

# An analytical approach to computing biomolecular electrostatic potential. I. Derivation and analysis

Andrew T. Fenley,<sup>1,a)</sup> John C. Gordon,<sup>2,b)</sup> and Alexey Onufriev<sup>1,2,c)</sup>

<sup>1</sup>Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA

<sup>2</sup>Department of Computer Science, Virginia Tech, Blacksburg, Virginia 24061, USA

(Received 24 August 2007; accepted 18 June 2008; published online 18 August 2008)

Analytical approximations to fundamental equations of continuum electrostatics on simple shapes can lead to computationally inexpensive prescriptions for calculating electrostatic properties of realistic molecules. Here, we derive a closed-form analytical approximation to the Poisson equation for an arbitrary distribution of point charges and a spherical dielectric boundary. The simple, parameter-free formula defines continuous electrostatic potential everywhere in space and is obtained from the exact infinite-series (Kirkwood) solution by an approximate summation method that avoids truncating the infinite series. We show that keeping all the terms proves critical for the accuracy of this approximation, which is fully controllable for the sphere. The accuracy is assessed by comparisons with the exact solution for two unit charges placed inside a spherical boundary separating the solute of dielectric 1 and the solvent of dielectric 80. The largest errors occur when the source charges are closest to the dielectric boundary and the test charge is closest to either of the sources. For the source charges placed within 2 Å from the boundary, and the test surface located on the boundary, the root-mean-square error of the approximate potential is less than 0.1 kcal/mol/|e| (per unit test charge). The maximum error is 0.4 kcal/mol/|e|. These results correspond to the simplest first-order formula. A strategy for adopting the proposed method for realistic biomolecular shapes is detailed. An extensive testing and performance analysis on real molecular structures are described in Part II that immediately follows this work as a separate publication. Part II also contains an application example. © 2008 American Institute of Physics. [DOI: 10.1063/1.2956497]

## I. INTRODUCTION

Electrostatic interactions are often a key factor in determining properties of biomolecules,<sup>1–5</sup> including their functions such as catalytic activity,<sup>6,7</sup> ligand binding,<sup>8,9</sup> complex formation,<sup>10</sup> proton transport,<sup>11</sup> as well as structure and stability.<sup>12,13</sup> In-depth studies of electrostatics-based phenomena in macromolecular systems require the ability to compute the potentials and fields efficiently and accurately on the atomic scale.<sup>2,14</sup> Within the framework of the so-called implicit or continuum solvent model,<sup>15–17</sup> the Poisson–Boltzmann (PB) approach is an exact way to compute the electrostatic potential  $\phi(\mathbf{r})$  produced by a molecular charge distribution  $\rho(\mathbf{r})$ . In many practical applications its linearized form is used, in which case the following equation or its equivalent must be solved:

$$\nabla \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) + \kappa^2\epsilon(\mathbf{r})\phi(\mathbf{r}), \quad (1)$$

where  $\epsilon(\mathbf{r})$  is the position-dependent dielectric constant, and the electrostatic screening effects of monovalent salt enter via the Debye–Hückel screening parameter  $\kappa$ .

Historically, the first quantitative approaches to computation and analysis of the electrostatic potential produced by

biomolecular charge distributions relied on analytical approximations<sup>18,19</sup> to Eq. (1), such as the famous model due to Kirkwood.<sup>19</sup> The use of these models led to unique insights into a number of important biophysical problems, for example, protein titration<sup>20</sup> and protein folding.<sup>21</sup> The limited accuracy resulting from the use of simplified shapes such as a sphere to represent the true complexity of a molecular surface was probably thought to be an inevitable drawback of these models and thus prompted the development of numerical approaches to solving the PB equation.

A prototypical numerical PB (NPB) method works by placing the molecule inside a bounding box or surface, defining a three-dimensional (3D) grid of points within it, and then solving for the  $\phi(\mathbf{r})$  at every grid point through iterating a set of self-consistent equations. Currently available tools<sup>22–26</sup> based on these methods produce accurate potential fields  $\phi(\mathbf{r})$  for any realistic charge distribution and molecular shape. The errors of these numerical solutions can be controlled, and, in principle, made arbitrarily small (albeit at an unrealistic computational cost), by adjusting parameters of the numerical models such as the finite-difference grid resolution and the size of the bounding box.

The NPB approaches have become the *de facto* accuracy standard in the field.<sup>27</sup> Despite their widespread acceptance, the methodology has several drawbacks relative to alternative analytical approaches. From the practical standpoint, the NPB methods are fundamentally more complex and gener-

<sup>a)</sup>Electronic mail: afenley@vt.edu.

<sup>b)</sup>Electronic mail: jogordo3@vt.edu.

<sup>c)</sup>Author to whom correspondence should be addressed. Electronic mail: alexey@cs.vt.edu.

ally more expensive computationally compared to closed-form analytical expressions. These differences are especially pronounced in dynamical simulation, where availability of analytical energy functions is particularly advantageous. Generally, the NPB framework does not offer as much freedom and ease in exploring parameter space of simple model systems and toy models and in making qualitative estimates. This ability may be critical for studies aimed at certain fundamental system nonspecific properties of biomolecular systems.<sup>21</sup>

The fundamental difference between NPB and analytical approaches such as the Kirkwood model is seen in the limiting case when  $\phi(\mathbf{r})$  needs to be estimated at a single point in space: The NPB methodology still requires that  $\phi(\mathbf{r})$  is found simultaneously at many points of a finite spacial domain, for example, at every node of a 3D cubic grid or two-dimensional (2D) surface.<sup>28,29</sup> The computational complexity of finding  $\phi(\mathbf{r})$  combined with technical difficulties associated with computing forces due to changes in the molecular surface motivated the search for alternative methods to be used in molecular dynamics (MD) to estimate electrostatic forces within the implicit solvent framework.

While a number of promising models were proposed,<sup>30–33</sup> perhaps the most successful of these analytical alternatives is the generalized Born (GB) approximation pioneered by Still *et al.*<sup>34</sup> around 1990. The model offers an analytical prescription for estimating the electrostatic part of the solvation free energy. The GB's original formulation applies to the zero ionic strength case (the Poisson equation). Later, a heuristic prescription was introduced that successfully adapted the GB approximation to handle the nonzero salt case.<sup>35</sup>

Unlike the infinite-series Kirkwood's solution,<sup>19</sup> the GB expression is a mathematically simple, closed-form formula. Importantly, the GB approximation is also aimed at working for arbitrary shapes, not just spherical as in Kirkwood's model. The algorithmic simplicity and computational efficiency of the original GB model, combined with accuracy improvements, have made it the method of choice in implicit solvent MD,<sup>15,17,36–56</sup> although promising NPB-based alternatives have also been recently tested.<sup>26,57</sup>

Despite the successes of the GB approximation, the model has its own serious drawbacks. First, fundamentally, the GB model does not, even in principle, permit a definition of continuous electrostatic potential everywhere in space: at best, it can only be used to define  $\phi(\mathbf{r})$  at the centers of the atoms.<sup>58</sup> This property is at odds with the very physical nature of electrostatic potential. In practice, the ability to compute the potential at any given point is critical for many applications. Second, unlike many important approximate approaches in physics, for example, the perturbation theory, or the NPB approach itself, the GB model is heuristic in nature and does not have an obvious "handle" that controls its accuracy, at least in principle. As a result, the physical origins of the observed deviations from the NPB reference are hard to trace.<sup>59</sup>

The goal of this work is to overcome these drawbacks and derive a simple analytical approximation of the Poisson equation that is closed form and controllable. Ideally, the

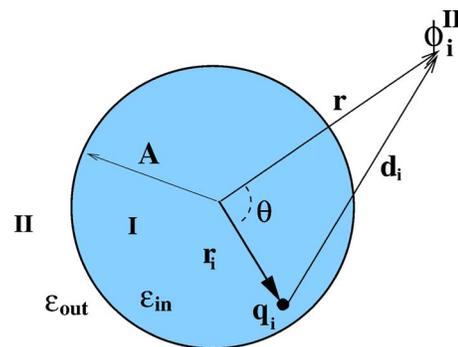


FIG. 1. (Color online) The boundary value problem for Eq. (1). A spherical boundary separates the inside region I, dielectric  $\epsilon_{in}$ , from the outside region II, dielectric  $\epsilon_{out}$ . The point of observation is specified by its spherical coordinates  $(r, \theta)$ ; the source charge is at  $(r_i, 0)$ . Here  $A$  is the radius of the sphere.

approximation should define physically admissible electrostatic potential everywhere in space and should provide a level of accuracy acceptable in practice.

In Part I of the study presented in this paper, we derive several candidates for such an approximation and thoroughly examine their behavior and physical nature on a simple geometry (sphere) for which an exact reference solution of the Poisson problem is available. We propose a candidate approximation for realistic biomolecular shapes and show how its parameters should be redefined once the spherical symmetry is abandoned.

In Part II of this work, which is a separate paper immediately following this one, we adapt the proposed approximation to handling the screening effects of salt and thoroughly test the resulting model on a large number of realistic biomolecules. We then demonstrate how the model might be useful in a concrete problem—a search for putative RNA binding sites on the surface of a viral capsid.

## A. Derivation of the analytical models

The geometric setup of the boundary value problem for the Poisson equation, Eq. (1) with  $\kappa=0$ , is shown in Fig. 1.

We follow Kirkwood<sup>19</sup> to obtain the exact infinite-series expressions for  $\phi(\mathbf{r})$  everywhere in space. The infinite-series solutions for region I (inside) is worked out in detail in Ref. 19, with  $\beta = \epsilon_{in} / \epsilon_{out}$ ,

$$\phi_i^I = \frac{1}{\epsilon_{in}} \frac{q_i}{d_i} + \left( \frac{1}{\epsilon_{out}} - \frac{1}{\epsilon_{in}} \right) \frac{q_i}{A} \sum_{l=0}^{\infty} \left[ \frac{1}{1 + \frac{l}{l+1}\beta} \right] \left( \frac{r_i r}{A^2} \right)^l P_l \cos \theta. \quad (2)$$

The solution for region II is worked out in detail in the appendix at the end of this paper. To summarize, we have arrived at the following solution to the Poisson equation for region II:

$$\phi_i^{II} = -\frac{q_i}{r} \left( \frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}} \right) \sum_{l=0}^{\infty} \left[ \frac{1}{1 + \frac{l}{l+1}\beta} \right] \left( \frac{r_i}{r} \right)^l P_l(\cos \theta) + \frac{q_i}{r} \frac{1}{\epsilon_{in}} \sum_{l=0}^{\infty} \left( \frac{r_i}{r} \right)^l P_l(\cos \theta). \quad (3)$$

Equations (2) and (3) satisfy the usual<sup>60</sup> continuity conditions at the boundary,

$$\phi_i^I(A) = \phi_i^II(A), \quad (4)$$

$$\epsilon_{\text{in}} \left. \frac{\partial \phi_i^I}{\partial r} \right|_A = \epsilon_{\text{out}} \left. \frac{\partial \phi_i^II}{\partial r} \right|_A. \quad (5)$$

The above solutions, Eqs. (2) and (3), of the Poisson equation are valuable since they are *exact*. Unfortunately, they are not very useful in practice since each one is dependent on two infinite series that converge slowly for charge distributions relevant to biomolecules. For example, the infinite series in Eq. (3) converge slowly when  $(r_i/r) \rightarrow 1$ . For the potential near the molecular surface, the ratio being close to 1 is a typical case in real molecules since charged groups are rarely buried due to a high desolvation penalty. As will be discussed below, tens or even hundreds of terms might need to be kept in order to approach well-converged sums. Thus, for practical applications where speed is a factor, something different needs to be done. Also, the infinite series itself or its partial sum is not particularly helpful in illuminating the physical properties of  $\phi(\mathbf{r})$ . A simple closed-form approximation that retains the key physics of the Poisson equation embedded in Eqs. (2) and (3) is what we are looking for. Below we present the detailed derivations for Eq. (3) and just list the end result derived from Eq. (2).

As discussed above, we need to avoid truncating the infinite series. Instead, we keep the  $l=0$  term unchanged and approximate  $l/(l+1) \approx \text{const} = \alpha$  for all  $l > 0$  terms in the first of the two infinite sums in Eq. (3). The approximation is both mathematically and physically motivated.

Mathematically, the approximation recasts the infinite series into a form that can be summed exactly into a closed-form simple formula. The specific algebraic form of  $\alpha$  is motivated by a relatively small variation of  $l/(l+1)$  for any  $l > 0$ :  $1/2 \leq l/(l+1) \leq 1$ .

Physically, this approximation maintains a dependence on the constant  $\beta$ , which encapsulates a specific contribution of the dielectric interface to the potential. While one can easily construct other algebraically “simple” approximations that would provide equal mathematical benefit, e.g.,  $(1 + (l/(l+1))\beta) \approx \text{const} = \alpha$  or  $(1 + (l/(l+1))\beta)^{-1} \approx \text{const} = \alpha$ , these would lose the explicit dependency on  $\beta$  and thus were not considered.

Upon setting  $l/(l+1) \approx \text{const} = \alpha$  for all  $l > 0$ , the infinite series in Eq. (3) is approximated as

$$\begin{aligned} & \sum_{l=0}^{\infty} \left[ \frac{1}{1 + \frac{l}{l+1}\beta} \right] \left( \frac{r_i}{r} \right)^l P_l(\cos \theta) \\ & \approx 1 + \frac{1}{1 + \alpha\beta} \sum_{l=1}^{\infty} \left( \frac{r_i}{r} \right)^l P_l(\cos \theta) \\ & \approx \frac{1}{1 + \alpha\beta} \left[ \sum_{l=0}^{\infty} \left( \frac{r_i}{r} \right)^l P_l(\cos \theta) + \alpha\beta \right]. \end{aligned} \quad (6)$$

We now define  $t = (r_i/r)$  and use the following identity:

$$\sum_{l=0}^{\infty} t^l P_l(\cos \theta) = \frac{1}{\sqrt{1 + t^2 - 2t \cos \theta}}, \quad (7)$$

to approximate the first term in Eq. (3) as

$$\begin{aligned} & \frac{1}{1 + \alpha\beta} \left[ \sum_{l=0}^{\infty} t^l P_l(\cos \theta) + \alpha\beta \right] \\ & \approx \frac{1}{1 + \alpha\beta} \left[ \frac{1}{\sqrt{1 + t^2 - 2t \cos \theta}} + \alpha\beta \right]. \end{aligned} \quad (8)$$

Since  $1/2 \leq l/(l+1) \leq 1$  for  $l > 0$ , a reasonable first guess for  $\alpha$  is the middle of the interval,  $\alpha = 0.75$ . Applying the same identity to the second infinite sum in Eq. (3) and combining the two terms yields the following *closed-form* approximate expression for  $\phi_i^II$ :

$$\begin{aligned} \phi_i^II \approx & -\frac{q_i}{r} \left( \frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \frac{1}{1 + \alpha\beta} \left[ \frac{1}{\sqrt{1 + t^2 - 2t \cos \theta}} + \alpha\beta \right] \\ & + \frac{q_i}{r} \frac{1}{\epsilon_{\text{in}}} \frac{1}{\sqrt{1 + t^2 - 2t \cos \theta}}. \end{aligned} \quad (9)$$

After algebraic manipulations, we arrive at the following analytical form for the electrostatic potential outside of the sphere, region II in Fig. 1. The corresponding expression for the inside space, region I is obtained in the same fashion. Below is the combined key result of this work,

$$\begin{aligned} \phi_i^I \approx & \frac{1}{\epsilon_{\text{in}}} \frac{q_i}{d_i} - \frac{q_i}{A} \left( \frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \frac{1}{1 + \alpha\beta} \\ & \times \left[ \frac{A^2}{\sqrt{(A^2 - r_i^2)(A^2 - r^2) + A^2 d_i^2}} + \alpha\beta \right], \end{aligned} \quad (10)$$

$$\phi_i^II \approx \frac{q_i}{\epsilon_{\text{out}}} \frac{1}{(1 + \alpha\beta)} \left[ \frac{(1 + \alpha)}{d_i} - \frac{\alpha(1 - \beta)}{r} \right]. \quad (11)$$

Since only the first term,  $l=0$ , in the exact infinite sums was kept intact throughout the derivations, the above expression can be referred to as the first-order approximation, although it shall not be confused with truncating the infinite sums. To demonstrate how the accuracy of this approximation can be controlled, at least in principle, we extend Eq. (8) to include the next two terms exactly. Due to the specific symmetry of the Legendre polynomial, retaining the  $l=1$  term exactly improves the accuracy only for antisymmetric charge distributions:  $\rho(\theta) = -\rho(-\theta)$  and the  $l=2$  term improves the accuracy for symmetric charge distributions:  $\rho(\theta) = \rho(-\theta)$ . Thus, the next order that is expected to produce overall improvements in accuracy is the third order according to the terminology just introduced,

$$\begin{aligned} \sum_{l=0}^{\infty} \frac{t^l P_l(\cos \theta)}{1 + \frac{l}{l+1}\beta} \approx & \frac{1}{1 + \alpha\beta} \left[ \frac{1}{\sqrt{1 + t^2 - 2t \cos \theta}} + \alpha\beta \right. \\ & \left. + \frac{\beta(\alpha - 1/2)}{1 + 1/2\beta} t P_1 + \frac{\beta(\alpha - 2/3)}{1 + 2/3\beta} t^2 P_2 \right]. \end{aligned} \quad (12)$$

After similar algebraic manipulations as before, we arrive at

the following third-order expression for the outside potential:

$$\phi_i^{\text{II}} \approx \frac{q_i}{\epsilon_{\text{out}}(1+\alpha\beta)} \left[ \frac{1}{d_i} - \frac{\alpha(1-\beta)}{r} - \frac{(\alpha-1/2)(1-\beta)}{r^2(1+1/2\beta)} r_i P_1 + \frac{(\alpha-2/3)(1-\beta)}{r^3(1+2/3\beta)} r_i^2 P_2 \right]. \quad (13)$$

An analogous third-order expression exists for the inside solution, but it will not be used in this work. An optimal  $\alpha$  for the third-order formula must lie in the interval  $\frac{3}{4} \leq \alpha \leq 1$ ; we choose the middle of the interval,  $\alpha=0.875$ , as a reasonable initial guess.

Higher-order approximations can be defined using the approach described above. Equation (14), shown below, represents the exactly summable  $k$ th-order approximation with  $k/(k+1) \leq \alpha \leq 1$  and  $k \geq 1$ ,

$$\begin{aligned} & \sum_{l=0}^{\infty} \left[ \frac{1}{1 + \frac{l}{l+1}\beta} \right] t^l P_l(\cos \theta) \\ & \approx \sum_{l=0}^{k-1} \left[ \frac{1}{1 + \frac{l}{l+1}\beta} \right] t^l P_l(\cos \theta) + \sum_{l=k}^{\infty} \left[ \frac{1}{1 + \alpha\beta} \right] t^l P_l(\cos \theta) \\ & = \frac{1}{1 + \alpha\beta} \left[ \frac{1}{\sqrt{1+t^2-2t\cos\theta}} \right] \\ & + \sum_{l=0}^{k-1} \left[ \frac{1}{1 + \frac{l}{l+1}\beta} - \frac{1}{1 + \alpha\beta} \right] t^l P_l(\cos \theta). \end{aligned} \quad (14)$$

### 1. Properties of the analytical approximations

We now establish some basic properties of the analytical approximations we have just derived.

*a. Relation to the Poisson equation* Each of the approximate formulas just derived satisfy the Poisson equation. For the first-order Eq. (11), this is seen immediately: The expression is the sum of two Coulomb potentials multiplied by constant prefactors. For Eq. (10) one can verify explicitly that  $\epsilon_{\text{in}} \nabla^2 \phi_i^{\text{I}}(\mathbf{r}) = -4\pi \delta(\mathbf{r}-r_i)$ . The statement remains true for all orders of the approximation. This is because each term in the original infinite-series solution satisfies the Poisson equation; the approximate expression contains the same terms, each multiplied by its own constant.

At first glance, the fact that the analytical approximations also satisfy the Poisson equation may seem to be at odds with the uniqueness theorem that guarantees just one solution of the Poisson problem for the specific boundary conditions. Careful examination of the behavior of our analytical approximations at the boundary resolves the apparent paradox: These analytical approximations satisfy only one of the two continuity equations at the boundary, specifically Eq. (4). The other condition, Eq. (5), is satisfied only approximately;  $(\epsilon_{\text{in}} \partial \phi_i^{\text{I}} / \partial r|_A - \epsilon_{\text{out}} \partial \phi_i^{\text{II}} / \partial r|_A)$  is strictly zero only for the exact infinite-series solution making the exact solution unique. Still, the fact that our analytical approximations satisfy the Poisson equation is reassuring, since it means that these analytical approximations retain some of the key physics of the problem. Their continuity across the

boundary makes this surface a natural location for simultaneously testing the accuracy of both the inside and outside solutions. For this purpose we will use  $\phi_i^{\text{II}}$  defined right outside the dielectric boundary (molecular surface).

The specific form of the approximate solution of order  $k=1$  we have just derived is peculiar: It is mathematically equivalent to the sum of scaled Coulomb potentials due to each source charge plus a scaled Coulomb potential due to the total charge of the system placed in the center of the solute sphere. The scaling factors are nontrivial, but do not depend on the geometry (size) of the solute. In contrast to the multipole expansion, the applicability domain of the approximation includes distances from the solute surface considerably smaller than the solute size  $\mathbf{A}$ .

*b. Accuracy* For the exact spherical geometry considered so far, the error of the analytical approximation for the potential due to a single charge inside the dielectric boundary originates solely from replacing the first infinite sum in Eq. (3) with the  $k$ th-order approximation shown in Eq. (14). A rigorous error bound for this approximation would provide useful general insights into the accuracy of the formulas we have proposed. Such an upper bound is derived in the appendix,

$$\begin{aligned} |\phi_{\text{approx}}^{\text{II}}(k) - \phi_{\text{exact}}^{\text{II}}| & \leq \left| \frac{q}{r} \left( \frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \left( \frac{t^k}{1-t} \right) \left( \frac{\beta}{1+\beta} \right) \right. \\ & \left. \times \left[ \frac{1}{(1+k)} \right] \right|. \end{aligned} \quad (15)$$

For any fixed order  $k$  of the approximation, the error decreases monotonically as the parameter  $t=r_i/r$  approaches zero, i.e., as the test charge moves away from the source. Specifically,  $|\phi_{\text{approx}}^{\text{II}}(k) - \phi_{\text{exact}}^{\text{II}}| = O(r^{-k})$  in the limit  $r \rightarrow \infty$ . Perhaps more interesting is the converse statement, that is, the error bound increases monotonically as the parameter  $t=r_i/r$  approaches unity. This corresponds to the point of observation approaching the source charge, Fig. 1. Obviously, the closer to the source, the larger the potential itself becomes, and so it is perhaps not so surprising that the absolute error of our approximation also increases. However, for any realistic molecular structure the error stays finite. This is because the largest value of  $t$  possible in real molecules is determined by the distance of closest approach of the center of the source and test charges to molecular surface, which is determined by the radius  $\rho_{\text{vdW}}$  of the atom carrying the charge. This physical restriction sets the “worst case” value of  $t$  to be  $(A - \rho_{\text{vdW}})/A$ , and thus suggests that in realistic structures the approximation be tested at a distance of 1–2 Å from the surface. For a fixed geometry of the source and test points,  $t = \text{const}$ , the error bound decreases with increasing order of the approximation  $k$  and approaches zero as  $k \rightarrow \infty$ .

The error bound discussed above does not describe the beneficial effects of error cancellation arising from a specific choice of  $\alpha$ . In particular, how much of an additional benefit do higher-order approximations,  $k > 1$ , provide? To investigate the accuracy of our approximations further we compare the approximate formulas directly with solutions that can be considered numerically exact.

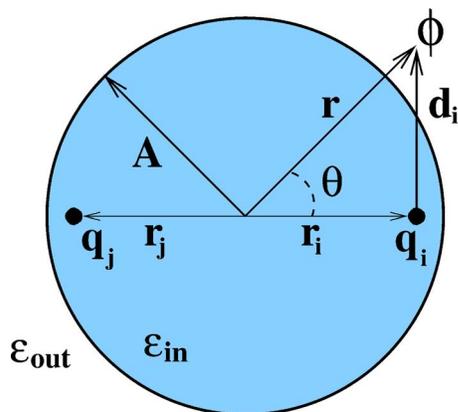


FIG. 2. (Color online) Setup of the test cases. Two unit charges are located on the diameter of a perfect sphere of radius  $A$ , equidistant from the center  $r_i=r_j$ . For the dipole case,  $q_i=-q_j$ , and for the dual positive case,  $q_i=q_j$ . The potential  $\phi(r, \theta)$  is computed at  $r=A$  for  $0 \leq \theta \leq \pi$ .

The exact solution of the Poisson equation on a sphere can be used to test the accuracy of our analytical approximations directly. In practice, we take the sum of the first  $N=1000$  terms in the infinite series in Eq. (3) to represent the exact solution. We use the test setup shown in Fig. 2. For a sphere of radius of  $15 \text{ \AA}$ , which is the size of a typical small protein, the partial sum converges to machine precision when  $\sim 100$  terms are retained, Figs. 3(a) and 3(b). For a larger sphere,  $100 \text{ \AA}$ , which is on the order of the size of a viral capsid, all  $\sim 1000$  terms are needed for the sum to converge to machine precision, Figs. 3(c) and 3(d). These plots demonstrate a key difference between our closed-form analytical approximations, Eqs. (11) and (13), and a brute-force approach in which the first  $N$  terms in the infinite series (3) are retained to approximate  $\phi(\mathbf{r})$ . Depending on the size of the sphere, tens to hundreds of terms will need to be retained to achieve the same level of accuracy provided by the closed-form approximations.

It should be stressed that the “controllability” of the approximations just derived strictly applies only in the case of a perfectly spherical dielectric boundary. In particular, one cannot *a priori* expect that  $\lim_{k \rightarrow \infty} |\phi_{\text{approx}}(k) - \phi_{\text{exact}}| = 0$  for realistic biomolecular structures. We speculate that one may use higher orders  $k > 1$  of the approximation to explore the limits of the sphere-based approach on different classes of realistic biomolecular shapes. Namely, for some shapes and/or regions of space one may observe systematic improvement in the accuracy with increasing  $k$ . For these shapes, one may consider the use of  $k > 1$  formulas. However, our first priority will be to adapt and test the basic  $k=1$  approximation on realistic biomolecular shapes. This is because the error analysis presented above for the spherical shape shows that the bulk of the agreement between the analytical approximations and the exact solution is already achieved within just the first-order approximation, Fig. 3. The next step, the third-order approximation given by Eq. (13), only marginally improves the agreement with the exact solution while substantially increasing the approximation’s complexity. This additional increase in complexity may not be justified, especially if one aims at using the

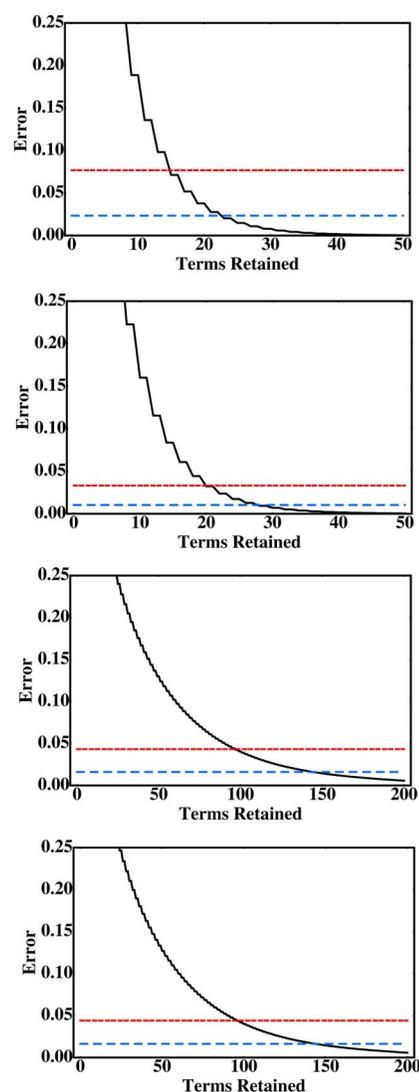


FIG. 3. (Color online) The root-mean-square error, in kcal/mol per unit charge, of the various approximations to the exact solution of the Poisson equation on a sphere. The functions plotted are the error of first-order ( $k=1$ ) analytical approximation, Eq. (11), with  $\alpha=0.750$  (double-dashed red line), the third-order ( $k=3$ ) analytical approximation, Eq. (13), with  $\alpha=0.875$  (dashed blue line), and a partial sum solution obtained by retaining the first  $N$  terms of Eq. (3) (black curve). The potentials are computed at the surface of the sphere over the interval  $0 \leq \theta \leq \pi$ ; the errors are computed with respect to the exact solution, which is the converged partial sum of Eq. (3). The test geometry is shown in Fig. 2. (a) Sphere  $A=15 \text{ \AA}$ , dipole charge distribution, and charges located at  $|r_i|=|r_j|=13 \text{ \AA}$ . (b) Sphere  $A=15 \text{ \AA}$ , dual positive charge distribution, and charges at  $|r_i|=|r_j|=13 \text{ \AA}$ . (c) Sphere  $A=100 \text{ \AA}$ , a dipole charge distribution, and  $|r_i|=|r_j|=98 \text{ \AA}$ . (d) Sphere with  $A=100 \text{ \AA}$ , a dual positive charge distribution, and  $|r_i|=|r_j|=98 \text{ \AA}$ .

formulas in applications where speed and stability of the algorithms are critical.

## 2. Setting parameters of the model

Later in this work we will present additional arguments for using the simpler Eqs. (10) and (11) for real biomolecules. At this point we need to decide what value of the parameter  $\alpha$  in Eqs. (10) and (11) is best. While we could simply take the *ad hoc* value of  $\alpha=0.75$  that was used in Fig. 3 above, we prefer to derive the optimal  $\alpha$  based on

more rigorous grounds. A physically justified choice of  $\alpha$  can come from the requirement that it minimizes the error between the approximate and exact  $\phi(\mathbf{r})$ . There are many reasonable ways to compare two scalar fields defined in 3D space (or 2D if one limits comparison to some Gaussian surface around the charge distribution, for example, the molecular surface). Here, we will use the following approach to set the value of  $\alpha$ : Require that the best  $\alpha$  minimizes the error in the solvation energy of a random charge distribution inside a sphere. We chose this strategy because comparing two real numbers is more straightforward than comparing two scalar fields. This comparison also allows us to make a connection between the current model and the previous ones such as the GB. To this end, we consider an arbitrary charge distribution and define the reaction field potential  $\Phi$  inside the sphere. The  $\Phi$  is given by the inside part of the analytical approximation, Eq. (10), less the Coulomb field:  $\Phi = \sum_i (\phi_i^I - 1/\epsilon_{\text{in}} q_i/d_i)$ . The electrostatic part of the solvation energy is then

$$\Delta G_{\text{el}} = \frac{1}{2} \sum_j q_j \Phi$$

$$\approx -\frac{1}{2} \left( \frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \frac{1}{1 + \alpha\beta} \sum_{ij} q_i q_j \left( \frac{1}{f_{ij}} + \frac{\alpha\beta}{A} \right), \quad (16)$$

with  $f_{ij} = A^{-1} \sqrt{A^2 d_{ij}^2 + (A^2 - r_i^2)(A^2 - r_j^2)}$ .

A closer look at the above expression reveals that it is equivalent to Eq. (3) of Ref. 61 which is the analytic linearized PB (ALPB) model developed in Refs. 33 and 61. Thus, the ALPB model with the above  $f_{ij}$  can be considered a special “discrete” case of the current first-order approximation, Eqs. (10) and (11), for  $\phi(\mathbf{r}_i)$  defined only at the location of the point charges  $q_i$ . This connection allows us to use the optimal value of  $\alpha = 32(3 \ln 2 - 2)/(3\pi^2 - 28) - 1 \approx 0.580127$  which was rigorously derived for the ALPB model.<sup>33</sup> This value of  $\alpha$  should be appropriate for a random charge distribution inside the sphere. One can also check explicitly that the GB model (on a sphere) is also just a particular case of the current theory in the limits  $\epsilon_{\text{out}} \rightarrow \infty$  or  $\alpha \rightarrow 0$ . In the  $\epsilon_{\text{out}} \rightarrow \infty$  limit, the analytical approximations, Eqs. (10), (11), and (13), all become exact solutions of the Poisson equation on a sphere.

With the rigorously justified choice of an optimal value for  $\alpha$ , our approximations, Eqs. (10) and (11), become parameter-free. Their performance for the entire range  $0 \leq \theta \leq \pi$  is compared to the exact solution on the surface of a sphere, Fig. 4. For comparison, the “Null model”—screened Coulomb potential  $1/\epsilon_{\text{out}} \sum_i (q_i/d_i)$  due to the same set of charges  $q_i$ —is also shown.

In agreement with the considerations presented above for the error bound, the largest errors of the approximation occur when the source charges are closest to the boundary and the test charge is closest to one of the sources. For the geometry used to produce the error curves in Fig. 4, these maximal errors for  $k=1$  approximation are  $\sim 0.4$  kcal/mol/ $|e|$  or  $\sim 10\%$  of the corresponding exact value. These are of the same order of what one may expect from a “typical” numerical solution of the PB equation for a

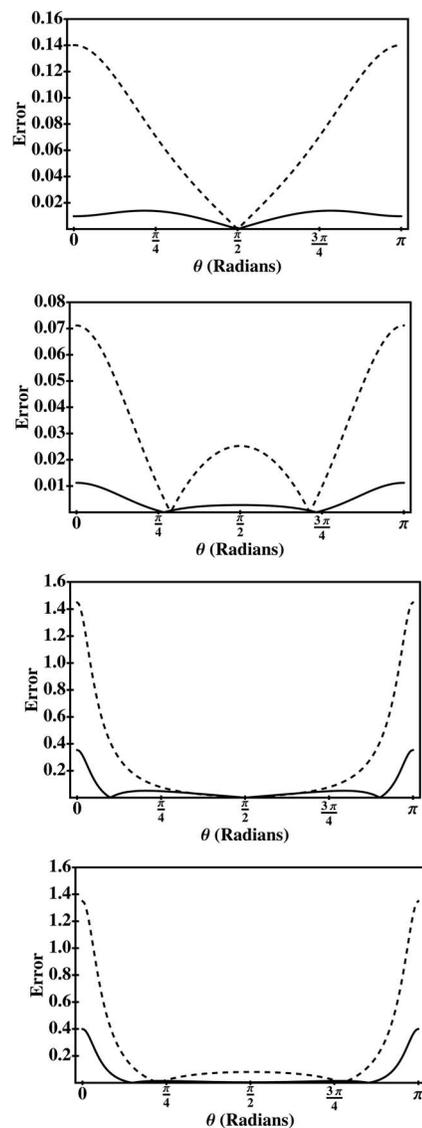


FIG. 4. Absolute error, in kcal/mol per unit charge, of the first-order analytical approximation, Eq. (11), with  $\alpha=0.580127$  (solid lines). The error is computed as the absolute difference between the analytical approximation and the exact solution (converged partial sum). For comparison, the absolute error of the screened Coulomb potential produced by the same charge distribution is also shown (dashed lines). The geometric setup is shown in Fig. 2. (a) Sphere  $A=15$  Å, dipole charge distribution, and unit charges located at  $|r_i|=|r_j|=6$  Å. (b) Sphere  $A=15$  Å, dual positive charge distribution, and unit charges at  $|r_i|=|r_j|=6$  Å. (c) Sphere  $A=15$  Å, dipole charge distribution, and unit charges located at  $|r_i|=|r_j|=13$  Å. (d) Sphere  $A=15$  Å, dual positive charge distribution, and charges at  $|r_i|=|r_j|=13$  Å.

similar test charge geometry. Namely, in an earlier study,<sup>62</sup> a geometric setup similar to ours and the same reference—numerically converged partial sum of the exact series solution for a sphere—was used to assess the accuracy of a finite-difference algorithm that was at the time implemented in the popular package DELPHI. The largest error reported in that study was  $\sim 15\%$  of the exact reference, for the source charge located 1 Å deep inside the dielectric boundary, and the test charges being 3 Å away from the source. One should be careful, however, not to overinterpret such comparisons between two fundamentally different approaches: The accuracy of both can be increased, albeit at additional computational expense. In the case of our analytical approxi-

mation this can be achieved by using its higher orders  $k > 1$ , while the accuracy of the NPB solutions can be improved through a variety of techniques that include focusing<sup>62</sup> or multigrid methods.<sup>24</sup>

The errors of the approximate electrostatic solvation energies  $\Delta G_{\text{el}}$  computed via Eq. (16) for our test geometries are appreciably smaller than the errors (per unit charge) in the potential itself. Namely, for the two source charge geometries described in Fig. 4 the maximum error in  $\Delta G_{\text{el}}$  is  $\sim 0.13$  kcal/mol or only 0.1% of the corresponding exact value. We therefore conclude that direct comparisons between approximate and exact potentials over the entire dielectric boundary is a more sensitive test of the accuracy of the type of approximation considered here. Although quite tedious, these comparisons may thus be preferred to “global metrics” such as  $\Delta G_{\text{el}}$ .

## B. Adaptation to nonspherical shapes

The key question now is how well our analytical approximation for the solution of the Poisson equation on a sphere will perform on shapes that are not exactly spherical. The extensive testing on realistic biomolecular shapes will be presented in Part II of this work that immediately follows this paper. Here, we conclude by showing how our model can be adapted to the nonspherical case.

The first step is to decide what order  $\mathbf{k}$  of the analytical expressions derived above is appropriate for realistic biomolecular shapes. We have already argued that since the first-order Eqs. (10) and (11) and the third-order Eq. (13) perform similarly against the exact solution, Fig. 3, the extra computational complexity of introducing dependencies on Legendre polynomials might be unwarranted. Therefore, we propose that the adaptation of our approximations for realistic molecular shapes begins with the  $k=1$ , Eqs. (10) and (11).

Next, we need to define all the geometrical parameters that enter Eqs. (11) and (10) for the nonspherical case. The distance from the point charge to the point of observation  $d_i$  does not present a problem as it translates directly to the nonspherical case. The distance from the center of the sphere to the observation point  $\mathbf{r}$  is less straightforward. Fortunately, we do have a physical parameter that characterizes the global shape of the structure and replaces the radius of the sphere in the general case—the so-called effective electrostatic radius that was introduced earlier.<sup>33</sup> Once this parameter is computed, which can be done analytically,<sup>61</sup> the  $\mathbf{r}$  distance can be defined as electrostatic radius plus (or minus, if the point of observation is inside the structure) the distance  $\mathbf{p}$  to molecular surface, see Fig. 5.

The above definition of the geometric parameters that enter formulas (10) and (11) for nonspherical geometries is attractive because it treats all regions of space on the same footing. This is why it will be used throughout this work, particularly in Part II. However, depending on specific application, one may find some more restrictive alternatives useful. We note in this respect that the accuracy of the outside solution, Eq. (11), is rather insensitive to the precise definition of  $r$ . This is because the maximum error of the approximation occurs closest to the source on the dielectric bound-

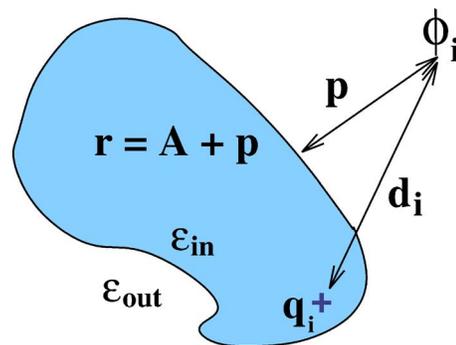


FIG. 5. (Color online) Definition of the geometric parameters that enter the analytical formulas (10) and (11) and can be used to compute the electrostatic potential  $\phi_i$  due to a single charge located inside an arbitrary biomolecule (in the absence of mobile ions). Here  $d_i$  is the distance from the point of observation where  $\phi_i$  needs to be computed, to the source charge  $q_i$ . The distance from the point of observation to the molecular surface is  $\mathbf{p}$  ( $\mathbf{p} < 0$  for points inside the boundary). The so-called effective electrostatic size of the molecule,  $\mathbf{A}$ , characterizes its global shape and is computed analytically as described in Ref. 61. The distance from the point of observation to the “center” of the molecule is then defined as  $r = \mathbf{A} + \mathbf{p}$ . Likewise the position of the charge  $r_i$  is defined as  $\mathbf{A}$  minus the distance of the charge to surface (not shown).

ary, and at this region the  $1/d_i$  terms dominate. To be specific, consider the following example. Suppose the goal is to get a quick estimate of just  $\phi_i^{\text{II}}$  (solvent space), then one can proceed by determining a meaningful geometric center of the structure, and then define  $r$  simply as the distance to it. Since, according to the main definition in Fig. 5,  $r$  cannot be less than  $\mathbf{A}$  for points outside the structure, one should set  $r = \mathbf{A}$  for all  $r \leq \mathbf{A}$ . For an overall neutral molecule,  $\sum_i q_i = 0$ , and the computation simplifies even further as the explicit dependence on  $r$  cancels from the total potential  $\sum_i \phi_i^{\text{II}}$  obtained via Eq. (11).

## II. CONCLUSIONS

In this study we have shown how the exact infinite-series solution of the Poisson equation for an arbitrary charge distribution inside a spherical dielectric boundary can be approximated by a simple analytical formula. We have derived such expressions for the potentials both inside and outside the dielectric boundary, for arbitrary internal and external dielectrics. Unlike the GB model, our model defines electrostatic potential everywhere in 3D space; this parameter-free approximate expression is itself a solution of the Poisson equation, which means that it retains some of the key physics of the problem. We show how an apparent contradiction with the uniqueness theorem of electrostatics is resolved. We have extensively tested the accuracy of the approximation against the exact infinite-series solution represented by its numerically converged partial sum. The errors are assessed for two source charges placed inside the spherical boundary separating the solute of dielectric 1 and the solvent of dielectric 80. We analyzed the errors resulting from several locations of the source charges on the opposite sides of the diameter of the sphere. For unit source charges placed within 2 Å from the boundary, and the test surface located on the boundary, we find the root-mean-square error of the approximate potential to be less than 0.1 kcal/mol/|e| (per unit test charge).

In agreement with the predictions based on a rigorously derived error bound, the largest errors in the approximate potential arise from configurations in which the source charge is closest to the dielectric boundary and the test charge is closest to the source. This maximum error of 0.4 kcal/mol/ $|e|$  or  $\sim 10\%$  of the exact value corresponds to the source charges being 2 Å apart in our test geometry, that is, less than a typical salt-bridge distance. The errors of the approximate electrostatic solvation energies computed via the approximation are noticeably smaller than the corresponding errors in the potential itself. Thus, direct comparisons between approximate and exact potential over the entire dielectric boundary, although tedious, appear to be a more sensitive test of the accuracy of the type of approximation considered here than comparisons based on solvation energy.

Just like the perturbation theory, our approximation is fully controllable, at least in the perfect spherical case considered in this work: it is rigorously shown that the error approaches zero with the increasing order of the approximation. However, unlike the perturbation theory, the approximation is not equivalent to a sum of the first few terms of the infinite-series solution: it effectively retains all of the terms, albeit approximately. To achieve the equivalent accuracy by a straightforward summation of the exact infinite-series solution, tens or even hundreds of terms would have to be retained for realistic charge distributions. While we cannot claim full controllability for realistic biomolecular shapes, we speculate that for some shapes and/or regions of space one may observe systematic improvement in the accuracy with increasing order of the approximation. These improvements are likely to be small though: for the perfectly spherical shape the bulk of the agreement between the analytical approximations and the exact solution is already achieved within just the first-order approximation. Thus, testing the first-order formulas on realistic molecular structures should be the first priority. These tests are performed in Part II of this study that immediately follows.

## ACKNOWLEDGMENTS

The authors thank Grigori Sigalov for reading the manuscript and providing valuable feedback. This work was supported by NIH Grant No. GM076121 and ASPIRES seed grant from Virginia Tech. A.T.F. acknowledges support from NSF IGERT Grant No. DGE-0504196.

## APPENDIX: DERIVATION DETAILS

### 1. Boundary value problem

The derivation refers to the setup shown in Fig. 1. The fixed charges exist only in region I, and so the corresponding Poisson equation is

$$\nabla^2 \phi_i^I = -\frac{q_i}{\epsilon_{in}} \frac{1}{|\mathbf{r} - r_i \hat{\mathbf{e}}_z|}, \quad (\text{A1})$$

where the point charge density  $\rho = q_i \delta(\mathbf{r} - r_i \hat{\mathbf{e}}_z)$  is placed on the  $z$ -axis at position  $r_i$ .

In region II,

$$\nabla^2 \phi_i^{II} = 0. \quad (\text{A2})$$

These two regions in the spherically symmetric case are  $0 \leq r \leq A$  and  $A \leq r < \infty$ , with the charge located on the  $z$ -axis, a distance  $r_i$  from the origin. The solution of the Poisson equation for region I, Eq. (A1), is the sum of Coulomb's potential due to the point charge  $q_i$  and the reaction field part. Due to azimuthal symmetry, the solution depends only on the angle  $\theta$  through Legendre polynomials  $P_l(\cos \theta)$ ,

$$\phi_i^I = \frac{q_i}{\epsilon_{in}} \frac{1}{|\mathbf{r} - r_i \hat{\mathbf{e}}_z|} + \sum_{l=0}^{\infty} B_l r^l P_l(\cos \theta). \quad (\text{A3})$$

Using the following definitions:

$$\text{if } r_i > r, \text{ then } r_i = r_{>} \text{ and } r = r_{<}, \quad (\text{A4})$$

$$\text{if } r_i < r, \text{ then } r_i = r_{<} \text{ and } r = r_{>},$$

and the well-known identity,<sup>60</sup>

$$\frac{q_i}{\epsilon_{in}} \frac{1}{|\mathbf{r} - r_i \hat{\mathbf{e}}_z|} = \frac{q_i}{\epsilon_{in}} \sum_{l=0}^{\infty} \frac{r_{<}^l}{r_{>}^{l+1}} P_l(\cos \theta), \quad (\text{A5})$$

the solution for region I is

$$\phi_i^I = \frac{q_i}{\epsilon_{in}} \sum_{l=0}^{\infty} \frac{r_{<}^l}{r_{>}^{l+1}} P_l(\cos \theta) + \sum_{l=0}^{\infty} B_l r^l P_l(\cos \theta). \quad (\text{A6})$$

No fixed charges are present in region II, which gives

$$\phi_i^{II} = \sum_{l=0}^{\infty} \frac{C_l}{r^{l+1}} P_l(\cos \theta), \quad (\text{A7})$$

where  $B$  and  $C$  are constants determined by the continuity conditions at the boundary  $r=A$ :  $\phi_i^I(A) = \phi_i^{II}(A)$  and  $\epsilon_{in} \partial \phi_i^I / \partial r|_A = \epsilon_{out} \partial \phi_i^{II} / \partial r|_A$ . For the remaining boundary condition, the continuity of the tangential components of the electric field  $\partial \phi_i / \partial \theta$  will be satisfied automatically for the unique exact solution of the Poisson equation.

The first boundary condition gives

$$\begin{aligned} \frac{q_i}{\epsilon_{in}} \sum_{l=0}^{\infty} \frac{r_i^l}{A^{l+1}} P_l(\cos \theta) + \sum_{l=0}^{\infty} B_l A^l P_l(\cos \theta) \\ = \sum_{l=0}^{\infty} \frac{C_l}{A^{l+1}} P_l(\cos \theta). \end{aligned} \quad (\text{A8})$$

Because of the orthogonality of the Legendre polynomials, the equality simplifies to a relation between  $B_l$  and  $C_l$ ,

$$\int_{-1}^1 P_l(x) P_m(x) dx = \frac{2}{2l+1} \delta_{lm} \quad (\text{A9})$$

or, after integration,

$$B_l = \frac{1}{A^{2l+1}} \left( C_l - \frac{q_i}{\epsilon_{in}} (r_i)^l \right). \quad (\text{A10})$$

The second boundary condition equates the normal components of the electric displacement fields of the two regions,

$$\begin{aligned}
& -\epsilon_{\text{out}} \sum_{l=0}^{\infty} (l+1) \frac{C_l}{A^{l+2}} P_l(\cos \theta) \\
& = \epsilon_{\text{in}} \left[ \sum_{l=0}^{\infty} l B_l A^{l-1} P_l(\cos \theta) - \frac{q_i}{\epsilon_{\text{in}}} \sum_{l=0}^{\infty} (l+1) \frac{r_i^l}{A^{l+2}} P_l(\cos \theta) \right].
\end{aligned} \tag{A11}$$

The orthogonality relation between the Legendre polynomials is used again to simplify Eq. (A11), thus providing the second relationship between  $B_l$  and  $C_l$ ,

$$C_l = \frac{\epsilon_{\text{in}}}{\epsilon_{\text{out}}} \left[ \frac{q_i r_i^l}{\epsilon_{\text{in}}} - \frac{l}{l+1} A^{2l+1} B_l \right]. \tag{A12}$$

Equations (A10) and (A12) are solved simultaneously to give independent expressions for  $B_l$  and  $C_l$ ,

$$B_l = \frac{q_i}{A^{2l+1}} r_i^l \left( \frac{1}{\epsilon_{\text{out}}} - \frac{1}{\epsilon_{\text{in}}} \right) \frac{1}{1 + \frac{l}{l+1} \beta}, \tag{A13}$$

$$C_l = q_i r_i^l \left( \frac{1}{\epsilon_{\text{out}}} - \frac{1}{\epsilon_{\text{in}}} \right) \frac{1}{1 + \frac{l}{l+1} \beta} + \frac{q_i}{\epsilon_{\text{in}}} r_i^l. \tag{A14}$$

Recall that the equation for region I is

$$\phi_i^I = \frac{q_i}{\epsilon_{\text{in}}} \sum_{l=0}^{\infty} \frac{r_i^l}{r_>} P_l(\cos \theta) + \sum_{l=0}^{\infty} B_l r^l P_l(\cos \theta). \tag{A15}$$

Let  $t=r_</r_>$ , then the equation for region I becomes

$$\phi_i^I = \frac{1}{\epsilon_{\text{in}}} \frac{q_i}{r_>} \sum_{l=0}^{\infty} t^l P_l(\cos \theta) + \sum_{l=0}^{\infty} B_l r^l P_l(\cos \theta). \tag{A16}$$

After summing up the first infinite series, Eq. (A16) becomes

$$\phi_i^I = \frac{1}{\epsilon_{\text{in}}} \frac{q_i}{r_>} \frac{1}{\sqrt{1+t^2-2t \cos \theta}} + \sum_{l=0}^{\infty} B_l r^l P_l(\cos \theta). \tag{A17}$$

Figure 1 represents the geometry definition and defines  $\cos \theta = (r_>^2 + r_<^2 - d_i^2) / (r_< \cdot r_>)$ . By replacing  $\cos \theta$  with this identity and simplifying the potential in region I,  $\phi_i^I$  becomes

$$\phi_i^I = \frac{1}{\epsilon_{\text{in}}} \frac{q_i}{d_i} + \left( \frac{1}{\epsilon_{\text{out}}} - \frac{1}{\epsilon_{\text{in}}} \right) \frac{q_i}{A} \sum_{l=0}^{\infty} \left[ \frac{1}{1 + \frac{l}{l+1} \beta} \right] \left( \frac{r_i r}{A^2} \right)^l P_l \cos \theta. \tag{A18}$$

To simplify the equation, define the dimensionless distance parameter  $t=(r_i r/A^2)$ . Then

$$\phi_i^I = \frac{1}{\epsilon_{\text{in}}} \frac{q_i}{d_i} + \left( \frac{1}{\epsilon_{\text{out}}} - \frac{1}{\epsilon_{\text{in}}} \right) \frac{q_i}{A} \sum_{l=0}^{\infty} \left[ \frac{1}{1 + \frac{l}{l+1} \beta} \right] t^l P_l \cos \theta. \tag{A19}$$

For region II, the dimensionless distance parameter is  $t=r_i/r$ ; substituting the result for  $C_l$  into Eq. (A7) yields the potential in region II,

$$\begin{aligned}
\phi_i^{II} = & -\frac{q_i}{r} \left( \frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \sum_{l=0}^{\infty} \left[ \frac{1}{1 + \frac{l}{l+1} \beta} \right] t^l P_l(\cos \theta) \\
& + \frac{q_i}{r} \frac{1}{\epsilon_{\text{in}}} \sum_{l=0}^{\infty} t^l P_l(\cos \theta).
\end{aligned} \tag{A20}$$

## 2. Error bound

The error of the approximate analytic solution for the potential in region II for a single charge in a sphere originates from replacing the first infinite sum in Eq. (3) with the  $k$ th-order approximation shown in Eq. (14). Since the terms with  $l < k$  in this approximation are exact, the error is

$$\begin{aligned}
|\phi_{\text{error}}^{II}(k)| & = |\phi_{\text{approx}}^{II}(k) - \phi_{\text{exact}}^{II}| \\
& = \left| -\frac{q}{r} \left( \frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \sum_{l=k}^{\infty} \left[ \frac{1}{1 + \alpha \beta} \right. \right. \\
& \quad \left. \left. - \frac{1}{1 + \frac{l}{l+1} \beta} \right] t^l P_l(\cos \theta) \right|.
\end{aligned} \tag{A21}$$

A relatively simple upper bound for the above infinite sum is available, which depends on the value of  $k$  chosen for the order of the approximation. First, notice that since  $|\sum ab| \leq \sum |a||b|$ , the above error is largest when all  $t^l P_l(\cos \theta)$  are largest and of the same sign, which occurs at  $\cos \theta = 0$  when  $P_l(\cos \theta) = 1$  ( $t \geq 0$  by definition). Then, since  $k/(k+1) < \alpha < 1$ ,  $l/(l+1) < 1$ , and  $l \geq k$  in Eq. (A21), one can check that  $[[1/(1+\alpha\beta) - 1/(1+(l/(l+1))\beta)]] \leq [1/(1+(k/(k+1))\beta) - 1/(1+\beta)]$ . This yields the following expression for the upper bound on  $|\phi_{\text{error}}^{II}(k)|$ :

$$|\phi_{\text{error}}^{II}(k)| \leq \left| \frac{q}{r} \left( \frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \right| \left[ \frac{1}{1 + \frac{k}{k+1} \beta} - \frac{1}{1 + \beta} \right] \sum_{l=k}^{\infty} t^l. \tag{A22}$$

After performing the summation of the geometric series in the above equation along with some algebraic manipulation, we arrive at

$$\begin{aligned}
|\phi_{\text{error}}^{II}(k)| & \leq \left| \frac{q}{r} \left( \frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \right| \left( \frac{t^k}{1-t} \right) \left( \frac{\beta}{1+\beta} \right) \\
& \quad \times \left[ \frac{1}{(1+k+k\beta)} \right].
\end{aligned} \tag{A23}$$

In reality,  $\beta$  is always positive, which allows us to also write

$$|\phi_{\text{error}}^{II}(k)| \leq \left| \frac{q}{r} \left( \frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \right| \left( \frac{t^k}{1-t} \right) \left( \frac{\beta}{1+\beta} \right) \left[ \frac{1}{(1+k)} \right]. \tag{A24}$$

In the important case of aqueous solvation,  $\beta \ll 1$ , this somewhat simpler expression has essentially the same numerical value as the one above it.

<sup>1</sup>M. Perutz, *Science* **201**, 1187 (1978).

<sup>2</sup>B. Honig and A. Nicholls, *Science* **268**, 1144 (1995).

<sup>3</sup>M. E. Davis and J. A. McCammon, *Chem. Rev. (Washington, D.C.)* **90**, 509 (1990).

- <sup>4</sup>N. A. Baker and J. A. McCammon, *Structural Bioinformatics* (Wiley, New York, 2002).
- <sup>5</sup>A. Warshel and J. Åqvist, *Annu. Rev. Biophys. Biophys. Chem.* **20**, 267 (1991).
- <sup>6</sup>A. Warshel, *Biochemistry* **20**, 3167 (1981).
- <sup>7</sup>A. Fersht, J. Shi, J. Knill-Jones, D. Lowe, A. Wilkinson, D. Blow, P. Brick, P. Carter, M. Waye, and G. Winter, *Nature (London)* **314**, 235 (1985).
- <sup>8</sup>G. Szabo, G. Eisenman, S. McLaughlin, and S. Krasne, *Ann. N.Y. Acad. Sci.* **195**, 273 (1972).
- <sup>9</sup>T. Douglas and D. R. Ripoll, *Protein Sci.* **7**, 1083 (1998).
- <sup>10</sup>F. B. Sheinerman, R. Norel, and B. Honig, *Curr. Opin. Struct. Biol.* **10**, 153 (2000).
- <sup>11</sup>A. Onufriev, A. Smondyrev, and D. Bashford, *J. Mol. Biol.* **332**, 1183 (2003).
- <sup>12</sup>A.-S. Yang and B. Honig, *Curr. Opin. Struct. Biol.* **2**, 40 (1992).
- <sup>13</sup>S. Whitten and B. Garcia-Moreno, *Biochemistry* **39**, 14292 (2000).
- <sup>14</sup>K. Chin, K. A. Sharp, B. Honig, and A. M. Pyle, *Nat. Struct. Biol.* **6**, 1055 (1999).
- <sup>15</sup>C. Cramer and D. Truhlar, *Chem. Rev. (Washington, D.C.)* **99**, 2161 (1999).
- <sup>16</sup>B. Roux and T. Simonson, *Biophys. Chem.* **78**, 1 (1999).
- <sup>17</sup>E. Gallicchio and R. Levy, *J. Comput. Chem.* **25**, 479 (2004).
- <sup>18</sup>K. Linderström-Lang, *C. R. Trav. Lab. Carlsberg* **15**, 1 (1924).
- <sup>19</sup>J. G. Kirkwood, *J. Chem. Phys.* **2**, 351 (1934).
- <sup>20</sup>C. Tanford and R. Roxby, *Biochemistry* **11**, 2192 (1972).
- <sup>21</sup>D. Stigter, D. O. Alonso, and K. A. Dill, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4176 (1991).
- <sup>22</sup>J. D. Madura, M. E. Davis, M. K. Gilson, R. C. Wade, B. A. Lutyl, and J. A. McCammon, *Rev. Comput. Chem.* **5**, 229 (1994).
- <sup>23</sup>D. Bashford, in *Scientific Computing in Object-Oriented Parallel Environments*, Lecture Notes in Computer Science, ISCOPE97 Vol. 1343, edited by Y. Ishikawa, R. R. Oldehoeft, J. V. W. Reynders, and M. Tholburn (Springer, Berlin, 1997), pp. 233–240.
- <sup>24</sup>N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037 (2001).
- <sup>25</sup>W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, and B. Honig, *J. Comput. Chem.* **23**, 128 (2002).
- <sup>26</sup>R. Luo, L. David, and M. Gilson, *J. Comput. Chem.* **23**, 1244 (2002).
- <sup>27</sup>N. A. Baker, *Curr. Opin. Struct. Biol.* **15**, 137 (2005).
- <sup>28</sup>M. Totrov and R. Abagyan, *Biopolymers* **60**, 124 (2001).
- <sup>29</sup>B. Lu, X. Cheng, J. Huang, and J. A. McCammon, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19314 (2006).
- <sup>30</sup>R. Abagyan and M. Totrov, *J. Mol. Biol.* **235**, 983 (1994).
- <sup>31</sup>J. J. Havranek and P. B. Harbury, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11145 (1999).
- <sup>32</sup>W. Cai, S. Deng, and D. Jacobs, *J. Comput. Phys.* **223**, 846 (2006).
- <sup>33</sup>G. Sigalov, P. Scheffel, and A. Onufriev, *J. Chem. Phys.* **122**, 094511 (2005).
- <sup>34</sup>W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, *J. Am. Chem. Soc.* **112**, 6127 (1990).
- <sup>35</sup>J. Srinivasan, M. Trevathan, P. Beroza, and D. Case, *Theor. Chem. Acc.* **101**, 426 (1999).
- <sup>36</sup>G. Hawkins, C. Cramer, and D. Truhlar, *Chem. Phys. Lett.* **246**, 122 (1995).
- <sup>37</sup>G. Hawkins, C. Cramer, and D. Truhlar, *J. Phys. Chem.* **100**, 19824 (1996).
- <sup>38</sup>M. Schaefer and M. Karplus, *J. Phys. Chem.* **100**, 1578 (1996).
- <sup>39</sup>D. Qiu, P. Shenkin, F. Hollinger, and W. Still, *J. Phys. Chem. A* **101**, 3005 (1997).
- <sup>40</sup>S. Efinger, C. Cortis, P. Shenkin, and R. Friesner, *J. Phys. Chem. B* **101**, 1190 (1997).
- <sup>41</sup>B. Jayaram, Y. Liu, and D. Beveridge, *J. Chem. Phys.* **109**, 1465 (1998).
- <sup>42</sup>A. Ghosh, C. Rapp, and R. Friesner, *J. Phys. Chem. B* **102**, 10983 (1998).
- <sup>43</sup>D. Bashford and D. Case, *Annu. Rev. Phys. Chem.* **51**, 129 (2000).
- <sup>44</sup>M. Lee, F. Salsbury, Jr., and C. Brooks III, *J. Chem. Phys.* **116**, 10606 (2002).
- <sup>45</sup>A. Felts, Y. Harano, E. Gallicchio, and R. Levy, *Proteins* **56**, 310 (2004).
- <sup>46</sup>B. Dominy and C. Brooks, *J. Phys. Chem. B* **103**, 3765 (1999).
- <sup>47</sup>L. David, R. Luo, and M. Gilson, *J. Comput. Chem.* **21**, 295 (2000).
- <sup>48</sup>V. Spassov, L. Yan, and S. Szalma, *J. Phys. Chem. B* **106**, 8726 (2002).
- <sup>49</sup>N. Calimet, M. Schaefer, and T. Simonson, *Proteins* **45**, 144 (2001).
- <sup>50</sup>V. Tsui and D. Case, *J. Am. Chem. Soc.* **122**, 2489 (2000).
- <sup>51</sup>T. Wang and R. Wade, *Proteins* **50**, 158 (2003).
- <sup>52</sup>A. Onufriev, D. Bashford, and D. Case, *Proteins* **55**, 383 (2004).
- <sup>53</sup>C. Simmerling, B. Strockbine, and A. Roitberg, *J. Am. Chem. Soc.* **124**, 11258 (2002).
- <sup>54</sup>H. Nymeyer and A. Garcia, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13934 (2003).
- <sup>55</sup>M. Lee and Y. Duan, *Proteins* **55**, 620 (2004).
- <sup>56</sup>D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, *J. Comput. Chem.* **26**, 1668 (2005).
- <sup>57</sup>N. V. Prabhu, P. Zhu, and K. A. Sharp, *J. Comput. Chem.* **25**, 2049 (2004).
- <sup>58</sup>A. Onufriev, D. Bashford, and D. Case, *J. Phys. Chem. B* **104**, 3712 (2000).
- <sup>59</sup>D. R. Roe, A. Okur, L. Wickstrom, V. Hornak, and C. Simmerling, *J. Phys. Chem. B* **111**, 1846 (2007).
- <sup>60</sup>J. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, New York, 1999).
- <sup>61</sup>G. Sigalov, A. Fenley, and A. Onufriev, *J. Chem. Phys.* **124**, 124902 (2006).
- <sup>62</sup>M. K. Gilson, K. A. Sharp, and B. H. Honig, *J. Comput. Chem.* **9**, 327 (1988).