

## New and Efficient Poisson–Boltzmann Solver for Interaction of Multiple Proteins

Eng-Hui Yap\* and Teresa Head-Gordon\*

*Department of Bioengineering, University of California, Berkeley, Berkeley, California 94720, and UCSF and UC Berkeley Joint Graduate Group in Bioengineering*

Received March 17, 2010

**Abstract:** We derive a new numerical approach to solving the linearized Poisson–Boltzmann equation (PBE) by representing the protein surface as a collection of spheres in which the surface charges can then be iteratively solved by new analytical multipole methods previously introduced by us [Lotan, I.; Head-Gordon, T. *J. Chem. Theory Comput.* **2006**, *2*, 541]. We show that our Poisson–Boltzmann semianalytical method, PB-SAM, realizes better accuracy, more flexible memory management, and at reduced cost relative to either finite difference or boundary element method PBE solvers. We provide two new benchmarks of PBE solution accuracy to test the numerical PBE solutions based on (1) arrays of up to hundreds of spherical low dielectric geometries with asymmetric charges in which mutual polarization is treated exactly and (2) two overlapping spheres with increasing charge asymmetry by solving the PB-SAM method to very high pole order. We illustrate the strength of the PB-SAM approach by computing the potential profile of an array of 60 T1-particle forming monomers of the bromine mosaic virus.

### Introduction

The formation of protein complexes is ubiquitous in a crowded, salty cellular environment. Since electrostatic forces dominate the earliest of protein–protein recognition events in the cell, various analytical and numerical continuum theories of bulk electrolytes have been adapted for use to describe protein complexation mechanisms on the supramolecular scale.<sup>2</sup> One popular continuum mean-field theory is the Poisson–Boltzmann (PB) treatment, which forms the basis of Gouy–Chapman theory<sup>3,4</sup> in electrochemistry, and under the low field linearized PB (LPB) approximation, the Debye–Hückel theory in solution chemistry<sup>5</sup> and Derjaguin–Landau–Verwey–Overbeek (DLVO) theory in colloid chemistry.<sup>6,7</sup> Numerous techniques for solving the PB equation exist,<sup>1,6</sup> including both analytical or numerical methods, and each has its drawbacks and its strengths.

Analytical methods typically allow rapid solution of the PB equation using multipole expansions under specialized geometries, such as spheres or cylinders. A complete PB

solution comprising one spherical macromolecule was developed by Kirkwood<sup>8</sup> more than 70 years ago, but generalization of this complete solution to two or more spherical macromolecules proved to be more difficult, and many different partial and approximate solutions have been proposed.<sup>9–12</sup> We have recently achieved a fundamental result in deriving an analytical PB solution for computing the screened electrostatic interaction between *arbitrary* numbers of spherical proteins of *arbitrarily* complex charge distributions, separated by *arbitrary* distance.<sup>1</sup> While such idealized protein geometries will typically be inappropriate for describing complexation on a supermolecular scale, this new analytical solution is a novel component of our new numerical PB solver for arbitrary protein shape. It also serves as a benchmark for the accuracy of the numerical solutions in certain idealized test cases.

By contrast, numerical methods (see ref 13 for a recent survey), such as finite-difference (FD)<sup>14–19</sup> and finite-element (FE)<sup>20–22</sup> methods, can handle arbitrary dielectric boundaries by solving for the PB potential on a 3-D grid or mesh. However, there are limitations of the FE or FD formulations, such as singularities in the potential solution because of point

\* To whom correspondence should be addressed. E-mail address: enghui@berkeley.edu (E.-H.Y.); tlhead-gordon@lbl.gov (T.H.-G.).

charges, that electric displacement continuity could not be enforced across dielectric boundaries (thereby reducing the solution accuracy and convergence rate), and forces must be estimated from finite-difference calculations.<sup>13</sup> But most importantly, the requirement that the solution be solved on a grid limits its practical application to spatial domains of either two to three typical macromolecules at reasonably high resolution ( $\sim 0.2$  Å) or to larger numbers of macromolecules with greatly diminished resolution and thus solution accuracy. For example, the PBE solution for an assembled 50S ribosomal subunit has been evaluated at 0.45 Å resolution,<sup>14</sup> at the limit of machine memory, but to describe the preceding assembly process that occur over much larger spatial distances, the spatial resolution and consequently the solution accuracy would greatly deteriorate. As such, computational and memory cost in FD and FE methods are strictly functions of the number of grid points and not of the number of macromolecules described.

Boundary element (BE) methods<sup>22–29</sup> are an attractive alternative since they satisfy both the Dirichlet and von Neuman boundary conditions by construction, singular charges can be correctly treated, and most importantly the 2D solutions on the macromolecular surface removes spatial resolution limitations imposed by the 3D grid of the FD or FE solvers. However increasing the number of boundary surface element results in an increasingly large dense matrix to be solved with severe memory requirements, a problem which scales with the number of macromolecules. Recent acceleration of the BE approach<sup>24,26</sup> by incorporating fast multipole methods have rendered BE computational times comparable to state-of-the-art software packages like the Adaptive Poisson–Boltzmann Solver (APBS)<sup>14</sup> based on FD solutions.

In this work, we derive a new numerical approach to solving the PB equation by combining the advantages of both the boundary element and our analytical model<sup>1</sup> formalism. In particular, we replace the discretization of the molecule surface into a large number (tens of thousands) of boundary elements, by a discretization involving a smaller number (tens to hundreds) of spheres. The surface charges can then be iteratively solved using analytical multipole methods.<sup>1</sup> We show that our Poisson–Boltzmann semianalytical method, PB-SAM, converges to the analytical solution with better accuracy and at greatly reduced cost relative to the readily available public domain PB solver APBS.<sup>14</sup> Furthermore, we define a high-quality benchmark using 140 poles to describe the electrostatic potential for two overlapping spheres that are models for the sharp features that are sometimes present in real protein geometries, in which we show that our PB-SAM solution converges to the correct solution with the same computational cost or better than the finite difference solution. Finally we illustrate the strength of the PB-SAM approach by computing the potential profile of an array of 60 T1-particle-forming monomers of the bromine mosaic virus (PDB code 1YC6).

## Theory

**Mathematical Preliminaries.** Our theory makes extensive use of the spherical harmonics (SH) family of functions. The

spherical harmonic function of order  $n$  and degree  $m$ , at polar angle  $\theta$  and azimuthal angle  $\phi$ , is defined per the convention from Gumerov and Duraiswami<sup>30</sup> as

$$Y_{nm}(\theta, \phi) = (-1)^m \sqrt{\frac{(n - |m|)!}{(n + |m|)!}} P_{n|m|}(\cos \theta) e^{im\phi} \quad (1)$$

where  $P_{nm}(x)$  is the *associated Legendre polynomial*. Note that this definition of  $Y_{nm}(\theta, \phi)$  differs from the common convention by a  $\sqrt{(2n + 1)/4\pi}$  factor. The complex conjugate of  $Y_{nm}(\theta, \phi)$  will be denoted as  $\overline{Y}_{nm}(\theta, \phi)$ .

We shall utilize two important properties of spherical harmonics: their addition theorems and orthogonality. Let  $\mathbf{r}_1 = [r_1, \theta_1, \phi_1]$  and  $\mathbf{r}_2 = [r_2, \theta_2, \phi_2]$  be two points in 3D space specified by spherical coordinates, where  $r_2 > r_1$ . The Euclidean distance  $|\mathbf{r}_1 - \mathbf{r}_2|$  between them then obeys the addition theorems:<sup>24,31</sup>

$$\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} = \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{r_1^n}{r_2^{n+1}} \overline{Y}_{nm}(\theta_1, \phi_1) Y_{nm}(\theta_2, \phi_2) \quad (2a)$$

and for the screened Yukawa potential eq 2a is modified to read as

$$\begin{aligned} \frac{e^{-\kappa|\mathbf{r}_1 - \mathbf{r}_2|}}{|\mathbf{r}_1 - \mathbf{r}_2|} = \\ \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{r_1^n}{r_2^{n+1}} \hat{i}_n(\kappa r_1) e^{-\kappa r_2} \hat{k}_n(\kappa r_2) \overline{Y}_{nm}(\theta_1, \phi_1) Y_{nm}(\theta_2, \phi_2) \end{aligned} \quad (2b)$$

where  $\kappa$  is the inverse Debye–Huckel screening length (described later), and  $\hat{k}_n(z)$  and  $\hat{i}_n(z)$  are *adapted modified spherical Bessel functions* defined as

$$\hat{k}_n(z) = \sqrt{\frac{2}{\pi}} \frac{e^{z z^{n+1/2}}}{(2n - 1)!!} K_{n+1/2}(z) \quad (3a)$$

$$\hat{i}_n(z) = \sqrt{\frac{\pi}{2}} \frac{(2n + 1)!!}{z^{n+1/2}} I_{n+1/2}(z) \quad (3b)$$

$I_n(z)$  and  $K_n(z)$  are the *modified Bessel functions of the first and second kind*, respectively. Detailed properties of  $\hat{k}_n(z)$  and  $\hat{i}_n(z)$  have been described in ref 1.

The spherical harmonic functions are also orthogonal over the surface of a unit sphere ( $S_1$ )

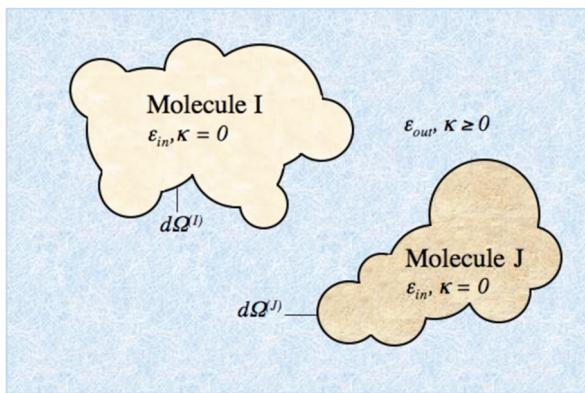
$$\int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} Y_{ls}(\theta, \phi) \overline{Y}_{nm}(\theta, \phi) \sin \theta d\theta d\phi = \frac{4\pi}{2n + 1} \delta_{nl} \delta_{ms} \quad (4a)$$

Hence a square-integrable function  $g(\theta, \phi)$  on  $S_1$  can be expanded using  $\{Y_{nm}\}$  as the basis set

$$g(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{2n + 1}{4\pi} G_{nm} Y_{nm}(\theta, \phi) \quad (4b)$$

with the coefficients  $G_{nm}$  determined through the reciprocal transform

$$G_{nm} = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} g(\theta', \phi') \overline{Y}_{nm}(\theta', \phi') \sin \theta' d\theta' d\phi' \quad (4c)$$



**Figure 1.** Setting up the boundary value problem. The example system is comprised of two proteins with arbitrary charge distribution, each represented as a collection of overlapping spheres to describe an arbitrarily shaped dielectric boundary containing no salt, immersed in a high dielectric salty continuum solvent. Salt screening effects are captured via the Debye–Hückel parameter  $\kappa$ .

**Setting up the Boundary Value Problem.** We seek to set up a boundary value problem for a system of  $N_{\text{mol}}$  macromolecules immersed in an implicit aqueous salty solvent. Figure 1 gives an example of the spatial domain for which we solve the linearized PB equation (LPBE). Each macromolecule  $I$  is embedded with  $N_C^{(I)}$  fixed partial charge and represented as a collection of  $N_S^{(I)}$  overlapping spheres with dielectric constant  $\epsilon_{\text{in}}$ . For simplicity, we consider in this paper the same  $\epsilon_{\text{in}}$  for all molecules, but the model can handle different dielectric constants. The solvent is treated as a continuum with dielectric constant  $\epsilon_{\text{out}}$ , with screening effects because of mobile ions captured via the inverse Debye length  $\kappa$ . The LPBE gives the potential  $\Phi$  at any point  $\mathbf{r}$  in space  $\mathcal{R}^3$  as

$$-\nabla[\epsilon(\mathbf{r})\nabla\Phi(\mathbf{r})] + \kappa^2\Phi(\mathbf{r}) = 4\pi\rho_{\text{fixed}}(\mathbf{r}) \quad (5)$$

where  $\epsilon$  is the relative dielectric function,  $\rho_{\text{fixed}}$  is the charge density from the fixed protein partial charges, and  $\kappa = \sqrt{8\pi\bar{n}e^2/\epsilon_{\text{out}}k_B T}$ , where  $\bar{n}$  is the bulk concentration of monovalent salt in the solution,  $e$  is the fundamental electronic charge,  $k_B$  the Boltzmann constant, and  $T$  the absolute temperature. Inside each macromolecule  $I$ , the potential  $\Phi_{\text{in}}^{(I)}(\mathbf{r})$  satisfies the Poisson equation

$$-\nabla^2\Phi_{\text{in}}^{(I)}(\mathbf{r}) = \rho_{\text{fixed}}^{(I)}(\mathbf{r})/\epsilon_{\text{in}} \quad (6a)$$

while in the region outside all macromolecules, the potential  $\Phi_{\text{out}}(\mathbf{r})$  satisfies the Helmholtz equation

$$\nabla^2\Phi_{\text{out}}(\mathbf{r}) - \kappa^2\Phi_{\text{out}}(\mathbf{r}) = 0 \quad (6b)$$

We first express the potential  $\Phi_{\text{in}}^{(I)}(\mathbf{r})$  anywhere inside molecule  $I$  as the sum of the potentials because of the embedded fixed charges and a single-layer of yet unknown reaction charges  $f^{(I)}(\mathbf{r})$  on the surface  $d\Omega^{(I)}$ <sup>23,32</sup>

$$\Phi_{\text{in}}^{(I)}(\mathbf{r}) = \sum_{\alpha=1}^{N_C^{(I)}} \frac{1}{|\mathbf{r} - \mathbf{r}_{\alpha}^{(I)}|} \frac{q_{\alpha}^{(I)}}{\epsilon_{\text{in}}} + \frac{1}{4\pi} \int_{d\Omega^{(I)}} \frac{1}{|\mathbf{r} - \mathbf{r}'|} f^{(I)}(\mathbf{r}') d\mathbf{r}' \quad (7)$$

In our new approach, the surface of molecule  $I$  is discretized into  $N_S^{(I)}$  spheres. We consider each sphere  $k$  of molecule  $I$  of radius  $a^{(I,k)}$  in turn, and all position vectors and coefficients are defined with the center of sphere  $k$  as the origin. We apply the first addition theorem (eq 2a) to eq 7 to obtain

$$\begin{aligned} \Phi_{\text{in}}^{(I,k)}(\mathbf{r}) = & \sum_{n=0}^{\infty} \sum_{m=-n}^n \left( \frac{E_{nm}^{(I,k)}}{r} \left( \frac{a^{(I,k)}}{r} \right)^n + \right. \\ & \left( \frac{r}{a^{(I,k)}} \right)^n \text{LE}_{nm}^{(I,k)}(\theta, \phi) + \sum_{n=0}^{\infty} \sum_{m=-n}^n \left( \left( \frac{r}{a^{(I,k)}} \right)^n (\text{LF}_{nm}^{(I,k)} + \right. \right. \\ & \left. \left. \text{LFS}_{nm}^{(I,k)}) \right) Y_{nm}^{(I,k)}(\theta, \phi) \end{aligned} \quad (8)$$

with the coefficients defined as

$$E_{nm}^{(I,k)} \equiv \sum_{\alpha=1}^{N_C^{(I,k)}} \frac{q_{\alpha}}{\epsilon_{\text{in}}} \left( \frac{r_{\alpha}}{a^{(I,k)}} \right)^n \overline{Y_{nm}^{(I,k)}}(\theta_{\alpha}, \phi_{\alpha}) \quad (8a)$$

$$\text{LE}_{nm}^{(I,k)} \equiv \sum_{\alpha=1}^{\bar{N}_C^{(I,k)}} \frac{q_{\alpha}}{\epsilon_{\text{in}} r_{\alpha}} \left( \frac{a^{(I,k)}}{r_{\alpha}} \right)^n \overline{Y_{nm}^{(I,k)}}(\theta_{\alpha}, \phi_{\alpha}) \quad (8b)$$

$$\text{LF}_{nm}^{(I,k)} \equiv \frac{1}{4\pi} \int_{d\Omega^{(I,k)}} \frac{f^{(I,k)}(\mathbf{r}')}{r'} \left( \frac{a^{(I,k)}}{r'} \right)^n \overline{Y_{nm}^{(I,k)}}(\theta', \phi') d\mathbf{r}' \quad (8c)$$

$$\text{LFS}_{nm}^{(I,k)} \equiv \frac{1}{4\pi} \int_{d\Omega^{(I,k)}} \frac{f^{(I,k)}(\mathbf{r}')}{r'} \left( \frac{a^{(I,k)}}{r'} \right)^n \overline{Y_{nm}^{(I,k)}}(\theta', \phi') d\mathbf{r}' \quad (8d)$$

Notice that we have scaled the terms with  $r_{\alpha}^n$  and  $r_{\alpha}^{n+1}$  dependence by  $(a^{(I,k)})^n$  and  $(a^{(I,k)})^{-n}$ , respectively. This is to avoid machine imprecision as  $n$  becomes large. Coefficients with  $(r_{\alpha}/a^{(I,k)})^n$  dependence, such as  $E_{nm}^{(I,k)}$ , are known as multipole (external) coefficients, while those with  $a^{(I,k)n}/r_{\alpha}^{n+1}$  dependence ( $\text{LE}_{nm}^{(I,k)}$ ,  $\text{LF}_{nm}^{(I,k)}$ , and  $\text{LFS}_{nm}^{(I,k)}$ ) are known as Taylor (local) coefficients. The first sum in eq 8 represents the potential due to fixed charges, where  $E_{nm}^{(I,k)}$  sums over  $N_C^{(I,k)}$  fixed charges *inside* sphere  $k$  of molecule  $I$ , while  $\text{LE}_{nm}^{(I,k)}$  sums over the remaining  $\bar{N}_C^{(I,k)}$  fixed charges *outside* sphere  $k$ . The second sum in eq 8 gives the potential resulting from the unknown surface charge  $f^{(I)}(\mathbf{r})$ ;  $\text{LFS}_{nm}^{(I,k)}$  and  $\text{LF}_{nm}^{(I,k)}$  account for reactive charges on sphere  $k$  and on other spheres in molecule  $I$ , respectively.

In the solvent region outside the molecules, the potential  $\Phi_{\text{out}}(\mathbf{r})$  can be represented as the sum of Yukawa potentials because of each molecule's yet unknown effective surface charges  $h^{(I)}(\mathbf{r})$ <sup>23,32</sup>

$$\Phi_{\text{out}}(\mathbf{r}) = \sum_{I=1}^{N_{\text{mol}}} \left( \frac{1}{4\pi} \int_{d\Omega^{(I)}} \frac{e^{-\kappa|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|} h^{(I)}(\mathbf{r}') d\mathbf{r}' \right) \quad (9)$$

The above equation valid for the *exposed* portion of sphere  $k$  of molecule  $I$ . Applying addition theorem 2 (eq 2b) to eq 9, the potential on the exposed surface can be expressed as

$$\Phi_{\text{out}}^{(I,k)}(\mathbf{r}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \left( \frac{H_{nm}^{(I,k)}}{r} \left( \frac{a^{(I,k)}}{r} \right)^n e^{-\kappa r} \hat{l}_n(\kappa r) + \left( \frac{r}{a^{(I,k)}} \right)^n \hat{l}_n(\kappa r) (\text{LH}_{nm}^{(I,k)} + \text{LHN}_{nm}^{(I,k)}) \right) Y_{nm}^{(I,k)}(\theta, \phi) \quad (10)$$

where the coefficients are defined as

$$H_{nm}^{(I,k)} \equiv \frac{1}{4\pi} \int_{d\Omega^{(i,k)}} h^{(I,k)}(\mathbf{r}') \left( \frac{r'}{a^{(I,k)}} \right)^n \hat{l}_n(\kappa r') \overline{Y_{nm}^{(I,k)}}(\theta', \phi') d\mathbf{r}' \quad (10a)$$

$$\text{LH}_{nm}^{(I,k)} \equiv \frac{1}{4\pi} \int_{d\Omega^{(I,k)}} \frac{h^{(I,k)}(\mathbf{r}')}{r'} \left( \frac{a^{(I,k)}}{r'} \right)^n e^{-\kappa r'} \hat{k}_n(\kappa r') \overline{Y_{nm}^{(I,k)}}(\theta', \phi') d\mathbf{r}' \quad (10b)$$

$$\text{LHN}_{nm}^{(I,k)} \equiv \sum_{J \neq I} \sum_{l=1}^{N_{\text{mol}}} \left( \frac{1}{4\pi} \int_{d\Omega^{(J,l)}} \frac{h^{(J,l)}(\mathbf{r}')}{r'} \times \left( \frac{a^{(I,k)}}{r'} \right)^n e^{-\kappa r'} \hat{k}_n(\kappa r') \overline{Y_{nm}^{(I,k)}}(\theta', \phi') d\mathbf{r}' \right) \quad (10c)$$

The multipole coefficient  $H_{nm}^{(I,k)}$  represents effective polarization charges on sphere  $k$  of molecule  $l$ 's exposed surface. The local coefficients  $\text{LH}_{nm}^{(I,k)}$  and  $\text{LHN}_{nm}^{(I,k)}$  represent effective polarization charges on other spheres in molecule  $I$  and on other molecules, respectively.

With eqs 8 and 10 in hand, we can impose boundary conditions at the dielectric boundary surface  $\mathbf{r}_E = (a^{(I,k)}, \theta_E, \phi_E) \in d\Omega_E^{(I,k)}$  between each sphere  $k$  in molecule  $I$  exposed to solvent:

$$\Phi_{\text{in}}^{(I,k)}(\mathbf{r}_E) = \Phi_{\text{out}}^{(I,k)}(\mathbf{r}_E) \quad (11a)$$

$$\varepsilon \frac{d\Phi_{\text{in}}^{(I,k)}}{dn} \Big|_{\mathbf{r}_E} = \frac{d\Phi_{\text{out}}^{(I,k)}}{dn} \Big|_{\mathbf{r}_E}, \quad \varepsilon = \varepsilon_{\text{in}}/\varepsilon_{\text{out}} \quad (11b)$$

The Dirichlet boundary condition (eq 11a) enforces potential continuity across the boundary

$$\sum_{n=0}^{\infty} \sum_{m=-n}^n (E_{nm}^{(I,k)} + a^{(I,k)}(\text{LE}_{nm}^{(I,k)} + \text{LF}_{nm}^{(I,k)} + \text{LFS}_{nm}^{(I,k)})) Y_{nm}^{(I,k)}(\theta_E, \phi_E) = \sum_{n=0}^{\infty} \sum_{m=-n}^n (H_{nm}^{(I,k)} e^{-\kappa a^{(I,k)}} \hat{k}_n(\kappa a^{(I,k)}) + a^{(I,k)} \hat{l}_n(\kappa a^{(I,k)}) (\text{LH}_{nm}^{(I,k)} + \text{LHN}_{nm}^{(I,k)})) Y_{nm}^{(I,k)}(\theta_E, \phi_E) \quad (12a)$$

while the von Neumann boundary condition (eq 11b) enforces electric displacement continuity

$$\begin{aligned} \varepsilon \sum_{n=0}^{\infty} \sum_{m=-n}^n (-(n+1)E_{nm}^{(I,k)} + nF_{nm}^{(I,k)} + na^{(I,k)}(\text{LE}_{nm}^{(I,k)} + \text{LF}_{nm}^{(I,k)})) Y_{nm}^{(I,k)}(\theta_E, \phi_E) &= \sum_{n=0}^{\infty} \sum_{m=-n}^n \times \\ &\left( H_{nm}^{(I,k)} e^{-\kappa a^{(I,k)}} [n\hat{k}_n(\kappa a^{(I,k)}) - (2n+1)\hat{k}_{n+1}(\kappa a^{(I,k)})] + \right. \\ &\left. a^{(I,k)} \left[ n\hat{l}_n(\kappa a^{(I,k)}) + \frac{(\kappa a^{(I,k)})^2 \hat{l}_{n+1}(\kappa a^{(I,k)})}{2n+3} \right] \times \right. \\ &\left. (\text{LH}_{nm}^{(I,k)} + \text{LHN}_{nm}^{(I,k)}) \right) Y_{nm}^{(I,k)}(\theta_E, \phi_E) \end{aligned} \quad (12b)$$

We have introduced  $F_{nm}^{(I,k)} \equiv a^{(I,k)} \text{LFS}_{nm}^{(I,k)}$ . We continue to simplify eqs 12a and 12b by rearranging

$$\begin{aligned} \sum_{n=0}^{\infty} \sum_{m=-n}^n (-H_{nm}^{(I,k)} e^{-\kappa a^{(I,k)}} \hat{k}_n(\kappa a^{(I,k)}) + F_{nm}^{(I,k)} + X\text{H}_{nm}^{(I,k)}) Y_{nm}^{(I,k)}(\theta_E, \phi_E) &= 0 \quad (13a) \end{aligned}$$

$$\begin{aligned} \sum_{n=0}^{\infty} \sum_{m=-n}^n (e^{-\kappa a^{(I,k)}} [n\hat{k}_n(\kappa a^{(I,k)}) - (2n+1)\hat{k}_{n+1}(\kappa a^{(I,k)})] H_{nm}^{(I,k)} - nE_{nm}^{(I,k)} + X\text{F}_{nm}^{(I,k)}) Y_{nm}^{(I,k)}(\theta_E, \phi_E) &= 0 \quad (13b) \end{aligned}$$

where

$$\begin{aligned} X\text{H}_{nm}^{(I,k)} &\equiv E_{nm}^{(I,k)} + a^{(I,k)}(\text{LE}_{nm}^{(I,k)} + \text{LF}_{nm}^{(I,k)}) - \\ &a^{(I,k)} \hat{l}_n(\kappa a^{(I,k)}) (\text{LH}_{nm}^{(I,k)} + \text{LHN}_{nm}^{(I,k)}) \quad (14a) \end{aligned}$$

$$\begin{aligned} X\text{F}_{nm}^{(I,k)} &\equiv a^{(I,k)} \left[ n\hat{l}_n(\kappa a^{(I,k)}) + \frac{(\kappa a^{(I,k)})^2 \hat{l}_{n+1}(\kappa a^{(I,k)})}{2n+3} \right] (\text{LH}_{nm}^{(I,k)} + \\ &\text{LHN}_{nm}^{(I,k)}) + (n+1)\varepsilon E_{nm}^{(I,k)} - nea^{(I,k)}(\text{LE}_{nm}^{(I,k)} + \text{LF}_{nm}^{(I,k)}) \quad (14b) \end{aligned}$$

The boundary equations above are valid on the solvent-exposed surfaces of sphere  $k$  on molecule  $I$ . We need another set of boundary equations on the buried surface  $\mathbf{r}_B = [a^{(I,k)}, \theta_B, \phi_B] \in d\Omega_B^{(I,k)}$ . We shall utilize the fact that there is no polarization charge on the buried surface, that is,  $f^{(I,k)}(\mathbf{r}_B) = h^{(I,k)}(\mathbf{r}_B) = 0$ , since there is no dielectric discontinuity. It follows that scaled versions of the charge distributions,  $\tilde{f}^{(I,k)}(\theta, \phi) \equiv (a^{(I,k)})^2 f^{(I,k)}(a^{(I,k)}, \theta, \phi)$  and  $\tilde{h}^{(I,k)}(\theta, \phi) \equiv (a^{(I,k)})^2 h^{(I,k)}(a^{(I,k)}, \theta, \phi)$ , are also zero on the buried surface. Separately, we can express  $\tilde{f}^{(I,k)}$  and  $\tilde{h}^{(I,k)}$  in terms of  $F_{nm}^{(I,k)}$  and  $H_{nm}^{(I,k)}$  using eqs 4c, 8d, and 10a

$$\tilde{f}^{(I,k)}(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{2n+1}{4\pi} F_{nm}^{(I,k)} Y_{nm}^{(I,k)}(\theta, \phi) \quad (15a)$$

$$\tilde{h}^{(I,k)}(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{2n+1}{4\pi} \frac{H_{nm}^{(I,k)}}{\hat{l}_n(\kappa a^{(I,k)})} Y_{nm}^{(I,k)}(\theta, \phi) \quad (15b)$$

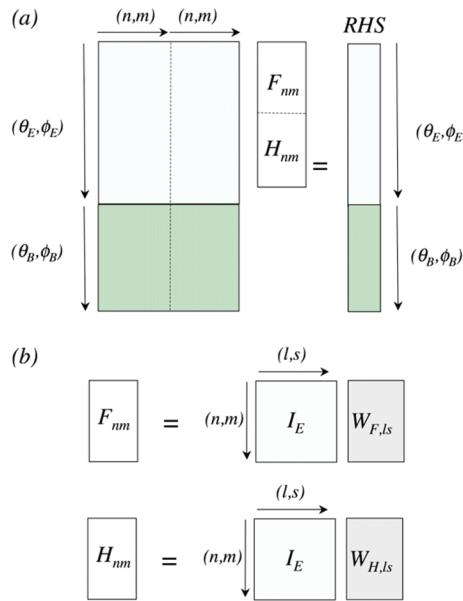
so the “zero-charge” requirement at the buried boundary can be imposed as

$$\sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{2n+1}{4\pi} F_{nm}^{(I,k)} Y_{nm}^{(I,k)}(\theta_B, \phi_B) = 0 \quad (16a)$$

$$\sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{2n+1}{4\pi} \frac{H_{nm}^{(I,k)}}{\hat{l}_n(\kappa a^{(I,k)})} Y_{nm}^{(I,k)}(\theta_B, \phi_B) = 0 \quad (16b)$$

Equations 13a, 13b, 16a, and 16b specified the complete boundary value problem, from which  $F_{nm}^{(I,k)}$  and  $H_{nm}^{(I,k)}$  can be solved.

**Solution of the Boundary Value Coefficients and Interaction Energy.** To solve for  $F_{nm}^{(i,k)}$  and  $H_{nm}^{(i,k)}$ , we need to cast the boundary value problem as a linear system of equations. The infinite expansion series must first be truncated at a maximum pole order  $p$ , chosen depending on the desired level of accuracy versus computational cost (see



**Figure 2.** Setting up the boundary equation (eqs 13a, 13b, 16a, and 16b): (a) as a linear least square solve problem and (b) as a matrix-vector multiply operation.

Results). The obvious approach is to set up the boundary equations as a linear least-squares problem (Figure 2a), by discretizing sphere  $k$  into  $M_B$  buried and  $M_E$  exposed grid points, and then finding solutions of vectors  $\mathbf{F}_{(l,k)}$  and  $\mathbf{H}^{(l,k)}$  that best satisfy the appropriate boundary equations on all grid points. Using the DGELSY routine (complete orthogonal factorization) in LAPACK for  $(M_E + M_B) = 10000$  and  $p = 60$ , each sphere is solved in approximately 10 min. This is computationally intractable if the LPBE needs to be solved repeatedly for tens to hundreds of spheres during dynamics simulations.

Instead, we formulated a novel approach that makes use of spherical harmonics' orthogonal property (eq 4). It converts the problem to a direct matrix-vector multiply operation (Figure 2b), which can be evaluated two-orders of magnitude faster than the LLS approach. We first add  $\sum_{n=0}^{\infty} \sum_{m=-n}^n (2n+1) ((H_{nm}^{(l,k)}) / (\hat{i}_n(\kappa a^{(l,k)})) Y_{nm}^{(l,k)}(\theta_E, \phi_E))$  to both sides of eq 13a and divide by  $4\pi$  to arrive at

$$\sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{2n+1}{4\pi} \frac{H_{nm}^{(l,k)}}{\hat{i}_n(\kappa a^{(l,k)})} Y_{nm}^{(l,k)}(\theta_E, \phi_E) = \tilde{w}_{H,\text{exposed}}(\theta_E, \phi_E) \quad (17a)$$

where

$$\tilde{w}_{H,\text{exposed}}(\theta, \phi) = \frac{1}{4\pi} \sum_{n=0}^{\infty} \sum_{m=-n}^n \left( H_{nm}^{(l,k)} \left( \frac{2n+1}{\hat{i}_n(\kappa a^{(l,k)})} - e^{-\kappa a^{(l,k)}} \hat{k}_n(\kappa a^{(l,k)}) \right) + F_{nm}^{(l,k)} + XH_{nm}^{(l,k)} \right) Y_{nm}^{(l,k)}(\theta, \phi) \quad (17b)$$

Similarly, we add  $\sum_{n=0}^{\infty} \sum_{m=-n}^n (2n+1) F_{nm}^{(l,k)} Y_{nm}^{(l,k)}(\theta_E, \phi_E)$  to both sides of eq 13b and then divide by  $4\pi$

$$\sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{2n+1}{4\pi} F_{nm}^{(l,k)} Y_{nm}^{(l,k)}(\theta_E, \phi_E) = \tilde{w}_{F,\text{exposed}}(\theta_E, \phi_E) \quad (18a)$$

$$\begin{aligned} \tilde{w}_{F,\text{exposed}}(\theta, \phi) &= \frac{1}{4\pi} \sum_{n=0}^{\infty} \sum_{m=-n}^n \times \\ &\left( e^{-\kappa a^{(l,k)}} [n \hat{k}_n(\kappa a^{(l,k)}) - (2n+1) \hat{k}_{n+1}(\kappa a^{(l,k)})] H_{nm}^{(l,k)} + \right. \\ &\left. (2n+1 - n\varepsilon) F_{nm}^{(l,k)} + X F_{nm}^{(l,k)} \right) Y_{nm}^{(l,k)}(\theta, \phi) \end{aligned} \quad (18b)$$

Equations 17a and 17b (and similarly 18a and 18b) now completely describe functions  $\tilde{w}_H(\theta, \phi)$  (and  $\tilde{w}_F(\theta, \phi)$ ) over the entire surface of sphere  $k$

$$\begin{aligned} \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{2n+1}{4\pi} \left[ \frac{H_{nm}^{(l,k)}}{\hat{i}_n(\kappa a^{(l,k)})} \right] Y_{nm}^{(l,k)}(\theta, \phi) &= \tilde{w}_H(\theta, \phi) = \\ \begin{cases} \tilde{w}_{H,\text{exposed}}(\theta, \phi), (\theta, \phi) \in \{\theta_E, \phi_E\} \\ 0, (\theta, \phi) \in \{\theta_B, \phi_B\} \end{cases} \end{aligned} \quad (19a)$$

$$\begin{aligned} \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{2n+1}{4\pi} [F_{nm}^{(l,k)}] Y_{nm}^{(l,k)}(\theta, \phi) &= \tilde{w}_F(\theta, \phi) = \\ \begin{cases} \tilde{w}_{F,\text{exposed}}(\theta, \phi), (\theta, \phi) \in \{\theta_E, \phi_E\} \\ 0, (\theta, \phi) \in \{\theta_B, \phi_B\} \end{cases} \end{aligned} \quad (19b)$$

The above equations now have the familiar form of spherical harmonic expansion of eq 4b, so we can directly evaluate the coefficients in square parentheses via the reciprocal transform eq 4c. We show below the derivation for  $\mathbf{H}^{(l,k)}$

$$\begin{aligned} \frac{H_{nm}^{(l,k)}}{\hat{i}_n(\kappa a^{(l,k)})} &= \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \tilde{w}_H(\theta', \phi') \overline{Y_{nm}^{(l,k)}}(\theta', \phi') \sin \theta' d\theta' d\phi' = \\ &\int_{\phi_E} \int_{\theta_E} \tilde{w}_{H,\text{exposed}}(\theta', \phi') \overline{Y_{nm}^{(l,k)}}(\theta', \phi') \sin \theta' d\theta' d\phi' \\ &= \int_{\phi_E} \int_{\theta_E} \left\{ \sum_{l=0}^{\infty} \sum_{s=-l}^l \left( H_{ls}^{(l,k)} \left( \frac{2l+1}{\hat{i}_l(\kappa a^{(l,k)})} - e^{-\kappa a^{(l,k)}} \hat{k}_l(\kappa a^{(l,k)}) \right) + \right. \right. \\ &\quad \left. \left. F_{ls}^{(l,k)} + X H_{ls}^{(l,k)} \right) Y_{ls}^{(l,k)}(\theta', \phi') \right\} \overline{Y_{nm}^{(l,k)}}(\theta', \phi') \sin \theta' d\theta' d\phi' \\ &= \sum_{l=0}^{\infty} \sum_{s=-l}^l I_{E,lsnm}^{(l,k)} \left( H_{ls}^{(l,k)} \left( \frac{2l+1}{\hat{i}_l(\kappa a^{(l,k)})} - e^{-\kappa a^{(l,k)}} \hat{k}_l(\kappa a^{(l,k)}) \right) + \right. \\ &\quad \left. F_{ls}^{(l,k)} + X H_{ls}^{(l,k)} \right) \end{aligned} \quad (20)$$

where  $I_E$ , the exposed surface integral matrix, is computed using quadrature method with  $M_{\text{grid}}$  uniform surface grid points

$$I_{E,lsnm}^{(l,k)} \equiv \frac{1}{4\pi} \int_{\phi_E} \int_{\theta_E} Y_{ls}^{(l,k)}(\theta', \phi') \overline{Y_{nm}^{(l,k)}}(\theta', \phi') \sin \theta' d\theta' d\phi' \approx \frac{1}{M_{\text{grid}}} \sum_{k=1}^{M_E} Y_{ls}^{(l,k)}(\theta_k, \phi_k) \overline{Y_{nm}^{(l,k)}}(\theta_k, \phi_k) \quad (21)$$

A similar transform to eq 20 can be written for  $F_{nm}^{(l,k)}$ . Finally, we truncate the series at pole order  $p$  to get the iterative equations

$$\begin{aligned} \frac{H_{nm}^{(l,k)}}{\hat{i}_n(\kappa a^{(l,k)})} &= \sum_{l=0}^p \sum_{s=-l}^l I_{E,lsnm}^{(l,k)} \left( H_{ls}^{(l,k)} \left( \frac{2l+1}{\hat{i}_l(\kappa a^{(l,k)})} - e^{-\kappa a^{(l,k)}} \hat{k}_l(\kappa a^{(l,k)}) \right) + \right. \\ &\quad \left. F_{ls}^{(l,k)} + X H_{ls}^{(l,k)} \right) \end{aligned} \quad (22a)$$

$$\begin{aligned} F_{nm}^{(I,k)} = & \sum_{l=0}^p \sum_{s=-l}^l I_{E,lsnm}^{(I,k)} (e^{-\kappa a^{(I,k)}} [lk_l(\kappa a^{(I,k)}) - \\ & (2l+1)k_{l+1}(\kappa a^{(I,k)})] H_{ls}^{(I,k)} + (2l+1-l\varepsilon) F_{ls}^{(I,k)} + X \mathbf{F}_{ls}^{(I,k)}) \end{aligned} \quad (22b)$$

Equations 22a and 22b, along with eqs 14a and 14b, represent a key result of this paper. The equations are iteratively evaluated, until the values of  $\mathbf{F}^{(I,k)}$  and  $\mathbf{H}^{(I,k)}$  converge to a stipulated tolerance. The operations are simply matrix-vector multiply,  $\mathbf{y} = \mathbf{Ax}$ , where the vector  $\mathbf{x}$  is constantly updated using the latest values of  $\mathbf{F}^{(I,k)}$  and  $\mathbf{H}^{(I,k)}$ . During computation, the surface integral coefficients  $I_{E,lsnm}^{(I,k)}$ , and fixed charge coefficients  $E_{nm}^{(I,k)}$  and  $LE_{nm}^{(I,k)}$  are precomputed for each sphere  $(I,k)$  prior to simulation, while  $LF_{nm}^{(I,k)}$ ,  $LH_{nm}^{(I,k)}$ , and  $LHN_{nm}^{(I,k)}$  are updated via multipole-to-local operations (see implementation section below).

In summary, our approach to solve the LPBE is as follows: (1) For each sphere  $k$  in molecule  $I$ , we apply the addition theorems to express the potentials  $\Phi_{in}(\mathbf{r})$  and  $\Phi_{out}(\mathbf{r})$  as spherical harmonic expansions containing unknown coefficients ( $F_{nm}^{(I,k)}$  and  $H_{nm}^{(I,k)}$ ) representing sphere  $k$ 's polarization charges. (2) We impose boundary conditions on the sphere surface to derive boundary equations. (3) We account for charges from other spheres and molecules by re-expanding their polarization coefficients ( $F_{nm}^{(J,l)}$  and  $H_{nm}^{(J,l)}$ ) about the center of sphere  $k$  using “multipole-to-local” operations. (4) We then solve the boundary equations for  $F_{nm}^{(I,k)}$  and  $H_{nm}^{(I,k)}$  iteratively using a novel fast iterative method (*inner-iteration*). (5) We repeat steps 1–4 for all other spheres (*outer-iteration*) until the convergence criteria is reached.

Convergence is monitored using relative change in  $\mathbf{H}^{(I,k)}$  between the  $t$ th and  $(t-1)$ th outer iterations

$$\mu_{H,t}^{(I,k)} \equiv \frac{\sum_{n=0}^p \sum_{m=-n}^n |H_{nm,t}^{(I,k)} - H_{nm,t-1}^{(I,k)}|}{\frac{1}{2} \sum_{n=0}^p \sum_{m=-n}^n |H_{nm,t}^{(I,k)}| + |H_{nm,t-1}^{(I,k)}|} \quad (23)$$

We now can calculate the interaction energies from converged values of  $\mathbf{H}$ . The interaction energy of sphere  $k$  is the inner product of its effective charge distribution with the potential due to external sources. The interaction energy  $W^{(I)}$  of each molecule  $I$  is the sum of interaction energies of its constituent spheres

$$W^{(I)} = \sum_{k=1}^{N_s^{(I)}} \langle LHN^{(I,k)}, \mathbf{H}^{(I,k)} \rangle = \sum_{k=1}^{N_s^{(I)}} \sum_{n=0}^p \sum_{m=-n}^n LHN_{nm}^{(I,k)} H_{nm}^{(I,k)} \quad (24)$$

**Implementation of Re-expansion Operations.** To solve for  $\mathbf{F}^{(I,k)}$  and  $\mathbf{H}^{(I,k)}$ , we need to first account for the polarization charges from all other spheres via  $\mathbf{LF}^{(I,k)}$ ,  $\mathbf{LH}^{(I,k)}$ , and  $\mathbf{LHN}^{(I,k)}$ . To do this, we convert source multipoles  $\mathbf{F}$  and  $\mathbf{H}$  from other spheres to target local expansions centered at  $\mathbf{c}^{(I,k)}$ . If the source and target spheres are well-separated (see criterion below), the re-expansion can be accomplished

analytically through multipole-to-local operators  $\mathbf{T}_0$  and  $\mathbf{T}_k$ . The procedure for computing coefficients of  $\mathbf{T}_0$  and  $\mathbf{T}_k$  has been previously detailed in ref 1. For *intra*-molecular re-expansions (i.e., from spheres  $j$  to center of sphere  $k$  in the same molecule  $I$ )

$$\mathbf{LF}^{(I,k)} = \sum_{j \neq k}^{N_s^{(I)}} \mathbf{T}_0^{(I,k)(I,j)} \mathbf{F}^{(I,j)}; \mathbf{LH}^{(I,k)} = \sum_{j \neq k}^{N_s^{(I)}} \mathbf{T}_k^{(I,k)(I,j)} \mathbf{H}^{(I,j)} \quad (25)$$

or *intermolecular* re-expansions (i.e., from spheres  $l$  on molecule  $J$  to center of sphere  $k$  in the same molecule  $I$ )

$$\mathbf{LHN}^{(I,k)} = \sum_{J \neq I} \sum_{l=1}^{N_{mol}^{(J)}} \mathbf{T}_k^{(I,k)(J,l)} \mathbf{H}^{(J,l)} \quad (26)$$

The analytical re-expansion operators are only valid when the target center  $\mathbf{c}^{(I,k)}$  lies outside the bounding sphere of the source charge distribution, so they cannot be used in cases where source and target spheres overlap. Nonetheless, the local expansions  $\mathbf{LF}^{(I,k)}$  and  $\mathbf{LH}^{(I,k)}$  are still well-defined and could be directly computed using discrete versions of eqs 8c and 10b: a procedure we termed “numerical re-expansion”, as described below. To our knowledge this method of circumventing the restriction by analytical re-expansion has not been previously documented.

We first discretize the surface of source sphere  $j$  uniformly into  $M_p$  patches, with each patch  $b$  centered at  $\mathbf{r}_b^{(I,j)} = [a^{(I,j)}, \theta_b^{(I,j)}, \phi_b^{(I,j)}]$ . We then compute the surface charge on the  $b$ th patch  $\tilde{q}_b^{(I,j)} = 4\pi\tilde{q}^{(I,j)}(\theta_b^{(I,j)}, \phi_b^{(I,j)})/M_p$ , where  $\tilde{q}^{(I,j)} = \tilde{f}^{(I,j)}$  or  $\tilde{h}^{(I,j)}$  from eqs 15a and 15b. The local expansions of sphere  $j$ 's multipoles recentered on  $k$  are then approximated from eqs 8c and 10b as

$$\mathbf{LF}_{nm}^{(I,k)} \approx \sum_{b=1}^{M_p} \frac{f_b^{(I,j)}}{r_b^{(I,k)}} \left( \frac{a^{(I,k)}}{r_b^{(I,k)}} \right)^n \overline{Y_{nm}}(\theta_b^{(I,k)}, \phi_b^{(I,k)}) \quad (27a)$$

$$\mathbf{LH}_{nm}^{(I,k)} \approx \sum_{b=1}^{M_p} \frac{h_b^{(I,j)}}{r_b^{(I,k)}} \left( \frac{a^{(I,k)}}{r_b^{(I,k)}} \right)^n e^{-\kappa r_b^{(I,k)}} \hat{k}_n(\kappa r_b^{(I,k)}) \overline{Y_{nm}}(\theta_b^{(I,k)}, \phi_b^{(I,k)}) \quad (28b)$$

where  $\mathbf{r}_b^{(I,k)} = \mathbf{r}_b^{(I,j)} - (\mathbf{c}^{(I,k)} - \mathbf{c}^{(I,j)})$ . The re-expansion becomes exact as  $M_p$  approaches infinity, although in practice we find that a value of  $M_p \approx 2.5p^2$  adequately captures features of the surface charge distributions. Numerical re-expansion is also used in cases where the source and target spheres are nonoverlapping but not well-separated, which we defined as when the distance between sphere surfaces is less than 5 Å. At such short distance, analytical re-expansion requires a high number of poles for a stipulated level of error. Since both computational time and memory for  $\mathbf{T}$  scales with  $p^3$  it is more efficient to perform the re-expansion using direct numerical method.

We have also derived a formula using Greengard's error bound<sup>33</sup> to adaptively determine the minimum pole order adequate for a re-expansion operation. To re-expand sphere  $(J,j)$ 's multipole to a local expansion at target center  $(I,k)$  within an error of  $\varepsilon_X$ , the pole order required is given by

$$p = \log\left(\frac{\sum_{\text{charges on } j} |\tilde{q}|}{\epsilon_X a^{(J,j)}(c - 1)}\right)/\log(c) - 1 \quad (29)$$

where  $c = (|\mathbf{c}^{(I,k)} - \mathbf{c}^{(J,j)}|)/a^{(I,j)} - 1$  and  $\tilde{q} = \tilde{f}$  or  $\tilde{h}$  are the surface polarization charges. The optimal pole order is calculated on the fly every outer iteration.

**Further Implementation Details.** The surface integral coefficients  $I_{E,\text{sum}}^{(I,k)}$  involve numerical quadratures that are precomputed for each sphere ( $I,k$ ); we have found that the number of quadrature points should scale with pole number as  $M_{\text{grid}} \approx 20p^2$ , which we found to be adequate for capturing the spatial features of the integrand in eq 21.

To prepare a target molecule for computation, we must discretize it into a collection of overlapping spheres. To do so, we first convert its PDB file to PQR format using the PDB2PQR webserver.<sup>12,13</sup> We then obtain its solvent excluded surface (SES) using MSMS<sup>34</sup> and a chosen probe radius  $r_p$  in Å. We proceed with a Monte Carlo search algorithm to find the minimum number of spheres and corresponding radii that satisfy the following criteria: (i) The sphere surface must be at least  $d$  (in Å) away from the outermost atom center. The distance  $d$  can be held constant or set to the van de Waals radius of each atom. (ii) The surface of the spheres cannot protrude more than  $t$  (in Å) from the SES surface. The search is terminated when each atom is encompassed by at least one sphere.

Finally, the code is implemented in C++ and is parallelized in a shared memory framework using openMP 2.0. Timings for PB-SAM and APBS for test cases in Results are based on single processor runs on an Intel(R) Xeon(R) CPU 2.27 GHz node with 24GB of physical memory; we did this to compare PB-SAM in a serial version against the APBS serial code. Parameters used for all APBS calculations are available in Supporting Information. Timings for Brome Mosaic virus calculations are performed on the same node using 8 processors.

## Results

**Nonoverlapping Spherical Test Cases.** We first assess the accuracy of PB-SAM and APBS finite difference solutions against analytical values for three test systems involving 2, 27, and 343 nonoverlapping spherical dielectric cavities (of diameter 20, 15, and 5 Å, respectively) with internal charges placed near the dielectric boundaries (Table 1). For large spheres, this corresponds to a highly asymmetric charge arrangement, while as sphere size decreases the charge

**Table 2.** Comparison of APBS against the Analytical Model for Test Systems Described in Table 1

test system	grid points	resolution (Å)	run time (s)	memory (GB)	overall relative error	maximum relative error
1	65 × 65 × 65	1.5625	3	0.08	19.7%	34.8%
1	129 × 129 × 129	0.7813	29	0.47	14.4%	24.7%
1	257 × 257 × 257	0.3906	142	3.50	11.2%	31.7%
1	513 × 513 × 513	0.1953	1315	27.8	4.9%	11.4%
2	513 × 513 × 513	0.1953	1216	27.8	1.9%	5.3%
3	513 × 513 × 513	0.1953	1421	27.8	1.1%	4.9%

distribution approaches a monopole. The exact analytical solution of the PBE for multiple nonoverlapping spheres has only become available recently.<sup>1</sup> In all cases, the salt concentration is set to 0.05 M, corresponding to  $\kappa = 0.07374$ . We chose a low salt concentration to show a worse case scenario for computational timings that cannot exploit an aggressive interaction cutoff that would be legitimate at higher physiological salt concentrations. Convergence is reached when the relative change  $\mu_{H,i}^{(I,k)}$  falls below  $10^{-2}$  for all spheres.

For test system 1 (two nonoverlapping spheres), we computed the APBS solutions at four different grid resolutions (0.19–1.56 Å) that are typically used in biomolecular applications, and compared the potential value over the entire surface against the analytical model, as well as reporting the corresponding memory requirements and timings (Table 2). The APBS timing scales linearly with the number of grid points because does the memory cost that largely reached the limit of 27 GB on our computing node at the highest resolution we tested. At the most coarse resolution, we find that the APBS error can be as high as ~20% of the theoretical result: as the APBS grid spacing decreases the APBS accuracy increases, reaching ~5% of the true value. The range of our computed errors for APBS agree with the error analysis performed by Moy et al.,<sup>35</sup> who found that the errors from using finite difference methods could range from a few percent at 0.5 Å to more than 100% at 2 Å. Given that these commonly used grid resolutions entailed such large errors, we feel that more systematic benchmarking should be done in the future to quantify accuracies in numerical Poisson–Boltzmann solutions to ascertain the impact on force and free energy computations. Using the highest resolution grid but increasing the number of spheres in test systems 2 and 3, the APBS solution gets corresponding better as the charge distribution simplifies, with average errors of ~2% and ~1%, respectively.

Table 3 shows that the corresponding results for our PB-SAM model. For each system, we report the PB-SAM results

**Table 1.** Spherical Test Systems for Comparison of APBS and PB-SAM to Analytical Model Solution in Tables 2 and 3<sup>a</sup>

test system	description	charge configuration [position from center], charge [e]
1	2 dielectric cavities of radius 20 Å	cavity 1 [18, 0, 0], +3 cavity 2 [-18, 0, 0], -3
2	27 dielectric cavities of radius 15 Å	all cavities [13, 0, 0], +1; [-13, 0, 0], -1 [0, 13, 0], +2; [0, -13, 0], -2 [0, 0, 13], +1; [0, 0, -13], -1
3	343 dielectric cavities of radius 5 Å	all cavities [3, 0, 0], +1; [-3, 0, 0], -1 [0, 3, 0], +2; [0, -3, 0], -2 [0, 0, 3], +1; [0, 0, -3], -1

<sup>a</sup> Cavities have surface-to-surface separation of 1 Å from one another.

**Table 3.** Comparison of PB-SAM against Analytical Model for Test Systems Described in Table 1

test system	number of multipoles	run time (s)	memory (GB)	overall relative error	maximum relative error
1	30	4.3	0.023	13.5%	17.6%
1	35	12.1	0.031	4.3%	4.6%
1	40	20.7	0.051	2.4%	1.9%
2	10	1.4	0.015	13.6%	26.7%
2	15	2.3	0.021	6.4%	11.8%
2	20	7.6	0.033	2.2%	4.1%
2	30	46.5	0.082	0.4%	4.4%
3	5	22.2	0.108	4.4%	9.6%
3	10	28.4	0.167	0.1%	0.3%

for various pole orders, to demonstrate how a user would be able to tune into a desired level of accuracy in PB-SAM's using the pole order. From Table 3, it is clear that we can quickly exceed the accuracy of the APBS solution at a fraction of the cost and memory requirements for all three systems. In all three test cases, very few poles ( $p \leq 40$ ) are needed to define a high accuracy solution, primarily because there are no problematic deep cusp dielectric geometries in the nonoverlapping sphere case.

**Overlapping Spherical Test Cases.** Our second comparison involves two overlapping spheres of various sizes. In this case, no analytical solution is known, but we can define a benchmark calculation based on a high quality PB-SAM solution computed at  $p = 140$  and  $M_p = 200\,000$  (PB-SAM140). The one-time precomputation of the surface integral matrix ( $\mathbf{I}_E$ ) scales with  $p^4$  both in timing and storage. We stopped generating  $\mathbf{I}_E$  at 140 poles, which took 48 h per sphere. Before using  $p = 140$  as a benchmark, we have confirmed that the overall relative difference between potential solved using  $p = 140$  and lesser pole orders decays with increasing pole order: the overall relative difference between the  $p = 130$  and  $p = 140$  solutions are less than 0.5%.

In Table 4, we compare the relative difference in surface potential against PB-SAM140 as sphere size increases. We considered the worst-case scenario by placing the positive charge close to the surface, at a fixed distance of 1.73 Å

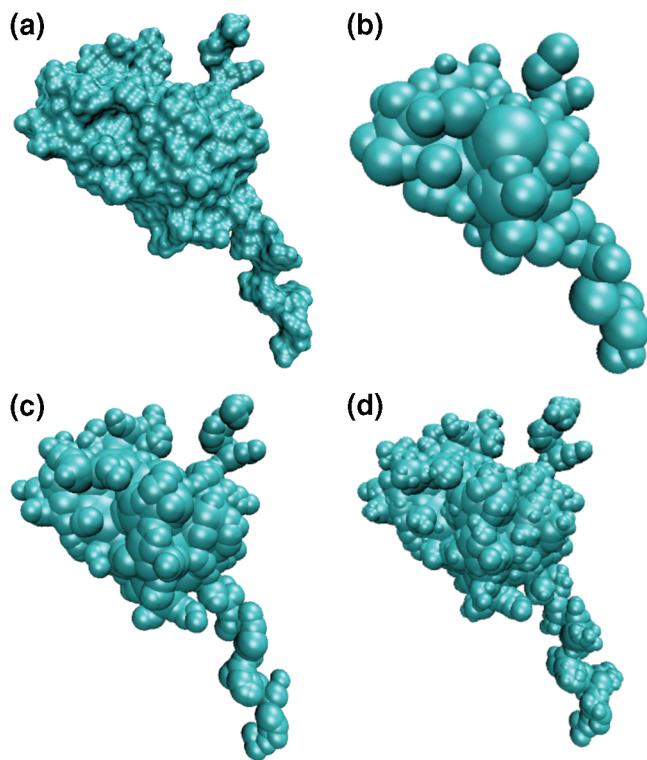
below the cusp region, so that as sphere size grows it results in higher asymmetry of the charge distribution. For each sphere radius, we first compared the surface potential computed by APBS against that of PB-SAM140 and then perform the same comparison between PB-SAM of various pole orders against PB-SAM140. For APBS at a maximum grid dimension fixed at 513,<sup>3</sup> as the system size increases the grid resolution and accordingly the solution accuracy deteriorate. On the other hand, the accuracy of PB-SAM is a function of the pole order and the size of constituent spheres (sphere resolution). Table 4 thus provides a handy estimate of PB-SAM error to help guide our choice of sphere resolutions and pole order for subsequent application to realistic protein cases. For the two overlapping sphere case, PB-SAM with 60 poles is able to achieve relative errors comparable to APBS with comparable total solve time, and with less memory requirements. We want to point out that the total solve time for PB-SAM reported in Table 4 is principally dominated by the one-time cost of surface integral computation (1140 s), while the actual time for solving the iterative equations, eqs 22a and 22b, are between 9 s and 2 min.

It is interesting to note that, for a fixed number of poles, PB-SAM's relative error increases with increasing sphere sizes. Since the boundary equations are formulated and solved in scaled representations, they should be independent of sphere sizes. The two potential sources of error are if  $M_p$  is insufficient in discriminating the positions of the source surface charges for numerical re-expansion, or the scaled fixed charge multipole  $E_{nm}^{(I,k)}$  decays more slowly with poles with increasing charge asymmetry. While we found that increasing  $M_p$  by a factor of 40 resulted in no change in potential, the term  $(r_\alpha/a^{(I,k)})^n$  in  $E_{nm}^{(I,k)}$  converges slower at large sphere radii; hence, more poles are needed to describe the corresponding increase in charge asymmetry. In practice, charges in realistic biomolecules are more evenly distributed; hence, their fixed charge multipoles will converge much faster. The convergence improves further when smaller spheres are used to define higher resolution dielectric

**Table 4.** Two Overlapping Spheres with Varying Sphere Sizes<sup>a</sup>

sphere size	APBS					PB-SAM				
	grid size (Å)	solve time (s)	memory (GB)	relative error	maximum relative error	pole order	solve time (s)	memory (GB)	relative error	maximum relative error
2	0.0195	960	27.8	0.6%	1.6%	20	1141	0.018	12.1%	16.1%
						30	1143	0.030	7.0%	9.2%
						40	1149	0.057	3.8%	5.6%
						60	1209	0.230	1.5%	2.1%
5	0.0391	1018	27.8	1.6%	3.5%	20	1141	0.018	14.4%	20.6%
						30	1143	0.030	7.7%	10.5%
						40	1148	0.057	5.5%	7.8%
						60	1315	0.229	2.3%	3.9%
15	0.1172	1,158	27.8	4.6%	9.7%	20	1141	0.018	21.7%	30.4%
						30	1143	0.030	12.5%	18.4%
						40	1148	0.057	8.6%	14.4%
						60	1223	0.229	4.3%	6.9%
50	0.3906	1,276	27.8	16.8%	32.8%	20	1141	0.018	41.8%	36.2%
						30	1142	0.030	27.7%	25.3%
						40	1148	0.057	19.6%	18.2%
						60	1180	0.229	11.3%	13.3%

<sup>a</sup> Comparison of the surface potential computed with APBS and PB-SAM ( $M_{\text{grid}} = 100k$ ,  $M_p = 2.5p^2$ ) against PB-SAM140.



**Figure 3.** Representations of 1YC6 monomer based on different discretization criteria: (a) the solvent-excluded surface computed using MSMS with  $p = 1.4 \text{ \AA}$ , (b) 107 spheres with  $p = 1 \text{ \AA}$ ,  $d = 1 \text{ \AA}$ ,  $t = 2 \text{ \AA}$ , (c) 354 spheres with  $p = 1 \text{ \AA}$ ,  $d = \text{atomic vdW radii}$ ,  $t = 1 \text{ \AA}$ , and (d) 712 spheres with  $p = 1 \text{ \AA}$ ,  $d = \text{atomic vdW radii}$ ,  $t = 0.5 \text{ \AA}$ .

boundaries. Hence Table 4 shows that PB-SAM's relative error decreases with smaller spheres and higher pole, and our simplified case with maximum charge asymmetry provides a worse case upper bound on the relative error for  $20 < p < 60$ . This will inform estimates of error in our calculation of the bromine mosaic virus in the following section.

**Bromine Mosaic Virus.** We have also applied our PB-SAM method to solve for the potential around a biological molecule, the  $T = 1$  particle of the brome mosaic virus (BMV) capsid (PDB code 1YC6). The virus has been shown to convert from  $T = 3$  (comprising of 180 monomers) to  $T = 1$  (comprising of 60 monomers) capsid under proteolytic conditions.<sup>36</sup> Each capsid protein monomer is comprised of 154 amino acids. To prepare the PDB file for calculation,

we converted chain A of the PDB file into PQR format using the PDB2PQR server,<sup>37,38</sup> which also assigned partial atomic charges using the AMBER 99 force field.<sup>39</sup> We then discretized the protein into a collection of overlapping spheres using an in-house algorithm (see implementation details in Methods). Using discretization criteria that varies in spatial resolution, we generated three representations of the protein monomer with 107, 354, and 712 spheres, and Figure 3 compares the dielectric boundary representation against the solvent excluded surface computed using MSMS<sup>34,40</sup> with probe radius  $r_p = 1.4 \text{ \AA}$ . The generation of 107, 354, and 712 spheres took 11, 30, and 43 min, respectively, which is typically a one-time cost if the dielectric representation does not change during the course of a Brownian dynamics simulation, for example. The resulting rendered dielectric boundary in each case were then used to generate a corresponding APBS solution on the maximum allowed grid points of  $513^3$  given our maximum memory of 24GB.

Table 5 describes the computational time and memory resources for PB-SAM to calculate the self-polarization of one 1YC6 monomer, and the mutual polarization of an array of 60 monomers that make up the unassembled BMV capsid. Since it is our intention to study the dynamics of BMV capsid assembly via Brownian dynamics in future work, we consider the breakdown of computational cost and memory as (1) a one time cost to prepare the surface integral of the chosen dielectric representation of the 1YC6 monomer, (2) the one-time cost to self-polarize each monomer, and (3) the cost to mutually polarize the 60 monomers. In the context of a Brownian dynamic simulation, Table 5 represents the cost of the initialization phase that will require “cold” guesses for  $\mathbf{F}$  and  $\mathbf{H}$  for steps 2 and 3, and the timings will be nonoptimal relative to later solutions that will provide better initial guesses as the dynamics algorithm proceeds as the capsid assembles.

The PB-SAM computational cost depends on the number of poles and number of spheres, and timings are faster or slower depending on how much of the calculation can be done in memory. We use Table 4 to guide our choices of pole order and sphere resolution. We will focus our PB-SAM solutions at a ~5–7% error by choosing pole order  $20 < p < 60$  and keeping average size of spheres of the dielectric boundary representation between 2–5 Å. For step 1, Table 5 shows the one time surface integral

**Table 5.** Computational Timing and Memory Resources Using PB-SAM for Capsid Assembly<sup>a</sup>

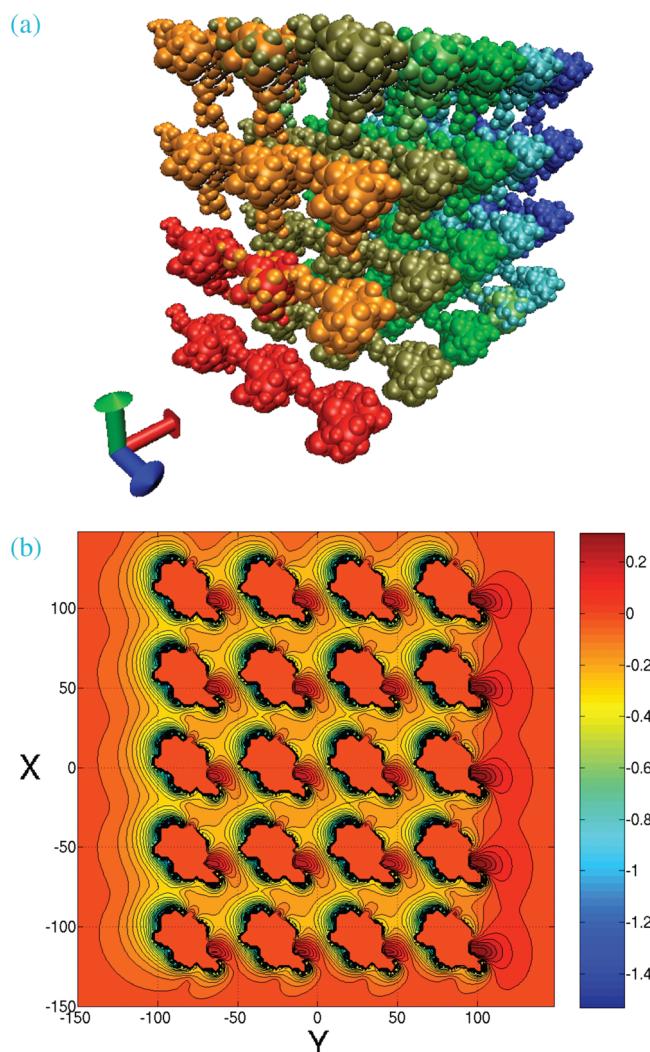
number and median sphere radius	poles	time to calculate surface integrals (s)	self-polarization		mutual-polarization	
			time (s)	memory (GB)	time (s)	memory (GB)
107 spheres 4.40 Å	40	1083	280	3.6	2589	4.4 <sup>b</sup>
	50	4131	552	7.2		
	60	12336	1180	13.3		
354 spheres 3.06 Å	30	423	603	7.8	9365	13.5 <sup>b</sup>
	40	2380	2091	7.1 <sup>b</sup>		
	50	9079	4934	17.1 <sup>b</sup>		
712 spheres 1.91 Å	20	70	271	8.6	16046	33.8 <sup>b</sup>
	30	802	1177	17.5		
	40	4508	3707	14.2 <sup>b</sup>		

<sup>a</sup> Self-polarization of 1YC6 monomer and mutual polarization of 60 monomers of BMV capsid for various dielectric representations (Figure 3). <sup>b</sup> Memory-saving mode.

cost of the 1YC6 monomer, which scales as  $O(M_{\text{grid}} p^4)$  (see Methods), varies between several minutes to several hours. However, a nice benefit is that as resolution increases the sphere size, and hence,  $M_{\text{grid}}$  decreases as do the number of needed poles, which together mitigates the time of calculating more spheres. The cost to self-polarize will depend on the available memory; in memory-saving mode the re-expansion operators  $T_0$  and  $T_K$  are computed on the fly, instead of being stored in memory and hence increase the cost of the calculation. In Table 5, the self-polarization timings are based on a “cold” guess of  $\mathbf{F}^{(I,k)} = 0$  and  $\mathbf{H}^{(I,k)}$  approximated using the fixed charges, and iterated until the relative change in  $\mathbf{H}^{(I,k)}$  falls below  $10^{-2}$  for all spheres.

Unlike the idealized test cases in Table 1, we do not have a benchmark “exact” solution for the 1YC6 monomer, since no analytical solution exists for nonspherical geometries, and computation of hundreds of surface integral matrices to  $p = 140$  for PB-SAM is currently intractable. Instead we can obtain error estimates of APBS and PB-SAM by looking up corresponding PBE solution parameters of grid size for APBS and pole and sphere size for PB-SAM in Table 4. For the 1YC6 monomer, the APBS result is necessarily evaluated at a low resolution of 0.22 Å on the basis of the maximum allowed grid points of  $513^3$  given our maximum memory of 24 GB; Table 4 suggests that the APBS relative error would be  $\sim 10\text{--}12\%$  for this system. For PB-SAM, the error estimate from Table 4 for spheres between 2–5 Å solved to pole order  $20 < p < 60$ , fall between 2.1 to 20.6%. Any direct quantitative comparison between the errors of the APBS and PB-SAM solutions is not possible since there is no exact benchmark for this case, and we can only provide the numerical *difference* between methods (which is  $\sim 10\%$ ) and not *error* between the two numerical solutions. However, using our error estimate from Table 4, it suggests that with 30–40 poles for the representations of 107 and 353 spheres and 20–30 poles for 712 spheres, PB-SAM is a higher quality solution at a comparable CPU cost and memory of the APBS solution.

Finally, we have evaluated the potential of an assembly of 60 copies of 1YC6 monomers in a  $5 \times 4 \times 3$  array, corresponding to a system size of  $165 \text{ \AA} \times 220 \text{ \AA} \times 275 \text{ \AA}$  (Figure 4). The array configuration is intended to mimic late stage assembly, at which the entire capsid system is compact and mutual polarization becomes significant and more difficult to converge (as opposed to the 60 monomers being well separated). All monomers were given the same initial guess of  $\mathbf{F}^{\text{self}}$  and  $\mathbf{H}^{\text{self}}$  from the converged self-polarization step, and the computational time and memory to calculate the total (self-and mutual) polarization is given in Table 5. The memory for the 712-sphere representation required 33 GB of virtual memory, which is not as efficient if it were able to fit in the available 24 GB of physical memory. The fact that the calculation of a high quality solution is doable on a single standard commodity node



**Figure 4.** Array of 60 virus monomers: (a) array configuration and (b) potential profile of a cross-section through the  $z = 0$  plane with twenty monomers. Contour lines at  $0.05 \text{ kT}$ .

is a strength of the PB-SAM approach, although further optimization will be explored in the future.

## Conclusion

We have developed a novel method for solving the linearized Poisson–Boltzmann equation by discretizing the protein surface as a collection of spheres, in which the surface charges can be iteratively solved by our recent analytical solution of the PBE equations for spherical geometries in which mutual polarization is treated exactly.<sup>1</sup> We have compared PB-SAM and the finite difference PB solver APBS against two new benchmarks never before available to compare numerical methods. First, we show that PB-SAM converges to the analytical solution of hundreds of spheres with better accuracy and at greatly reduced cost relative to APBS. Second, the PB-SAM solution using 140 poles allows us to define a high quality benchmark to describe the electrostatic potential for two overlapping spheres that are models for cusp-like features of protein active sites, in which we show that our PB-SAM solution converges to the correct solution with the same computational cost or better than the finite difference solution. Finally we illustrate the strength

of the PB-SAM approach by computing the potential profile of a close configuration of 60 T1-particle forming monomers of the bromine mosaic virus (PDB code 1YC6), with clear improvements in accuracy relative to other numerical PB solutions, given a fixed hardware configuration of physical memory.

Further development is necessary to enable PB-SAM's application in large-scale Brownian dynamic simulations. The current version of PB-SAM expends significant computational time solving eqs 22a and 22b iteratively. This step was implemented simply as repeated calls to the BLAS matrix-vector multiply routine *dgemv* but can be accelerated by preconditioning eqs 22a and 22b and using a more sophisticated linear system solving method, such as generalized minimal residual method. We also noted during our benchmarking studies that when our current convergence criterion is relaxed, the resulting surface potential is unchanged, so there is room explore a less stringent but adequate convergence criterion. Finally, forces and torques are required for Brownian dynamic simulation. We have derived in reference<sup>1</sup> how forces and torques can be computed analytically for spherical dielectrics. The same formulation can be extended to the overlapping sphere representation in PB-SAM via superposition, which is ongoing work in our lab.

**Acknowledgment.** We gratefully acknowledge support from an NIH Multiscale grant and NERSC for computational resources, and Dr. Itay Lotan for providing source code for ref 1 and clarifications.

**Supporting Information Available:** Parameters used for all APBS calculations. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

## References

- (1) Lotan, I.; Head-Gordon, T. *J. Chem. Theory Comput.* **2006**, 2, 541.
- (2) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, 90, 509.
- (3) Chapman, D. L. *Philos. Mag., Ser. 6* **1913**, 25, 475.
- (4) Gouy, G. *J. Physique* **1910**, 9, 456.
- (5) Debye, P.; Huckel, E. *Phys. Z.* **1923**, 24, 185.
- (6) Derjaguin, B.; Landau, L. *Prog. Surf. Sci.* **1993**, 43, 30.
- (7) Verwey, E. J. W. *Philips Res. Rep.* **1945**, 1, 33.
- (8) Kirkwood, J. G. *J. Chem. Phys.* **1934**, 2, 351.
- (9) Fenley, A. T.; Gordon, J. C.; Onufriev, A. *J. Chem. Phys.* **2008**, 129.
- (10) McClurg, R. B.; Zukoski, C. F. *J. Colloid Interface Sci.* **1998**, 208, 529.
- (11) Phillips, G. D. *J. Chem. Phys.* **1974**, 60, 983.
- (12) Sader, J. E.; Lenhoff, A. M. *J. Colloid Interface Sci.* **1998**, 201, 233.
- (13) Lu, B. Z.; Zhou, Y. C.; Holst, M. J.; McCammon, J. A. *Comput. Phys.* **2008**, 3, 973.
- (14) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, 98, 10037.
- (15) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, 12, 435.
- (16) Rocchia, W.; Alexov, E.; Honig, B. *J. Phys. Chem. B.* **2001**, 105, 6507.
- (17) Zhou, Z. X.; Payne, P.; Vasquez, M.; Kuhn, N.; Levitt, M. *J. Comput. Chem.* **1996**, 17, 1344.
- (18) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, 62, 187.
- (19) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R.; McCammon, J. A. *Comput. Phys. Commun.* **1995**, 91, 57.
- (20) Chen, L.; Holst, M. J.; Xu, J. C. *SIAM J. Numer. Anal.* **2007**, 45, 2298.
- (21) Holst, M.; Baker, N.; Wang, F. *J. Comput. Chem.* **2001**, 22, 475.
- (22) Zhou, H. X. *Biophys. J.* **1993**, 65, 955.
- (23) Bordner, A. J.; Huber, G. A. *J. Comput. Chem.* **2003**, 24, 353.
- (24) Boschitsch, A. H.; Fenley, M. O.; Zhou, H. X. *J. Phys. Chem. B* **2002**, 106, 2741.
- (25) Juffer, A. H.; Botta, E. F. F.; Vankeulen, B. A. M.; Vanderploug, A.; Berendsen, H. J. C. *J. Comput. Phys.* **1991**, 97, 144.
- (26) Lu, B. Z.; Cheng, X. L.; Huang, J. F.; McCammon, J. A. *J. Chem. Theory Comput.* **2009**, 5, 1692.
- (27) Lu, B. Z.; Cheng, X. L.; Huang, J. F.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, 103, 19314.
- (28) Lu, B. Z.; McCammon, J. A. *J. Chem. Theory Comput.* **2007**, 3, 1134.
- (29) Zauhar, R. J.; Morgan, R. S. *J. Mol. Biol.* **1985**, 186, 815.
- (30) Gumerov, N. A.; Duraiswami, R. *SIAM J. Sci. Comput.* **2003**, 25, 1344.
- (31) Abramowitz, M.; Stegun, I. A. *Handbook of Mathematical Functions*; U.S. Department of Commerce: Washington, D.C., 1968; Vol. 5.
- (32) Chen, G.; Zhou, J. *Boundary Element Methods*, 1st ed.; Academic Press, London, San Diego, 1992.
- (33) Cheng, H.; Greengard, L.; Rokhlin, V. *J. Comput. Phys.* **1999**, 155, 468.
- (34) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, 38, 305.
- (35) Moy, G.; Corry, B.; Kuyucak, S.; Chung, S. H. *Biophys. J.* **2000**, 78, 2349.
- (36) Lucas, R. W.; Kuznetsov, Y. G.; Larson, S. B.; McPherson, A. *Virology* **2001**, 286, 290.
- (37) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res.* **2007**, 35, W522.
- (38) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res.* **2004**, 32, W665.
- (39) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, 21, 1049.
- (40) Sanner, M. F. *J. Mol. Graphics Modell.* **1999**, 17, 57.