) = \left[\frac{(2l+3)}{R^{2l+3}}\right]^{1/2} r^l Y_{lm}(\theta, \phi), \tag{A1}$$

where Y_{lm} is the spherical harmonics which, in terms of associated Legendre functions, P_l^m, is given by

$$Y_{lm}(\theta, \phi) = (-1)^m \left[\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}\right]^{1/2} P_l^m(\cos\theta)e^{im\phi}. \tag{A2}$$

Through the linear combination of Y_{lm} and $Y_{l,-m}$, Eq. (A1) becomes

$$b_{lm}(\mathbf{r})=\left[\frac{(2l+1)(2l+3)}{2\pi(1+\delta_{m0})R^{2l+3}}\frac{(l-m)!}{(l+m)!}\right]^{1/2}r^l$$

$$\times\cos m\phi P_l^m(\cos\theta)\qquad(A3)$$

and

$$b_{l,-m}(\mathbf{r})=\left[\frac{(2l+1)(2l+3)}{2\pi R^{2l+3}}\frac{(l-m)!}{(l+m)!}\right]^{1/2}r^l$$

$$\times\sin m\phi P_l^m(\cos\theta),\qquad(A4)$$

where m goes from 0 to l.

The charge distribution in the spherical inner region can be expressed as

$$\rho(\mathbf{r})=\sum_{l=0}^{L}\sum_{m=-l}^{l}c_{lm}b_{lm}(\mathbf{r}),\qquad(A5)$$

where L is the upper limit of the spherical multipole and the coefficients c_{lm} are equivalent to the multipole moments Q_{lm} defined in Eq. (17) because the basis functions are orthonormal. For the force calculations, the first derivatives of Eqs. (A3) and (A4) are taken with respect to the position of atom α in the spherical coordinate,

$$\nabla_\alpha b_{lm}=N_{lm}\left[\hat{\mathbf{r}}lr_\alpha^{l-1}\cos m\phi_\alpha P_l^m\right.$$

$$-\hat{\theta}r_\alpha^{l-1}\sin\theta_\alpha\cos m\phi_\alpha P_l^{m'}$$

$$\left.-\hat{\phi}r_\alpha^{l-1}\frac{m}{\sin\theta_\alpha}\sin m\phi_\alpha P_l^m\right]\qquad(A6)$$

and

$$\nabla_\alpha b_{l,-m}=N_{l,-m}\left[\hat{\mathbf{r}}lr_\alpha^{l-1}\sin m\phi_\alpha P_l^m\right.$$

$$-\hat{\theta}r_\alpha^{l-1}\sin\theta_\alpha\sin m\phi_\alpha P_l^{m'}$$

$$\left.+\hat{\phi}r_\alpha^{l-1}\frac{m}{\sin\theta_\alpha}\cos m\phi_\alpha P_l^m\right],\qquad(A7)$$

where N_{lm} and $N_{l,-m}$ are the normalization constant given in Eqs. (A3) and (A4), respectively, $(r_\alpha,\theta_\alpha,\phi_\alpha)$ is the spherical coordinate of atom α,

$$\nabla_\alpha=\hat{\mathbf{r}}\frac{\partial}{\partial r_\alpha}+\hat{\theta}\frac{1}{r_\alpha}\frac{\partial}{\partial\theta_\alpha}+\hat{\phi}\frac{1}{r_a\sin\theta_\alpha}\frac{\partial}{\partial\phi_\alpha},\qquad(A8)$$

and $P_l^{m'}$ is the first derivatives of P_l^m which, through the recurrence relation,[52] is given

$$(1-x^2)^{1/2}P_l^{m'}(x)=P_l^{m+1}(x)-x(1-x^2)^{-(1/2)}P_l^m(x),\qquad(A9)$$

where $x=\cos\theta_\alpha$. The transformation matrix to the Cartesian coordinate is

$$\begin{pmatrix}\sin\theta_\alpha\cos\phi_\alpha & \cos\theta_\alpha\cos\phi_\alpha & -\sin\phi_\alpha\\ \sin\theta_\alpha\sin\phi_\alpha & \cos\theta_\alpha\sin\phi_\alpha & -\cos\phi_\alpha\\ \cos\theta_\alpha & -\sin\theta_\alpha & 0\end{pmatrix}.$$

2. Basis set in a orthorhombic inner region

The orthonormal basis functions for a variety of multipole charge distributions inside a orthorhombic inner region are defined as the products of the Legendre polynomials, $\{P_l\}$, in X, Y, and Z directions.

$$b_{lmn}(\mathbf{r})=\left[\frac{(2l+1)}{L_x}\frac{(2m+1)}{L_y}\frac{(2n+1)}{L_z}\right]^{1/2}P_l\left(\frac{2}{L_x}x\right)$$

$$\times P_m\left(\frac{2}{L_y}y\right)P_n\left(\frac{2}{L_z}z\right),\qquad(A10)$$

where L_x, L_y, and L_z are the lengths of the orthorhombic inner region in each direction. In this expression x, y, and z are scaled because the Legendre polynomials form an orthonormal basis set between -1 and 1.

The charge distribution in the orthorhombic inner region can be expressed as

$$\rho(\mathbf{r})=\sum_{l=0}^{L}\sum_{m=0}^{M}\sum_{n=0}^{N}c_{lmn}b_{lmn}(\mathbf{r}),\qquad(A11)$$

where L, M, and N are the upper limits of the Legendre Polynomials in each direction and the coefficients c_{lmn} are equivalent to the multipole moments Q_{lmn}. For the force calculations, the first derivatives of Eq. (A10) are directly taken with respect to the position of atom α in each direction.

[1] T. P. Straatsma, H. J. C. Berendsen, and J. P. M. Postma, J. Chem. Phys. **89**, 5876 (1988).
[2] J. S. Bader and D. Chandler, J. Phys. Chem. **96**, 6423 (1992).
[3] P. H. Hunenberger and J. A. McCammon, J. Chem. Phys. **110**, 1856 (1999).
[4] P. H. Hunenberger and J. A. McCammon, Biophys. Chem. **78**, 69 (1999).
[5] H. L. Friedman, Mol. Phys. **29**, 1533 (1975).
[6] M. Berkowitz and J. A. McCammon, Chem. Phys. Lett. **90**, 215 (1982).
[7] C. L. Brooks III and M. Karplus, J. Chem. Phys. **79**, 6312 (1983).
[8] A. Brunger, C. L. Brooks III, and M. Karplus, Chem. Phys. Lett. **105**, 495 (1984).
[9] A. Brunger, C. L. Brooks III, and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **82**, 8458 (1985).
[10] J. W. Essex and W. L. Jorgensen, J. Comput. Chem. **16**, 951 (1995).
[11] A. Warshel and G. King, Chem. Phys. Lett. **121**, 124 (1985).
[12] G. King and A. Warshel, J. Chem. Phys. **91**, 3647 (1989).
[13] J. A. C. Rullmann and P. Th. van Duijnen, Mol. Phys. **61**, 293 (1987).
[14] H. Alper and R. Levy, J. Chem. Phys. **99**, 9847 (1993).
[15] L. Wang and J. Hermans, J. Phys. Chem. **99**, 12001 (1995).
[16] D. Beglov and B. Roux, J. Chem. Phys. **100**, 9050 (1994).
[17] B. Roux, D. Beglov, and W. Im, in *Treatments of Electrostatics Interactions in Computer Simulations of Condensed Media*, edited by L. R. Pratt and G. Hummer (Santa Fe Workshop, California, 1999), p. 473.
[18] G. Basu, A. Kitao, A. Kuki, and N. Go, J. Phys. Chem. B **102**, 2076 (1998).
[19] G. Basu, A. Kitao, A. Kuki, and N. Go, J. Phys. Chem. B **102**, 2085 (1998).
[20] B. Roux, Biophys. J. **71**, 2076 (1996).
[21] M. Nina, D. Beglov, and B. Roux, J. Phys. Chem. B **101**, 5239 (1997).
[22] Y. Sugita and A. Kitao, Proteins **30**, 388 (1988).
[23] Y. Sugita and A. Kitao, Biophys. J. **75**, 2178 (1988).
[24] Y. Sugita, A. Kitao, and N. Go, Folding Des. **27**, 173 (1988).
[25] A. Kitao, S. Hayward, and N. Go, Proteins **33**, 496 (1988).
[26] D. Mohanty, R. Elber, D. Thirumalai, D. Beglov, and B. Roux, J. Mol. Biol. **272**, 423 (1997).
[27] S. Marchand and B. Roux, Proteins **33**, 265 (1988).
[28] J. G. Kirkwood, J. Chem. Phys. **2**, 351 (1934).
[29] T. Simonson, G. Archontis, and M. Karplus, J. Phys. Chem. B **101**, 8349 (1997).

[30] G. Archontis, T. Simonson, D. Moras, and M. Karplus, J. Mol. Biol. **275**, 823 (1998).

[31] T. Simonson, J. Phys. Chem. B **104**, 6509 (2000).

[32] D. A. Doyle, J. M. Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon, Science **280**, 69 (1998).

[33] A. D. Mackerell Jr., D. Bashford, M. Bellot *et al.*, J. Phys. Chem. B **102**, 3586 (1998).

[34] K. A. Sharp and B. Honig, Annu. Rev. Biophys. Biophys. Chem. **19**, 301 (1990).

[35] M. K. Gilson and B. Honig, Proteins **4**, 7 (1988).

[36] I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig, Proteins **1**, 47 (1986).

[37] A. Nicholls and B. Honig, J. Comput. Chem. **12**, 435 (1991).

[38] W. Im, D. Beglov, and B. Roux, Comput. Phys. Commun. **111**, 59 (1998).

[39] W. Im, S. Seefeld, and B. Roux, Biophys. J. **79**, 788 (2000).

[40] B. Roux, Biophys. J. **73**, 2980 (1997).

[41] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).

[42] M. K. Gilson, K. Sharp, and B. Honig, J. Comput. Chem. **9**, 327 (1987).

[43] S. Berneche and B. Roux, Biophys. J. **78**, 2900 (2000).

[44] M. Born, Z. Phys. **1**, 45 (1920).

[45] M. K. Gilson, M. E. Davis, B. A. Luty, and J. A. McCammon, J. Phys. Chem. **97**, 3591 (1993).

[46] R. H. Stote, D. J. States, and M. Karplus, J. Chim. Phys. **88**, 2419 (1991).

[47] L. David and M. J. Field, J. Comput. Chem. **18**, 343 (1997).

[48] T. Simonson, D. Perahia, and A. T. Brunger, Biophys. J. **59**, 670 (1991).

[49] T. Simonson and D. Perahia, Biophys. J. **61**, 410 (1992).

[50] G. Monard and K. M. Merz Jr., Acc. Chem. Res. **32**, 904 (1999).

[51] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1962).

[52] W. R. Smythe, *Static and Dynamic Electricity* (McGraw–HIll, New York, 1968).