Incorporating variable dielectric environments into the generalized Born model

Grigori Sigalov, Peter Scheffel, and Alexey Onufriev
Department of Computer Science, Virginia Tech, Blacksburg, Virginia 24061

(Received 10 November 2004; accepted 16 December 2004; published online 3 March 2005)

A generalized Born (GB) model is proposed that approximates the electrostatic part of macromolecular solvation free energy over the entire range of the solvent and solute dielectric constants. The model contains no fitting parameters, and is derived by matching a general form of the GB Green function with the exact Green's function of the Poisson equation for a random charge distribution inside a perfect sphere. The sphere is assumed to be filled uniformly with dielectric medium ϵ_{in}, and is surrounded by infinite solvent of constant dielectric ϵ_{out}. This model is as computationally efficient as the conventional GB model based on the widely used functional form due to Still *et al.* [J. Am. Chem. Soc. **112**, 6127 (1990)], but captures the essential physics of the dielectric response for all values of ϵ_{in} and ϵ_{out}. This model is tested against the exact solution on a perfect sphere, and against the numerical Poisson–Boltzmann (PB) treatment on a set of macromolecules representing various structural classes. It shows reasonable agreement with both the exact and the numerical solutions of the PB equation (where available) considered as reference, and is more accurate than the conventional GB model over the entire range of dielectric values.
 [DOI: 10.1063/1.1857811]

I. INTRODUCTION

An accurate description of the solvent environment is essential for realistic biomolecular simulations, but may become very expensive computationally. The analytic generalized Born (GB) approximation is a computationally effective way to calculate the main contribution to the total solvation free energy of the molecule—its electrostatic part ΔG_{el}. The methodology has become popular,[1–13] especially in molecular dynamics applications[14–24] due to its relative simplicity and computational efficiency, compared to the more standard numerical solution of the Poisson–Boltzmann (PB) equation.

Both the GB and linear PB approximations share the same underlying physics of the continuum dielectric model in which discrete solvent molecules are replaced by an infinite continuum medium with the dielectric properties of the solvent: despite the fundamental nature of the approximation, the model has enjoyed considerable success in calculating various macromolecular properties.[6,25–30]

Since the first publications on the GB model,[1] it has almost invariably been written in the form

$$\Delta G_{el} \approx \sum_{ij} \Delta G_{ij}^{GB} = -\frac{1}{2}\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}}\right)\sum_{ij} \frac{q_i q_j}{f^{GB}(r_{ij}, R_i, R_j)}, \tag{1}$$

where it is assumed that the molecule is filled uniformly with material of dielectric constant ϵ_{in} and is surrounded by a solvent of dielectric value ϵ_{out} (e.g., 80 for water at 300 K). The sum above is over all pairs of atoms i and j, with r_{ij} representing the distance between them, and q_i being the charge at the center of atom i. So-called *effective Born radius* of atom i is denoted by R_i; it is related to the degree of the atom's burial inside the low dielectric region. Note that $\Delta G_{el} = \Delta G_{el}(\epsilon_{in}, \epsilon_{out})$ calculated via the GB formula above corresponds to the electrostatic free energy of transferring the molecule from a medium with the same dielectric constant as the interior of the molecule, ϵ_{in}, into the medium of dielectric ϵ_{out}, see Fig. 1. The most commonly used form of f^{GB} is the one due to Still:[1]

$$f^{GB} = [r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_i R_j)]^{1/2}, \tag{2}$$

although other expressions have been proposed,[4,31] which approach the same limits as the above formula for $r_{ij} \to 0$ and $r_{ij} \to \infty$. For a hypothetical molecule with a single charge q_i

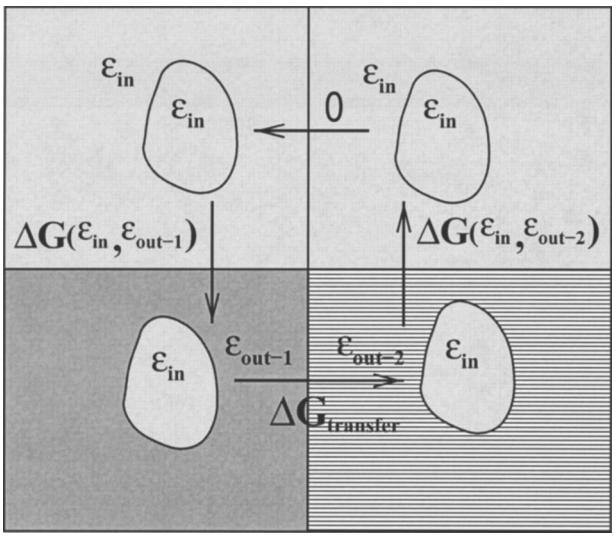

FIG. 1. The thermodynamic cycle used to calculate the electrostatic part of the free energy $\Delta G^{transf}(1 \to 2)$ of transferring a molecule from a medium with dielectric constant ϵ_{in} into a medium of dielectric constant ϵ_{out}. The interior of the molecule has dielectric ϵ_{in}.

at $\mathbf{r}_i$, Eq. (1) takes on a particularly simple form

$$\Delta G_{ii}^{\mathrm{el}} \approx \Delta G_{ii}^{\mathrm{GB}} = -\frac{1}{2}\left(\frac{1}{\epsilon_{\mathrm{in}}} - \frac{1}{\epsilon_{\mathrm{out}}}\right)\frac{q_i^2}{R_i}. \tag{3}$$

Here the reader can recognize the famous expression due to Born for the solvation energy of a single ion of radius R_i. Reasonable agreement of this formula with experiment for simple monovalent ions was arguably one of the first successes of the implicit solvent model based on continuum electrostatics. The key premise of the GB approximation is that if $\Delta G_{ii}^{\mathrm{el}}$ [and hence R_i via the inverse of Eq. (3)] is available for every atom i in the molecule, one can use these effective radii in Eq. (1) to get an estimate of ΔG_{el}. Note that for a molecule of N atoms, a computation via Eq. (1) involves estimation of $\frac{1}{2}N(N+1)$ terms: N diagonal self-energy terms $\Delta G_{ii}^{\mathrm{GB}}$ and $\frac{1}{2}N(N-1)$ off-diagonal "cross" terms $\Delta G_{ij}^{\mathrm{GB}} = \Delta G_{ji}^{\mathrm{GB}}$. Thus, knowledge of only N self-energy terms $\Delta G_{ii}^{\mathrm{GB}}$, or equivalently, values of R_i for each of the N atoms in the molecule gives, via Eq. (1), an estimate of the remaining $\sim\frac{1}{2}N^2$ charge-charge interaction energies $\Delta G_{ij}^{\mathrm{GB}}$. In our view, this is the key assumption and the essence of the GB model. It was shown earlier[31] that if the self-energies $\Delta G_{ii}^{\mathrm{el}}$ and hence R_i are computed accurately enough using the PB approach, then Eqs. (1) and (3) give very reasonable, relative to the PB treatment, estimates of the cross-terms $\Delta G_{ij}^{\mathrm{GB}}$ for biomolecules in water ($\epsilon_{\mathrm{in}}=1, \epsilon_{\mathrm{out}}=80$).

There are many situations when it is desirable to have an efficient estimate of ΔG_{el}, as well as of all of the charge-charge interaction energies, $\Delta G_{ij}^{\mathrm{el}}$, as a function of both the internal and external dielectric constants. For example, in *pK* calculations, a value for the ϵ_{in} larger than 1 may be appropriate,[32] e.g., $\epsilon_{\mathrm{in}}=4$. This higher value takes into account, albeit very approximately, the polarizability of the molecular interior. The external dielectric constant may also be different from its commonly used value of 80 for water at room temperature: the temperature may be different or the solvent may not be pure water. Another situation when one may need $\Delta G_{\mathrm{el}}(\epsilon_{\mathrm{in}}, \epsilon_{\mathrm{out}})$ is in calculations of the free energy of transfer $\Delta G^{\mathrm{transf}}$ of the molecule between two different media with dielectrics $\epsilon_{\mathrm{out}\ 1}$ and $\epsilon_{\mathrm{out}\ 2}$, respectively, since ΔG_{el} is a part of $\Delta G^{\mathrm{transf}}$ (along with the nonpolar contributions). The value of $\Delta G^{\mathrm{transf}}$ can be directly measured experimentally and therefore used in testing and parametrizing of an implicit solvent model such as the GB. Since neither $\epsilon_{\mathrm{out}\ 1}$ nor $\epsilon_{\mathrm{out}\ 2}$ may not, in general, be equal to ϵ_{in}, Eq. (1) needs to be used twice in order to compute the electrostatic component of $\Delta G^{\mathrm{transf}}$. First, one computes $\Delta G_{\mathrm{el}}(\epsilon_{\mathrm{in}}, \epsilon_{\mathrm{out}\ 1})$ and then $\Delta G_{\mathrm{el}}(\epsilon_{\mathrm{in}}, \epsilon_{\mathrm{out}\ 2})$ to obtain $\Delta G_{\mathrm{el}}^{\mathrm{transf}}(1\rightarrow 2) = \Delta G_{\mathrm{el}}(\epsilon_{\mathrm{in}}, \epsilon_{\mathrm{out}\ 1}) - \Delta G_{\mathrm{el}}(\epsilon_{\mathrm{in}}, \epsilon_{\mathrm{out}\ 2})$; refer to the thermodynamic cycle in Fig. 1.

The explicit dependence of the ΔG_{el} on the value of the internal dielectric constant ϵ_{in} is in the prefactor in Eq. (1), and so when it is combined with Eq. (2) to calculate the transfer energy, this dependence cancels out, leading to $\Delta G_{\mathrm{el}}^{\mathrm{transf}}(1\rightarrow 2)$ being independent of ϵ_{in} (providing that the effective Born radii depend only on the molecular geometry, which is the usual assumption). The following simple argument shows that this cannot generally be true: the distribution of the polarization charge on the molecular surface does depend on ϵ_{in}, therefore affecting $\Delta G_{\mathrm{el}}^{\mathrm{transf}}(1\rightarrow 2)$. (The only exception is a spherical ion with a single charge in its center, in which case the Born formula corresponds to the exact solution of the Poisson equation.) This argument is confirmed by the numerical calculations (see below), which show a large discrepancy between the $\Delta G_{\mathrm{el}}(\epsilon_{\mathrm{in}}, \epsilon_{\mathrm{out}})$ values computed within the GB models based on Eq. (1) and PB models. The purpose of this work is to introduce the correct functional dependence of ΔG_{el} upon both the internal and external dielectrics, based on rigorous physical principles.

Formally, the key principle of the GB model can be written in terms of the Green function[33] of the corresponding Poisson equation (we do not consider the effects of salt here)

$$\nabla[\epsilon(\mathbf{r})\,\nabla\,\mathbf{G}(\mathbf{r}_i,\mathbf{r}_j)] = -4\pi\delta(\mathbf{r}_i-\mathbf{r}_j), \tag{4}$$

whose solution is

$$\mathbf{G}(\mathbf{r}_i,\mathbf{r}_j) = \frac{1}{|\mathbf{r}_i-\mathbf{r}_j|} + \mathbf{F}(\mathbf{r}_i,\mathbf{r}_j), \tag{5}$$

where the effects of the nontrivial molecular boundary are embedded in the reaction field component of the Green function $\mathbf{F}(\mathbf{r}_i,\mathbf{r}_j)$. The latter can be used to find ΔG_{el} via

$$\Delta G_{\mathrm{el}} = \frac{1}{2}\sum_{ij}\mathbf{F}(\mathbf{r}_i,\mathbf{r}_j)q_iq_j. \tag{6}$$

We now postulate the generalized Born model in its most general form

$$\mathbf{F}(\mathbf{r}_i,\mathbf{r}_j) \approx \mathbf{F}_{\mathrm{GB}}(\Delta G_{ii}^{\mathrm{el}}, \Delta G_{jj}^{\mathrm{el}}, r_{ij}), \tag{7}$$

where $\Delta G_{ii}^{\mathrm{el}}$ and $\Delta G_{jj}^{\mathrm{el}}$ are self-energy contributions to the electrostatic part of the solvation free energy of charges q_i and q_j in the molecule. The goal of this study is to propose a functional form of $\mathbf{F}_{\mathrm{GB}}$ that would incorporate a realistic dependence on dielectric environments.

Our strategy is as follows: start with the exact Green function for a simple geometry and then use it on the left-hand side of Eq. (7) to suggest a computationally facile form of $\mathbf{F}_{\mathrm{GB}}$. A natural choice of the model geometry for globular molecules is a perfect sphere with point charges distributed inside—the exact solution of the Poisson equation for this system was found a long time ago by Kirkwood,[34] and will serve as a starting point for this model.

In this study we assume that the self-energy contributions $\Delta G_{ii}^{\mathrm{el}}$ can be computed accurately and in a way that does not depend upon any particular choice of the functional form for $\mathbf{F}_{\mathrm{GB}}$ in Eq. (7). While a number of approximate strategies exist to compute this quantity, or equivalently the effective Born radius, we use the essentially exact numerical solution of the Poisson equation for this purpose. While expensive, and hardly practical in applications, this approach was found to be useful in theoretical studies,[31] as it eliminates one possible source of error and allows us to focus on the functional form of the GB equation.

This paper has the following structure. Our GB method (GBϵ) is derived from Kirkwood's exact solution of the Poisson equation for a sphere in Sec. II. We show how the perfect sphere solution can be transferred to the case of an arbitrary globular molecule by expressing the appropriate variables

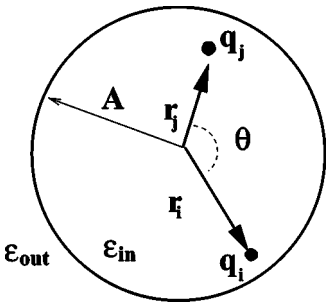

FIG. 2. Illustration for Eqs. (8) and (9): a sphere of dielectric ϵ_{in} of radius A with two charges, q_i and q_j at positions $\mathbf{r}_i$ and $\mathbf{r}_j$ relative to the sphere's center. The sphere is surrounded by infinite medium of dielectric ϵ_{out}.

via local geometry parameters. Most mathematical details of the derivation are placed in the Appendixes, while physics is discussed in Sec. II. The GBϵ method is tested on a set of structures described in Sec. III. The accuracy and performance of the method are compared to those of the existing PB and conventional GB methods in Sec. IV. The physical origin of unusually high effective dielectric values sometimes observed in both the PB and the GB models is discussed. We also explore the question of whether or not effective Born radii should be considered as dependent on the solvent or solute dielectrics. The conclusion is given in Sec. V.

II. THEORY: ϵ-DEPENDENT GENERALIZED BORN METHOD

A. Kirkwood's solution for an ideal sphere

We begin by recasting Kirkwood's well-known solution[34] of the Poisson equation for a spherical molecule in terms of its Green's function. For further convenience, we separate the self-contribution $\mathbf{F}(\mathbf{r}_i,\mathbf{r}_i)$ from the interaction part $\mathbf{F}(\mathbf{r}_i,\mathbf{r}_j)$:

$$\mathbf{F}(\mathbf{r}_i,\mathbf{r}_i)^{\text{sphere}} = -\frac{1-\beta}{A\epsilon_{\text{in}}}\sum_{l=0}^{\infty}\frac{t_{ii}^l}{1+\frac{l}{l+1}\beta}, \tag{8}$$

$$\mathbf{F}(\mathbf{r}_i,\mathbf{r}_j)^{\text{sphere}} = -\frac{1-\beta}{A\epsilon_{\text{in}}}\sum_{l=0}^{\infty}\frac{t_{ij}^l P_l(\cos\theta)}{1+\frac{l}{l+1}\beta}, \tag{9}$$

where $t_{ij}=r_i r_j/A^2$, $r_i=|\mathbf{r}_i|$ being the atom's position relative to the center of the sphere, A is molecule's radius, θ is the angle between $\mathbf{r}_i$ and $\mathbf{r}_j$, and $\beta=\epsilon_{\text{in}}/\epsilon_{\text{out}}$ (see Fig. 2).

This solution is *exact* for a sphere, and may be expected to be reasonably accurate for many realistic globular molecules. The dependence upon the internal dielectric constant enters in a nontrivial way (in the denominator of the summand, via β), and, unlike in the case of the traditional GB theory based upon Eq. (1), this dependence does not simply cancel out when the transfer energy is computed.

B. Local geometry approach for arbitrarily shaped molecules

The Kirkwood equations [Eqs. (8) and (9)] are exact for a hypothetical molecule with a perfectly spherical boundary. However, they cannot be immediately used for realistic molecules. The problem is that it is not at all obvious how the values of A and $\mathbf{r}_i$, and therefore t_{ij} and θ, should be defined for a nonspherical molecule. While the distance between the atoms $r_{ij}=|\mathbf{r}_i-\mathbf{r}_j|$ can be easily computed for any pair of atoms in the molecule, the definition of individual $\mathbf{r}_i$ (Fig. 2) requires the knowledge of the molecule's center, which may not be defined unambiguously for a molecule of irregular shape. As long as $\mathbf{r}_i$ and $\mathbf{r}_j$ are not clearly defined, the angle between them, θ, is not known, either. Last but not the least, a suitable definition of the "electrostatic radius" A must be given in the general case of nonspherical geometries. Our strategy is to express these key quantities—A and $\mathbf{r}_i$—through the precomputed values of $\Delta G_{ii}^{\text{el}}$.

To this end, we consider two separate limiting cases of Kirkwood's solution for the self-energy part, Eq. (8): $\beta\to\infty$ and $\beta\to 0$. In the first case only the $l=0$ term survives, while in the other limit we have a geometric series that does not depend on l and can be summed exactly. Recalling that $\Delta G_{ii}^{\text{el}}=\frac{1}{2}\mathbf{F}(\mathbf{r}_i,\mathbf{r}_i)q_i^2$, we can formally express A and r_i via the corresponding $\Delta G_{ii}^{\text{el}}(\beta\to\infty)$ and $\Delta G_{ii}^{\text{el}}(\beta\to 0)$:

$$\Delta G_{ii}^{\text{el}}(\beta\to\infty) = -\frac{q_i^2}{2}\left(\frac{1}{\epsilon_{\text{in}}}-\frac{1}{\epsilon_{\text{out}}}\right)\frac{1}{A}, \tag{10}$$

$$\begin{aligned}\Delta G_{ii}^{\text{el}}(\beta\to 0) &= -\frac{q_i^2}{2}\left(\frac{1}{\epsilon_{\text{in}}}-\frac{1}{\epsilon_{\text{out}}}\right)\frac{1}{A(1-t_{ii})}\\ &= -\frac{q_i^2}{2}\left(\frac{1}{\epsilon_{\text{in}}}-\frac{1}{\epsilon_{\text{out}}}\right)\frac{1}{A-r_i^2/A}.\end{aligned} \tag{11}$$

Once the value of A is determined from Eq. (10), r_i can be computed for every atom via Eq. (11), and we can then use the cosine rule to uniquely define

$$\cos\theta = (r_i^2+r_j^2-r_{ij}^2)/2r_i r_j. \tag{12}$$

To summarize, we define the effective local geometry around every pair of charges, based on the known self-energy of each charge embedded in the molecule. Therefore, provided that $\Delta G_{ii}^{\text{el}}(\beta\to\infty)$ and $\Delta G_{ii}^{\text{el}}(\beta\to 0)$ are known, we can express all of the input parameters of Eq. (9) through these quantities, and compute all the pair interactions $\Delta G_{ij}^{\text{GB}}(\epsilon_{\text{in}},\epsilon_{\text{out}})$, which is the goal of the GB theory. This is in the spirit of, but going beyond, the conventional GB model in which the key quantity—the effective Born radius—is computed from the self-energy via Eq. (3). To follow the GB convention of expressing $\Delta G_{ii}^{\text{el}}$ via quantities with the dimension of length, we compare Eq. (3) with Eq. (11) to relate r_i directly to the corresponding effective Born radius:

$$r_i^2 = A(A-\tilde{R}_i), \tag{13}$$

where we use $\tilde{R}$ to denote the effective Born radii computed in the limit $\beta=\epsilon_{\text{in}}/\epsilon_{\text{out}}\to 0$. Note that this quantity is close, but not exactly equal to the effective radius computed at the

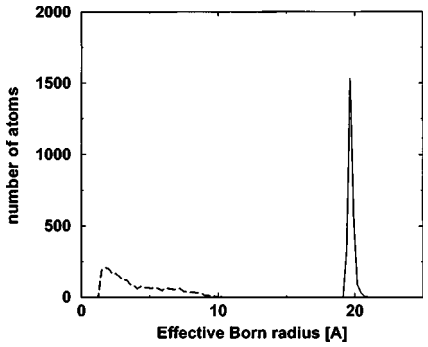

FIG. 3. Distribution of perfect effective Born radii for myoglobin in media with $\epsilon_{in}/\epsilon_{out}=10^4$ (solid line) and $\epsilon_{in}/\epsilon_{out}=1/80$ (dashed line). When $\epsilon_{in} \gg \epsilon_{out}$, all effective radii are essentially the same and equal to the molecule's electrostatic radius.

standard solvation conditions $\epsilon_{in}=1$, $\epsilon_{out}=80$, as often used in the GB literature. We will see later that it is $\tilde{R}$ that should be used in the GB formulas of the kind given by Eq. (1). To sum up, if A and $\tilde{R}_i$ have been computed, the values of t_{ij} and $\cos\theta$ needed in the Green function [Eqs. (8) and (9)] can be expressed through them by simple algebra.

While the meaning and the ways of computing of the effective Born radii have been extensively discussed in the literature, in this work we introduce the concept of the *electrostatic radius A*. A discussion about its physical meaning is therefore required. The defining equation, Eq. (10), is quite curious, as it effectively states that $\Delta G_{el}(\beta\to\infty)$ is independent of the charge distribution within the molecule. For a perfect sphere this result follows directly from Kirkwood's solution, and for an arbitrary shape can be qualitatively explained as follows. In the $\epsilon_{in}\to\infty$ limit (molecule filled with conductor) the electric field of the charges inside must be completely screened out by the polarization charges in the immediate vicinity, to ensure that the electric field inside the bulk of the molecule is zero. Therefore, the polarization charges that build up on the dielectric boundary do not "see" the charges inside, their distribution is determined by the geometry of the molecular boundary only, and is completely independent of the positions of the atomic charges. The only relationship between them follows from the overall neutrality of the dielectric, which necessitates that Σ(polarization charges)$=\Sigma_i q_i$. The total electrostatic energy of the system is expressed through the volume integral of the electric field, and since the electric field inside is zero, only the outside field due to polarization charges on the boundary will contribute. Therefore, the solvation energy of such (hypothetical) molecule will depend only on its total charge and not on how this charge is distributed inside. Note that the electrostatic radius of the molecule is nothing else but the limiting value of the effective Born radii of its atoms in the limit $\epsilon_{in}/\epsilon_{out}\to\infty$. However, while N separate computations have to be performed to compute all of effective Born radii for a molecule of N atoms, only one computation is, in principle, necessary to estimate A, since this quantity is independent of the position of the probe charge q_i inside the molecule. The point is illustrated in the Fig. 3 below, where the distribution of perfect,[31] i.e., PB-based, effective radii is shown in the $\epsilon_{in}/\epsilon_{out}\ll 1$ and $\epsilon_{in}/\epsilon_{out}\gg 1$ limits for a realistic protein myoglobin. The distribution of effective radii for $\epsilon_{in}/\epsilon_{out}=10^4$ is strongly peaked around the electrostatic radius $A=19.72$ Å, with the small spread likely due to numerical uncertainties.

C. GBε: Improved GB model with ε-dependent Green's function

One of the key advantages of the GB method is that it is an *analytical* formula, which is important when computation time is critical, in particular, in molecular dynamics applications. In what follows, we present an effective, accurate, yet simple enough approximation to Kirkwood's solution [Eqs. (8) and (9)].

Let us rewrite the series in Eq. (9) in the following way:

$$\sum_{l=0}^{\infty}\frac{t_{ij}^l P_l(\cos\theta)}{1+\frac{l}{l+1}\beta}=1+\sum_{l=1}^{\infty}\frac{t_{ij}^l P_l(\cos\theta)}{1+\frac{l}{l+1}\beta}$$
$$\approx 1+\sum_{l=0}^{\infty}\frac{t_{ij}^l P_l(\cos\theta)}{1+\alpha\beta}-\frac{1}{1+\alpha\beta}$$
$$=\frac{1}{1+\alpha\beta}\left[\sum_{l=0}^{\infty}t_{ij}^l P_l(\cos\theta)+\alpha\beta\right], \qquad (14)$$

where we have made an approximation $l/(l+1)\approx\alpha=\text{const}$. This approximation is reasonable for $l>0$, since in this case $l/(l+1)$ only varies between $1/2$ and 1. A value that minimizes the mean-square error of this approximation for a large spherical molecule has been analytically found to be $\alpha\approx 0.57\pm 0.01$, the spread being due to weak dependence of α on β (see Appendix A). We later show that $\alpha=0.57$ is also a good choice for realistic molecules.

Using the well-known identity for Legendre's polynomials [Eq. (B8), Appendix B] to sum the series on the right-hand side of Eq. (14), we obtain the following expression:

$$\mathbf{F}(\mathbf{r}_i,\mathbf{r}_j)^{\text{sphere}}\approx-\frac{1}{A\epsilon_{in}}\frac{1-\beta}{1+\alpha\beta}$$
$$\times\left[\frac{1}{\sqrt{1+t_{ij}^2-2t_{ij}\cos\theta}}+\alpha\beta\right]. \qquad (15)$$

We now use t_{ij}, r_i, and $\cos\theta$ as defined by the local geometry considerations [Eqs. (10)–(12)] along with Eq. (13) to express the arguments of the above equation through the electrostatic radius A and the set of effective Born radii $\tilde{R}_i$:

$$A^2(1-2t_{ij}\cos\theta+t_{ij}^2)=A^2-2r_i r_j\cos\theta+\frac{r_i^2 r_j^2}{A^2}$$
$$=r_{ij}^2+\left(A-\frac{r_i^2}{A}\right)\left(A-\frac{r_j^2}{A}\right)$$
$$=r_{ij}^2+\tilde{R}_i\tilde{R}_j, \qquad (16)$$

which leads to

$$\mathbf{F}(\mathbf{r}_i,\mathbf{r}_j)^{\text{sphere}} \approx -\left(\frac{1}{\epsilon_{\text{in}}}-\frac{1}{\epsilon_{\text{out}}}\right)\frac{1}{1+\alpha\beta} \times\left[\frac{1}{\sqrt{r_{ij}^2+\tilde{R}_i\tilde{R}_j}}+\frac{\alpha\beta}{A}\right]. \tag{17}$$

The functional form $f_{ij}^{\text{sphere}}=\sqrt{r_{ij}^2+R_iR_j}$ in the above equation has been derived from Green's function of the Poisson equation for a sphere, and is therefore[35] the best form for a hypothetical perfectly spherical molecule. However, for realistic molecules, we find it to be slightly inferior to the conventional (Still's) form of f_{ij}^{GB} from Eq. (2). This is likely due to the fact that, for an elongated molecule, more of the electric field lines between a pair of distant charges go through the high dielectric region, effectively reducing the pairwise interaction. The conventional form of f_{ij}^{GB} takes this into account, at least to some extent, by allowing for steeper decay of the interaction with charge-charge distance. In what follows we will be using the conventional form of f_{ij}^{GB} from Eq. (2), unless otherwise specified.

To summarize, we suggest the following form for $\Delta G_{\text{el}}^{ij}=\frac{1}{2}\mathbf{F}(\mathbf{r}_i,\mathbf{r}_j)q_iq_j$ to be used for realistic molecules:

$$\Delta G_{\text{el}}^{ij}=-\frac{1}{2}\left(\frac{1}{\epsilon_{\text{in}}}-\frac{1}{\epsilon_{\text{out}}}\right)\frac{q_iq_j}{1+\alpha\epsilon_{\text{in}}/\epsilon_{\text{out}}} \times\left[\frac{1}{f_{ij}^{\text{GB}}(r_{ij},\tilde{R}_i,\tilde{R}_j)}+\frac{\alpha\epsilon_{\text{in}}/\epsilon_{\text{out}}}{A}\right], \tag{18}$$

where $f_{ij}^{\text{GB}}=\sqrt{r_{ij}^2+\tilde{R}_i\tilde{R}_j\exp(-r_{ij}^2/4\tilde{R}_i\tilde{R}_j)}$ for any i and j, including $i=j$. It has not escaped our notice that, while the GB theory itself has undergone considerable evolution since the early nineties when this particular form of f_{ij}^{GB} was first introduced by Still *et al.*, the function itself appears to remain superior or equally good compared to other forms that have been proposed in the literature.

Parameter α has been found by rigorous derivation and therefore it is *not* a free model parameter; it should be kept in mind though that the value $\alpha\approx 0.57$ minimizes the approximation error for very large spherical molecules with random charge distribution. For smaller molecules of complex shape or with peculiar charge distribution the value of α that brings about the best results may be slightly different. However, as will be shown below, Eq. (18) is not very sensitive to α, therefore $\alpha=0.57$ is a reasonable estimate for a generic case.

The key difference between the traditional form of $\Delta G_{ij}^{\text{GB}}$, Eq. (1), and our GB$\epsilon$ model based on expression Eq. (18) is that the latter has an extra prefactor and an additional term proportional to the ratio $\beta=\epsilon_{\text{in}}/\epsilon_{\text{out}}$ and inversely proportional to the electrostatic radius A. The presence of this term reflects the qualitative difference between the two models and explains why the traditional formulation does not capture the right dielectric dependence, especially when the ratio β is large. In fact, while the conventional formula agrees with the current model in the limit $\beta\to 0$, it is very much different in the other extreme, $\beta\to\infty$. In the latter case Eq. (18) provides the correct physical asymptotics:

$$\Delta G_{\text{el}}=\sum_{ij}\Delta G_{ij}^{\text{el}}\to\frac{1}{2A\epsilon_{\text{out}}}\sum_{ij}q_iq_j=\frac{1}{2A\epsilon_{\text{out}}}\left(\sum_i q_i\right)^2, \quad \beta\to\infty. \tag{19}$$

This agrees with our earlier observation that in this limit the electrostatic part of solvation energy does not depend on the charge distribution inside the molecule.

While Eq. (17) is an approximation to the exact solution for a sphere, one can show (see Appendix B) that it agrees with the exact formula given by Eq. (21) in the limits $\beta\ll 1$, $\beta\gg 1$, and, quite trivially, for $\beta=1$. Perhaps, it is then not so surprising that Eq. (17) gives an excellent agreement with the exact solution for a sphere [see Fig. 5(b)], while Eq. (18) provides a realistic approximation for globular molecules over the entire range of β (see Sec. IV). At the same time, the computational expense associated with Eq. (18) is not expected to be noticeably different from that of the conventional formula, Eq. (1).

The current model is expected to be effective in molecular dynamics (MD) calculations. While Eq. (18) is more accurate than the conventional Eq. (1), it differs from it in only two aspects: the constant prefactor, which has no effect at all on the computation speed, and the term $\alpha\beta/A$. The latter term depends only on the global shape of the molecule through A, which varies slowly and may need to be recalculated only once in a while, if at all. It may not even be unreasonable to assume $dA/dr_{ij}\approx 0$, and leave the corresponding term out of the force calculations completely, especially in the case of aqueous solvation $\beta\ll 1$. Thus, MD based on Eq. (18) may only be insignificantly slower, while essentially more accurate, than that utilizing the conventional GB of Eq. (1).

D. Exact solution for a spherical molecule

We have so far been able to sum the infinite series of the Kirkwood equation in an approximate fashion. While we will show in Sec. IV that this approximation is reasonable not only for a sphere but for realistic molecules as well, it is still desirable to see if one could obtain a well-behaved analytical Green's function without any approximations, or within such an approximation that allows full control of its accuracy and becomes exact in some parameter limit. Such a solution would provide a reference point for the current model, and is valuable even if its computational cost is relatively high. The obvious choice is to retain a large number of terms in the infinite series corresponding to Kirkwood's equations (8) and (9). In principle, the more terms that are kept, the closer the partial sum is to the exact one. However, Eqs. (8) and (9) cannot be used in this very form for this purpose because the series involved converge too slowly when t approaches 1, which happens when the atoms are close to the molecular surface (taken to coincide with the dielectric boundary). In fact, both of the Green functions in the original Kirkwood form of Eqs. (8) and (9) diverge in the limit $t\to 1$. One can understand the origin of this divergence by invoking the image charge considerations: as the charge approaches the di-

electric boundary, its image approaches the same boundary from the other side, causing the interaction between them to diverge.

It is therefore critical to improve convergence of the series by resumming them in an appropriate manner. A possible solution is given below, the derivation being detailed in Appendix B [see Eqs. (B13) and (B15)]:

$$\mathbf{F}(\mathbf{r}_i,\mathbf{r}_i)^{\text{sphere}} = -\frac{1}{\epsilon_{\text{in}} A t_{ii}} \frac{1-\beta}{1+\beta}\left[\frac{t_{ii}}{1-t_{ii}} - \frac{\beta}{1+\beta}\ln(1-t_{ii}) + \sum_{n=2}^{\infty}\left(\frac{\beta}{1+\beta}\right)^n \text{Li}_n(t_{ii})\right], \tag{20}$$

$$\mathbf{F}(\mathbf{r}_i,\mathbf{r}_j)^{\text{sphere}} = -\frac{1}{\epsilon_{\text{in}} A t_{ij}} \frac{1-\beta}{1+\beta}\left[\frac{t_{ij}}{\sqrt{1-2t_{ij}\cos\theta+t_{ij}^2}} + \frac{\beta}{1+\beta}\ln\frac{t_{ij}-\cos\theta+\sqrt{1-2t_{ij}\cos\theta+t_{ij}^2}}{1-\cos\theta} + \sum_{n=2}^{\infty}\left(\frac{\beta}{1+\beta}\right)^n Q_n(t_{ij},\cos\theta)\right], \tag{21}$$

where polylogarithm function $\text{Li}_n(t)$ and its generalization $Q_n(t,x)$ are defined in Appendix B [Eqs. (B6) and (B7)]. It is shown in Appendix B that Eqs. (20) and (21) coincide when $i=j$. Though Eq. (20) is a particular case of Eq. (21), the former is worth keeping for numerical implementation.

Series (20) and (21) converge for *any* β, the convergence being fastest for small β. This case corresponds to low-dielectric molecule immersed into high-dielectric solvent, e.g., a protein in water. In the opposite case of large β, which means $\epsilon_{\text{in}} \gg \epsilon_{\text{out}}$, a computationally efficient series representation also exists [see Appendix B, Eqs. (B25) and (B27)]:

$$\mathbf{F}(\mathbf{r}_i,\mathbf{r}_i)^{\text{sphere}} = -\frac{1}{\epsilon_{\text{in}} A}\frac{1-\beta}{(1+\beta)^2}\left[\beta + \frac{1}{1-t_{ii}} - \frac{\beta}{1+\beta}\ln(1-t_{ii}) - \beta\sum_{n=2}^{\infty}\frac{(-1)^n}{(1+\beta)^n}\text{Li}_n(t_{ii})\right], \tag{22}$$

$$\mathbf{F}(\mathbf{r}_i,\mathbf{r}_j)^{\text{sphere}} = -\frac{1}{\epsilon_{\text{in}} A}\frac{1-\beta}{(1+\beta)^2}\left[\beta + \frac{1}{\sqrt{1-2t_{ij}\cos\theta+t_{ij}^2}} - \frac{\beta}{1+\beta}\ln\frac{1-t_{ij}\cos\theta+\sqrt{1-2t_{ij}\cos\theta+t_{ij}^2}}{2} - \beta\sum_{n=2}^{\infty}\frac{(-1)^n}{(1+\beta)^n}Q_n^{\dagger}(t_{ij},\cos\theta)\right], \tag{23}$$

where function $Q_n^{\dagger}(t,x)$ is defined in Appendix B [Eq. (B19)]. Equations (22) and (23) are optimized for large β and converge for any positive β. Again, Eqs. (22) and (23) coincide for $i=j$.

No approximations have been made so far in derivation of the optimized equations. Both pairs of Eqs. (20), (21) and (22), (23) are still exact and mathematically equivalent to Eqs. (8) and (9) for a spherical molecule. The corresponding series converge for any β, θ, and any $t_{ij}<1$, and can be well approximated by partial sums with a reasonable number of terms kept. Varying the number of terms retained, one can adjust the accuracy or time of computation. The singularities of Kirkwood's original formulas [Eqs. (8) and (9)] are now localized within the first two terms of the each regularized equation. Both of these terms are given by simple, closed-form expressions, well suited for numerical computations. The functions $\text{Li}_n(t)$, $Q_n(t,x)$, and $Q_n^{\dagger}(t,x)$ do not have any singularities for $n\geqslant 2$ [see Eq. (B20)] and can be computed numerically with any desirable accuracy via recursive relationships, Eqs. (B10) and (B23). We find that estimating ΔG_{el} through Eqs. (20) and (21) [or Eqs. (22) and (23)] is only a few times more costly than through the conventional GB formula, Eq. (1), but captures the key physics of the dielectric response missed by the conventional model. The existence of well-converging equations proves critical for obtaining a fast yet accurate numerical solution.

III. METHODS

A. Structures

We have carried performance tests on macromolecules representing different structural classes: native myoglobin (PDB ID 2MB5), thioredoxin (2TRX), hen egg lysozyme (0LZ5), β-hairpin (2GB1), and B-DNA (10 base-pair duplex). We have also used an artificial spherical molecule referred to as "sphere-15 Å." This is a sphere of radius 15 Å filled with uniform dielectric ϵ_{in} with 11 charges $q_i=+1$ located at points $(\pm 6,0,0)$, $(0,\pm 6,0)$, (0,0,6), $(\pm 12,0,0)$, $(0,\pm 12,0)$, and $(0,0,\pm 12)$ (all coordinates are in angstroms). Since the Kirkwood equation is exact for sphere, we are able to use sphere-15 Å as a benchmark to test both the GB and PB numerical methods.

B. Poisson–Boltzmann calculations

Numerical Poisson–Boltzmann solvers are used in two ways here. To calculate the electrostatic part of solvation free energy of each test molecule, MEAD (Ref. 36) and DELPHI-II (Refs. 25 and 37) have been used, with a cubic box and 0.25 Å grid spacing. The convergence criterion used by MEAD is set to its default value. Ionic strength is zero. Unless otherwise stated, all of the GB vs PB comparisons are done using MEAD energies as reference.

To compute the perfect effective Born radii of each atom in the test structure, a Poisson problem is set up and solved having the dielectric boundary shape of the full molecule present, but keeping only the charge of that particular atom. The van der Waals radii of Bondi[38] and a solvent-probe radius of 1.4 Å are used to define molecular surface,[39] which is taken as the dielectric boundary. The accumulation of these solutions gives the necessary Green's function information for use in Eq. (6) for the full Poisson solvation energy and Eq. (3) for the perfect effective Born radii. The computer program PEP developed by Beroza[26] and available via the Internet (ftp://ftp.scripps.edu/case/beroza/pep) is used for the setup and solution of these Poisson problems. The finest grid spacing used in all calculations is 0.07 Å, decreasing from 4 Å in eight steps of focusing on the atom in question. We

TABLE I. Exact (PB-based) and heuristic electrostatic radii of some molecules and respective solvation energies.

	A		ΔG_{el}	
Molecule	PEP	Heuristic	PB	Heuristic A
Myoglobin	19.0	18.5	−2969.2	−2969.4
DNA	13	13.3	−4701.8	−4701.0
Thioredoxin	17.7	15.3	−1559.5	−1559.7
Lysozyme	18.2	24.3	−2053.1	−2051.6

have verified, by comparison with the exact solution on a perfect sphere, that PEP-based calculations do produce physically reasonable results even in the case $\epsilon_{in} \gg \epsilon_{out}$, where the two other methods failed. However, the PEP algorithm is so computationally intense that we did not use it as a reference for all values of ϵ_{in} and ϵ_{out}. For a perfect sphere of given radius A, the effective radii are computer exactly via Eq. (13).

C. Simple estimation of the effective electrostatic radius

The principal way of calculating the effective electrostatic radii is given by Eq. (10), and this is how they are computed in this work (A is the limiting value of the effective Born radii for $\beta \gg 1$). However, the PEP calculations involved are intense, and it would be useful to have an analytical way of estimating A. That will involve the same kind of derivation, based on the expressions for the electrostatic energy density, that leads to a closed form expression for the effective Born radii.[8] Here we do not attempt to make the formal derivations; rather, we propose a simple and quick, albeit heuristic way to obtain a rough estimate of the effective electrostatic radius based on simple physical considerations. We will also show directly that, at least in the case of aqueous solvation of realistic biomolecules, the resulting solvation energy is rather insensitive to variations in A.

For a typical globular molecule, the solvation energy depends on its overall radius, $\Delta G_{el} \sim 1/R$. In contrast, for a thin and long cylindrical structure (e.g., the DNA), the characteristic length scale is the cylinder radius, $\Delta G_{el} \sim 1/r$. Generally, one may consider that a cylinder of height h and radius r circumscribes the molecule in question. The formula $1/A = 1/2[1/r + 1/(1/2)h]$ gives back the radius of the sphere for a spherical molecule and r for a long cylindrical one. To use this heuristics in practice on realistic structures, we first define a trend line to act as the axis of rotation for the cylinder. Next, the radius of the cylinder is calculated by computing the distance from each atom to the axis. We find that defining the radius to include only 75% of the atoms within the cylinder gives better results compared to the rigorous electrostatic estimates of A. The top and bottom sides are defined so as to include all atoms, thus defining the cylinder height h.

The arguments and the formula above are not meant to be precise, but to give a quick way of estimating A. However, when compared to the correct PEP values, the results are reasonable, as shown in Table I. The ΔG_{el} values are computed using $\epsilon_{in}=1$ and $\epsilon_{out}=80$, with either the numerical PB or GBϵ [Eq. (18)] using the heuristic values of A.

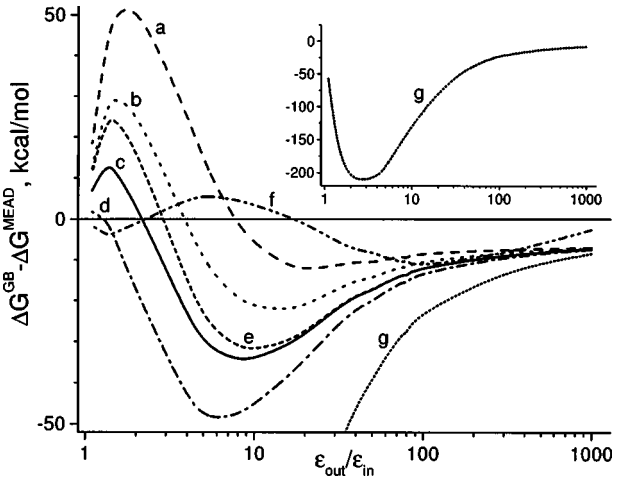

FIG. 4. Solvation energies for myoglobin: difference between various methods and the particular numerical PB used as reference (MEAD). (a)–(d) GBϵ: (a) $\alpha=0.75$, (b) $\alpha=0.65$, (c) $\alpha=0.57$, (d) $\alpha=0.50$; (e) GBϵ(exact), (f) another popular PB solver (DELPHI), (g) conventional GB.

IV. RESULTS AND DISCUSSION

The goal of this section is to assess the accuracy of the GBϵ model relative to the PB treatment, and compare its performance with the conventional GB model. To this end, the electrostatic part of the solvation free energy is calculated for realistic molecules representing various structural classes: DNA, myoglobin, thioredoxin, lysozyme, and β-hairpin, as well as an artificial perfectly spherical "molecule" of radius 15 Å (sphere-15 Å). All the structures and details of the calculations were described in the "Methods." We use the following naming convention here when referring to the GB models considered here: (i) GBϵ(exact)—the method based on Eq. (21) (for $\beta<1$) or Eq. (23) (for $\beta>1$). Strictly speaking, the model is only exact for a perfect sphere, in the limit of infinite number of terms kept. In practice, the terms are summed up until the result is essentially exact on a sphere (to within 10^{-6} of the exact solution); (ii) GBϵ—the approximation given by Eq. (18) (with $\alpha=0.57$ unless otherwise stated); and (iii) the conventional GB method [Eq. (1)].

A. Effect of parameter α on GBϵ accuracy

Before applying the GBϵ model to the various structures described above, we want to see how the value of the parameter α, theoretically found to be $\alpha \approx 0.57$ in the limit of an infinitely large molecule [see Appendix A, Eqs. (A13)–(A15)], works for a realistic biomolecule of finite size—myoglobin (see Fig. 4). In particular, we explore the sensitivity of the solvation energy to variation of α. As suggested by Eq. (14), reasonable values of α may be between 0.5 and 1. For comparison, we also compute ΔG^{GB} using GBϵ(exact) and conventional GB, the latter being equivalent to GBϵ with $\alpha=0$.

The results (Fig. 4) indicate that GBϵ with $\alpha \approx 0.57$ generates a plot very close to both the numerical PB and GBϵ(exact). Over a range of β values the difference between GBϵ and PB (MEAD) is comparable to the difference between solvation energies predicted by two different PB solvers: DELPHI-II and MEAD, especially around $\beta=1/80$ cor-

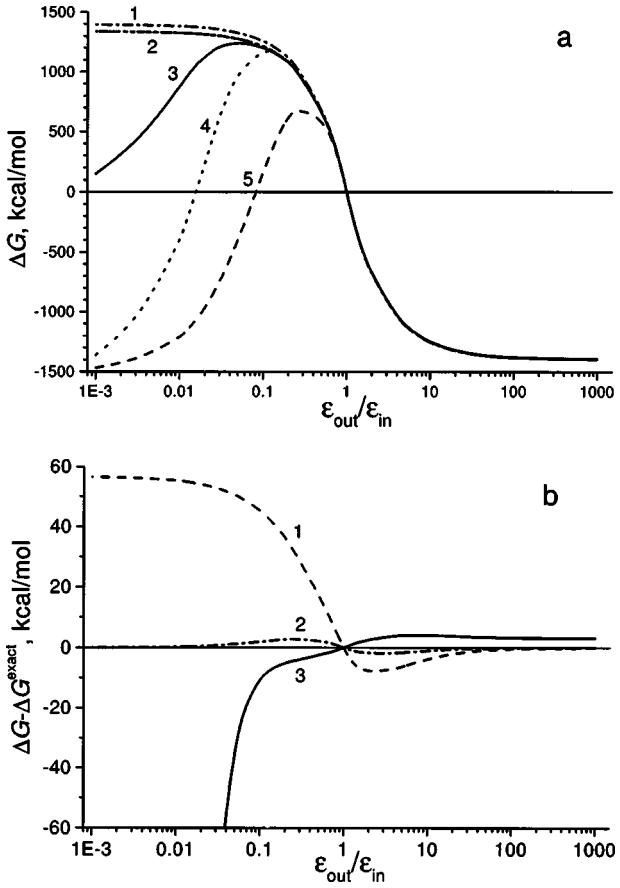

FIG. 5. Solvation energy for sphere-15 Å: (a) calculated by (1) conventional GB, (2) the exact solution and GBϵ with $\alpha=0.57$ (the curves virtually coincide), (3) MEAD, (4) DELPHI with 5000 maximum iterations, (5) DELPHI with 1000 maximum iterations; (b) difference between exact values and (1) conventional GB, (2) GBϵ with $\alpha=0.57$, (3) MEAD.

responding to the typical case of aqueous solvation. At the same time, the conventional GB shows considerable deviation form the PB.

A conclusion can be made that our "first principle" value of $\alpha\approx 0.57$ is a reasonable choice for globular molecules. This result is confirmed by further analysis presented below.

B. Solvation energies for an ideal sphere

Since Kirkwood's equation is exact for a sphere, we can use the model molecule sphere-15 Å described above to test not only the GB but also the numerical PB solvers. In this section, ΔG_{el} is calculated using the GBϵ method as well as two conventional numerical PB packages: MEAD and DELPHI-II. Note that GBϵ(exact) is actually exact for a sphere because it is mathematically equivalent to Kirkwood's equation (within the desired accuracy). Therefore, the GBϵ(exact) data set is used here as a reference. The results are shown in Fig. 5. All methods exhibit good agreement with the exact solution in the $\epsilon_{\mathrm{out}}>\epsilon_{\mathrm{in}}(\beta^{-1}>1)$ domain. The conventional GB deviates from the exact solution in the opposite limit, which is best seen in Fig. 5(b). The physically important asymptotics $\Delta G^{\mathrm{solv}}\sim(\Sigma q_i)^2/A$ in the limit $\beta\to\infty$ is not met by conventional GB equation. Curiously, both of the numerical PB methods used here (with the reasonable input parameters and default convergence criteria) fail to produce even a qualitative correspondence with the exact solution in this limit. Exploring the origins of this behavior is beyond the scope of this paper; it is possible that modifications to the corresponding algorithms can alleviate the problem. At the same time, the GBϵ curve can only be distinguished from the exact solution on the fine-scale plot [see Fig. 5(b)]. The GBϵ plot coincides with GBϵ(exact) at $\beta\to 0$, $\beta=1$, and $\beta\to\infty$, and is closer to the exact solution than any other method—including numerical PB—considered here in the entire possible range of β. Of course, one should keep in mind that this strong statement is only valid for a perfectly spherical molecule.

C. Performance of the GBϵ model on realistic molecules

We continue to explore the performance of the GBϵ model on realistic biomolecules, Fig. 6. Since no exact solution of the PB equation is available for arbitrarily shaped biomolecules, we use numerical PB (MEAD) as reference. For each structure, we compute ΔG_{el} with GBϵ using the first principle $\alpha=0.57$, as well as the value of α that gives the best agreement with the numerical PB for the given molecule. We also compare GBϵ and GBϵ(exact), which allows us to answer the following important question: is it likely that we can improve GBϵ for an arbitrary realistic molecule by keeping higher order terms in Kirkwoods's exact solution before making the key approximation that allowed us to sum the corresponding infinite series and arrive at Eq. (18)? Needless to say, GBϵ(exact) is not exact in this case, but it is the best one can do assuming underlying spherical geometry.

As seen from Fig. 6, the GBϵ methods give equally good agreement with the numerical PB for both very large β^{-1} and for $\beta\to 1$. (Note that since we have found that the reference PB solvers we use here are inadequate for $\beta>1$, we do not consider this domain for the realistic molecules.) The curves resulting from the best choice of α for the given molecule are close to those obtained with the first principle $\alpha=0.57$, suggesting that the latter is suitable for the generic case. Setting α to certain particular value for all structures makes the GBϵ model parameter-free.

Meanwhile, the conventional GB method gives a large error in the range of $2\lesssim\beta^{-1}\lesssim 20$. Note that while for three of the four structures the GBϵ(exact) method is most accurate of the three GB methods tested, it is not the case for lysozyme, Fig. 6(d). This may not be so surprising given that among the four structures lysozyme differs most from the underlying model: while the other three structures are basically convex, lysozyme has a distinct binding pocket. The important conclusion here is that we are reaching the limit of the accuracy here, at least within the GB model based on Green's function derived for a perfect sphere: the extra computational expense of the more elaborate GBϵ(exact) summation may not pay off. However, the GBϵ(exact) method has a clear way of controlling its accuracy, at least for molecules whose shape is close to spherical: the more terms in the series Eqs. (21) are retained, the higher is the accuracy, albeit at a larger computational cost. Still, while the GBϵ(exact) method is mathematically equivalent to Kirk-

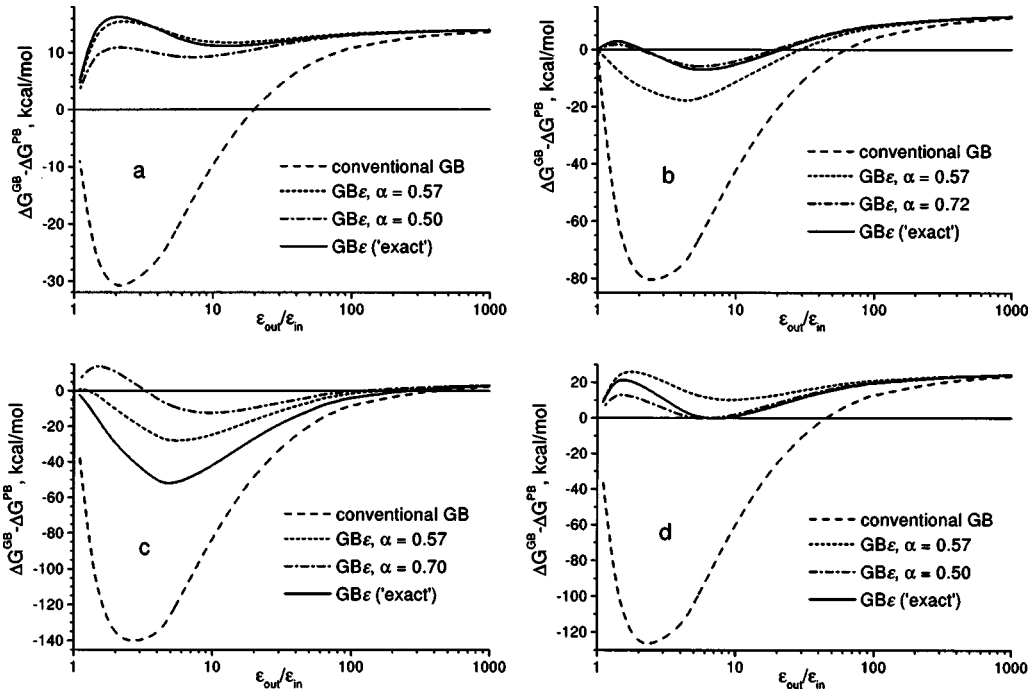

FIG. 6. Solvation energy for (a) β-hairpin, (b) DNA, (c) lysozyme, and (d) thioredoxin calculated by conventional GB, GBϵ(exact) and GBϵ methods, relative to numerical PB.

wood's original solution, the series in Eqs. (20)–(23) converge much faster than those in Eqs. (8) and (9), especially for small or large β. Therefore, a relatively small number of terms may be sufficient, leading to reasonable computational costs.

D. Effective dielectric constant: Can it be larger than ϵ_{out}?

Consider two charges immersed in low dielectric medium ϵ_{in} surrounded by high-dielectric medium ϵ_{out}. Part of the electrostatic field lines are located completely within the low-dielectric medium, while the other lines lie partly beyond it. Then it could be expected that the Coulomb interaction energy of these charges should be between $q_i q_j / r_{ij} \epsilon_{out}$ and $q_i q_j / r_{ij} \epsilon_{in}$. In other words, it seems natural for the charges to "feel" some effective medium with dielectric constant ϵ_{eff} such that $\epsilon_{in} < \epsilon_{eff} < \epsilon_{out}$.

It has been observed,[8] however, that in PB calculations ϵ_{eff} may be larger than ϵ_{out} for some charge pairs. Figure 7 illustrates the reason for such a behavior. If $\epsilon_{in} \ll \epsilon_{out}$, the electric field lines are "drawn" into the surrounding high-dielectric medium, thus shifting ϵ_{eff} towards higher values. In addition, the lines go a much longer way in ϵ_{out} medium, which makes the charges feel as if they were farther apart than they actually are.

The ability of an approximate theory to reproduce this nontrivial behavior may be used as a test of the theory's physical consistency. Let us compare conventional GB and GBϵ methods in this respect. Based on Coulomb's law, we define ϵ_{eff} as follows:

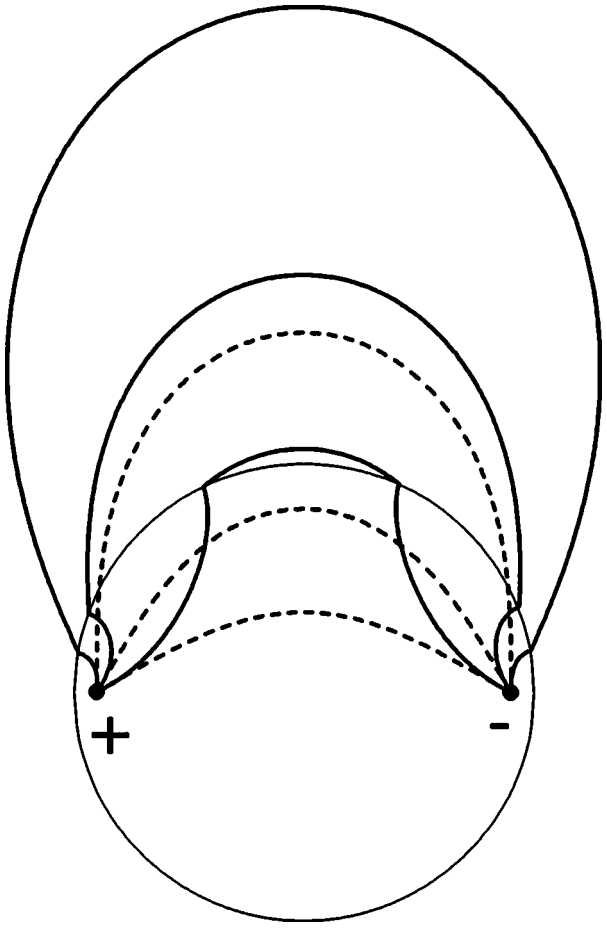

FIG. 7. Electric field lines between two charges, $+q$ and $-q$, located inside a spherical "molecule" with dielectric ϵ_{in} surrounded by medium with dielectric ϵ_{out}. Solid lines, $\beta = 1/80$; dashed lines, $\beta = 1$, e.g., infinite uniform medium.

$$\frac{q_i q_j}{\epsilon_{\text{eff}} r_{ij}} = \frac{q_i q_j}{\epsilon_{\text{in}} r_{ij}} + \Delta G^{ij} = \frac{q_i q_j}{\epsilon_{\text{in}} r_{ij}} + q_i q_j \mathbf{F}(\mathbf{r}_i, \mathbf{r}_j), \tag{24}$$

where we note that the full interaction energy is a sum of the vacuum and solvation parts.

If the conventional GB method is used [see Eq. (1)], then

$$\epsilon_{\text{eff}} = \frac{\epsilon_{\text{in}}}{1 - (1-\beta)\omega}, \quad \omega = \frac{r_{ij}}{f_{ij}^{\text{GB}}}. \tag{25}$$

Since $r_{ij} \leqslant f_{ij}^{\text{GB}}$ [see Eq. (2)], it follows that $\omega \leqslant 1$ and therefore $\epsilon_{\text{eff}} \leqslant \epsilon_{\text{out}}$. So, the conventional GB method cannot yield values $\epsilon_{\text{eff}} > \epsilon_{\text{out}}$, in principle.

Meanwhile, when $\mathbf{F}(\mathbf{r}_i, \mathbf{r}_j)$ is calculated using Eq. (17) or Eq. (18),

$$\omega = \left(\frac{r_{ij}}{f_{ij}^{\text{GB}}} + \alpha\beta \frac{r_{ij}}{A} \right) / (1 + \alpha\beta)$$

$$= \frac{r_{ij}}{f_{ij}^{\text{GB}}} + \frac{\alpha\beta}{1+\alpha\beta} \left(\frac{r_{ij}}{A} - \frac{r_{ij}}{f_{ij}^{\text{GB}}} \right). \tag{26}$$

So, if $f_{ij}^{\text{GB}} > A$, it may happen that $\omega > 1$. This means that GBϵ method may exhibit $\epsilon_{\text{eff}} > \epsilon_{\text{out}}$ in contrast to conventional GB. For example, the highest ϵ_{eff} for a sphere is achieved when the two atoms are at the largest possible separation of $r_{ij} = 2A$, in which case

$$\frac{\epsilon_{\text{eff}}}{\epsilon_{\text{out}}} = \left(1 - \alpha \frac{1-\beta}{1+\alpha\beta} \right)^{-1}. \tag{27}$$

In the typical case of $\epsilon_{\text{in}} = 1$, $\epsilon_{\text{out}} = 80$, and with $\alpha = 0.57$, the maximum value of ϵ_{eff} predicted by the GBϵ is ca. 180. The fact that ϵ_{eff} may exceed ϵ_{out} within the GBϵ model has also been confirmed (results not shown) by direct calculations for the realistic test molecules described above.

E. Do effective Born radii depend on dielectric?

The key premise of the GB method is that it yields a good estimate of the electrostatic part of the solvation free energy provided that the effective Born radii R_i are correct.[31] However, when one calculates the solvation energy as a function of the solvent/solute dielectric, a natural question arises: should the effective radii depend on ϵ_{in} and ϵ_{out}? As we have already seen, the derivation of the GBϵ theory introduced here implies that the effective radii $\tilde{R}$ need to be computed only once at $\beta = \epsilon_{\text{in}}/\epsilon_{\text{out}} \to 0$, and using an ϵ-dependent set of effective radii would be wrong. Still, one may wonder if the use of ϵ-dependent radii may rescue the conventional GB theory based on Eqs. (1) and (2), yielding an approximation for ΔG_{el} over the entire range of ϵ_{in} and ϵ_{out} as good as the GBϵ, but without the extra physics of GBϵ? To answer this question we have computed ΔG_{el} for a perfect sphere using GBϵ method [Eq. (17)] and the conventional theory [Eq. (1)] with ϵ-dependent set of effective radii, comparing both results to the exact solution available in this case (see Fig. 8). To obtain the explicit dependence of the R_i on the dielectric constants, we first compute the exact $\Delta G_{ii}^{\text{el}}$ via Kirkwood's solution and then use this value to obtain R_i (β) from Eq. (3).

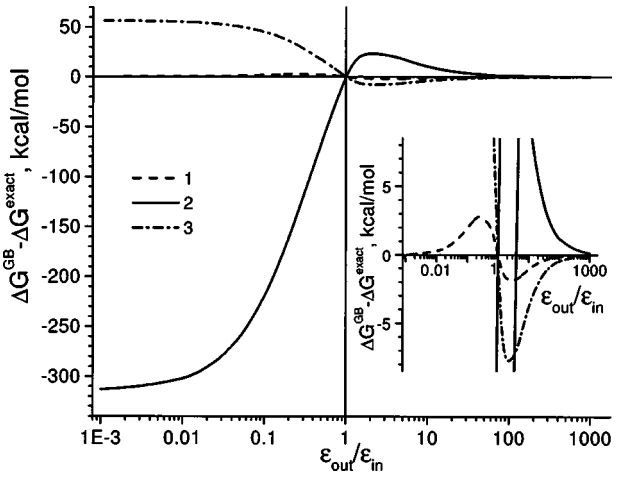

FIG. 8. Solvation energies for sphere-15 Å relative to exact treatment: (1) GBϵ, (2) conventional GB with ϵ-dependent set of effective radii, (3) conventional GB with the same set of fixed effective radii as in (1) (computed exactly for $\epsilon_{\text{in}} = 1$, $\epsilon_{\text{out}} \to \infty$). Since GB$\epsilon$ curve almost coincides with the axis, same curves are presented in the inset to show details.

The results shown in Fig. 8 clearly indicate that the use of ϵ-dependent radii set does not improve the performance of the conventional GB formula—its accuracy in computing ΔG_{el} is far worse than that of the GBϵ. If anything, using $\tilde{R}_i = R(\beta \to 0)$ in the conventional formula gives better results than the ϵ-dependent R_i's (dashed-dotted curve in Fig. 8), although still far from GBϵ and the exact solution. As previously stated, this is because the conventional formula [Eq. (1)] lacks some essential physics. In fact, one can easily see that Eq. (1) is the asymptotic case of the GBϵ [Eq. (17)] in the limit $\epsilon_{\text{in}}/\epsilon_{\text{out}} \to 0$, but as $\epsilon_{\text{in}}/\epsilon_{\text{out}}$ becomes finite, the conventional theory deviates from the exact solution. In contrast, GBϵ is always close to it, coinciding with the exact solution in both limits $\epsilon_{\text{in}}/\epsilon_{\text{out}} \to \infty$ and $\epsilon_{\text{in}}/\epsilon_{\text{out}} \to 0$.

While ϵ-independent effective Born radii is the correct choice for globular molecules and one dielectric boundary, this may or may not be the case in systems of more complex geometry or more than two dielectric media. In particular, a GB theory with ϵ-dependent R_i's was reported as being successfully applied to modeling of peptides and proteins in presence of a membrane.[16,40]

V. CONCLUSION

In this work we have proposed a generalized Born model GBϵ applicable in the entire range of solvent/solute dielectrics. The model contains no fitting parameters, and its main formula is derived as an approximation to the exact Green function for a perfect sphere. Relative to the conventional GB model based on Still's equation, the present approximation has only one extra term. However, unlike the conventional GB model, our model captures the essential physics of the dielectric response for all values of ϵ_{in} and ϵ_{out}.

The GBϵ model is first tested on a charge distribution inside a perfect sphere of 15 Å radius: the solvation energies agree to within 3 kcal/mol with the exact values obtained from the Kirkwood solution over the entire range of solute and solvent dielectrics. Curiously, we have found that some conventional PB solvers, such as MEAD or DELPHI, predict

unphysical solvation energies in the domain $\epsilon_{out} < \epsilon_{in}$. We have continued testing the current GB model against numerical PB on realistic molecules representing various structural classes: myoglobin, B-DNA, β-hairpin, lysozyme, and thioredoxin. In all cases, the current GB model has outperformed the conventional model considerably. Notably, in the most important region $\epsilon_{out} \gg \epsilon_{in}$, the error in the calculated solvation energy, relative to the PB solver used as a reference (MEAD), is comparable to the difference between the values predicted by another common solver (DELPHI) and the same reference.

Another line of evidence in support of the claim that some essential physics has been introduced into the current model comes from the fact that, unlike the conventional GB formula, the GBϵ model is capable of reproducing the unusually high values of effective dielectric constants. It has been known for a long time that the PB equation predicts, somewhat counterintuitively, values of the effective dielectric constant (between some pairs of charges in molecules) to be larger than that of the solvent dielectric. The inability of the conventional model to account for that behavior often served as a basis of critiques of its underlying physical principles. We have shown how these high effective values are generated within the current GBϵ model, and explained their physical origin.

Finally, we note that the current model is particularly well suited to be the basis of the implicit solvent representation in molecular dynamics simulations. Its formula is no more computationally complex than the conventional GB formula which has been successfully used in many popular MD packages.

ACKNOWLEDGMENT

The authors are thankful to Andrew Fenley for useful comments.

APPENDIX A: DERIVATION OF OPTIMAL α

Let us find the *best*, in a sense described below, value α to be used in approximate Eq. (17) or Eq. (18). Denote the exact (e) and approximate (a) series in Eq. (14) as

$$g_{ij}^{(e)} = \sum_{l=0}^{\infty} \frac{t_{ij}^l P_l(\cos\theta)}{1 + \frac{l}{l+1}\beta},$$

$$g_{ij}^{(a)} = \frac{1}{1+\alpha\beta}\left[\sum_{l=0}^{\infty} t_{ij}^l P_l(\cos\theta) + \alpha\beta\right]. \tag{A1}$$

While α may be found exactly for every particular pair (i,j) from equation $g_{ij}^{(e)} = g_{ij}^{(a)}$, all such α's will be different in a general case. Since we need one universal constant value instead, we will seek α such that the square-average error $\langle (g_{ij}^{(e)} - g_{ij}^{(a)})^2 \rangle$ is minimized. The average is taken on a sphere of unit radius with a weight function $w(t,\theta) = \pi(1-t^2)\sin\theta$ to take into account the system's geometry and symmetry:

$$\langle (g_{ij}^{(e)} - g_{ij}^{(a)})^2 \rangle = \pi \int_0^1 dt \int_0^{\pi} d\theta (1-t^2)\sin\theta (g_{ij}^{(e)} - g_{ij}^{(a)})^2. \tag{A2}$$

Atom sizes are neglected in this derivation, which makes the result exact in the limit of a large molecule: (molecule size)/(atom size) $\to \infty$. It should be therefore a reasonable approximation for large biomolecules such as proteins or DNA.

To find the error's minimum, we differentiate the penalty function:

$$\begin{aligned}\frac{\partial}{\partial\alpha}[(g_{ij}^{(e)} - g_{ij}^{(a)})^2] &= -2(g_{ij}^{(e)} - g_{ij}^{(a)})\frac{\partial}{\partial\alpha} g_{ij}^{(a)} = -2(g_{ij}^{(e)} - g_{ij}^{(a)}) \\ &\times \left[\frac{\beta}{1+\alpha\beta} - \frac{g_{ij}^{(a)}(1+\alpha\beta)\beta}{(1+\alpha\beta)^2}\right] \\ &= \frac{2\beta}{1+\alpha\beta}(g_{ij}^{(e)} - g_{ij}^{(a)})(g_{ij}^{(a)} - 1) \\ &= 0. \end{aligned} \tag{A3}$$

In the explicit form

$$g_{ij}^{(e)} - g_{ij}^{(a)} = \sum_{l=1}^{\infty} t_{ij}^l P_l(\cos\theta)\left(\frac{1}{1+\frac{l}{l+1}\beta} - \frac{1}{1+\alpha\beta}\right), \tag{A4}$$

$$g_{ij}^{(a)} - 1 = \frac{1}{1+\alpha\beta}\sum_{m=1}^{\infty} t_{ij}^m P_m(\cos\theta), \tag{A5}$$

therefore the master equation for α takes the following closed form upon dropping constant multipliers:

$$\int_0^1 dt \int_0^{\pi} d\theta (1-t^2)\sin\theta \left[\sum_{l=1}^{\infty}\sum_{m=1}^{\infty} t^{l+m} P_l(\cos\theta) P_m(\cos\theta) \times \left(\frac{1}{1+\frac{l}{l+1}\beta} - \frac{1}{1+\alpha\beta}\right)\right] = 0. \tag{A6}$$

Since the variables t and θ are separable, one can easily integrate to find

$$\begin{aligned}\int_0^1 dt(1-t^2)t^{l+m} &= \frac{1}{l+m+1} - \frac{1}{l+m+3} \\ &= \frac{2}{(l+m+1)(l+m+3)}, \end{aligned} \tag{A7}$$

TABLE II. Values of α for select β.

β	1/1000	1/80	1/10	1/2	2	10	80	1000
α	0.580 112	0.579 941	0.578 725	0.574 613	0.567 834	0.561 772	0.559 536	0.559 200

$$\int_0^\pi d\theta P_l(\cos\theta)P_m(\cos\theta)\sin\theta = \frac{2}{l+m+1}\delta_{lm}, \tag{A8}$$

where δ_{lm} is the Kronecker's delta. Thus, the double sum in Eq. (A6) is reduced to a single series

$$\sum_{l=1}^{\infty} \frac{2}{(2l+1)(2l+3)} \frac{2}{2l+1}\left(\frac{1}{1+\frac{l}{l+1}\beta} - \frac{1}{1+\alpha\beta}\right) = 0. \tag{A9}$$

Let us rewrite this equation as $S(\beta) - s/(1+\alpha\beta) = 0$, where

$$S(\beta) = \sum_{l=1}^{\infty} \frac{1}{(2l+1)^2(2l+3)} \frac{1}{1+\frac{l}{l+1}\beta} = \sum_{l=1}^{\infty} \frac{l+1}{(2l+1)^2(2l+3)(l+1+l\beta)}$$
$$= \frac{(1-\beta)\left[s+\beta\left(3s+\frac{4}{3}\right)+\beta^2\right] + \beta(1+\beta)\left[\psi\left(\frac{1}{2}\right) - \psi\left(\frac{1}{1+\beta}\right)\right]}{(1-\beta)^2(1+3\beta)}, \tag{A10}$$

$$s = \sum_{l=1}^{\infty} \frac{1}{(2l+1)^2(2l+3)} = \frac{\pi^2}{16} - \frac{7}{12} \approx 0.033\,516\,9, \tag{A11}$$

and $\psi(z)$ is the logarithmic derivative of gamma function, $\psi(z) = \Gamma'(z)/\Gamma(z)$. Finally we obtain

$$\alpha = \frac{1}{\beta}\left(\frac{s(1-\beta)^2(1+3\beta)}{(1-\beta)\left[s+\beta\left(3s+\frac{4}{3}\right)+\beta^2\right] + \beta(1+\beta)\left[\psi\left(\frac{1}{2}\right) - \psi\left(\frac{1}{1+\beta}\right)\right]} - 1\right). \tag{A12}$$

In the key limit cases the above expression is simplified as

$$\alpha(\beta\to 0) = \frac{32(3\ln 2 - 2)}{3\pi^2 - 28} - 1 \approx 0.580\,127, \tag{A13}$$

$$\alpha(\beta\to 1) = \frac{4(3\pi^2 - 28)}{42\zeta(3) + 3\pi^2 - 76} - 1 \approx 0.571\,412, \tag{A14}$$

$$\alpha(\beta\to \infty) = \frac{3(3\pi^2 - 28)}{164 - 9\pi^2 - 96\ln 2} \approx 0.559\,170, \tag{A15}$$

where $\zeta(n)$ is Riemann's zeta function. Function $\alpha(\beta)$ is monotonous and has no singularities. More numerical values are presented in Table II. As one can see, α varies only slightly over the entire range of β values.

APPENDIX B: SUMMATION OF KIRKWOOD SERIES

According to Kirkwood,[34] the electrostatic energy from a pair of interacting charges to solvation energy of a spherical molecule is

$$\Delta G_{ij}^{\mathrm{el}} \sim \sum_{l=0}^{\infty} \frac{t^l P_l(x)}{1+\frac{l}{l+1}\beta}, \tag{B1}$$

where $t = r_i r_j / A^2, 0 < t < 1, x = \cos\theta, \theta$ being the angle between vectors $\mathbf{r}_i$ and $\mathbf{r}_j$ (see Fig. 2), and $\beta = \epsilon_{\mathrm{in}}/\epsilon_{\mathrm{out}} > 0$; $P_l(x)$ is Legendre's polynomial. In the case of self-term contribution, the series is as follows:

$$\Delta G_{ii}^{\mathrm{el}} \sim \sum_{l=1}^{\infty} \frac{t^l}{1+\frac{l}{l+1}\beta}. \tag{B2}$$

Note that both series diverge at $t\to 1$, thus precluding effective numerical implementation of the equations. The goal of this appendix is to regularize the series in Eqs. (B1) and (B2) for the purpose of numerical calculations.

1. Case $\epsilon_{\mathrm{in}} < \epsilon_{\mathrm{out}}$

First, consider the case $\beta = \epsilon_{\mathrm{in}}/\epsilon_{\mathrm{out}} < 1$. Let us develop the following Taylor's series:

$$\frac{1}{1+\frac{l}{l+1}\beta} = \frac{1}{1+\beta-\frac{\beta}{l+1}}$$
$$= \frac{1}{1+\beta}\frac{1}{1-\frac{1}{l+1}\frac{\beta}{1+\beta}}$$
$$= \frac{1}{1+\beta}\sum_{n=0}^{\infty}\left(\frac{\beta}{1+\beta}\right)^n \frac{1}{(l+1)^n}. \tag{B3}$$

Substitution of Eq. (B3) into Eq. (B2) yields

$$\sum_{l=0}^{\infty}\frac{t^l}{1+\frac{l}{l+1}\beta}=\frac{1}{1+\beta}\sum_{l=0}^{\infty}t^l\sum_{n=0}^{\infty}\left(\frac{\beta}{1+\beta}\right)^n\frac{1}{(l+1)^n}$$

$$=\frac{1}{(1+\beta)t}\sum_{n=0}^{\infty}\sum_{l=0}^{\infty}\left(\frac{\beta}{1+\beta}\right)^n\frac{t^{l+1}}{(l+1)^n}$$

$$=\frac{1}{(1+\beta)t}\sum_{n=0}^{\infty}\left(\frac{\beta}{1+\beta}\right)^n\mathrm{Li}_n(t), \tag{B4}$$

where $\mathrm{Li}_n(t)$ is the polylogarithm function

$$\mathrm{Li}_n(t)=\sum_{l=0}^{\infty}\frac{t^{l+1}}{(l+1)^n}. \tag{B5}$$

In particular, $\mathrm{Li}_0(t)=t/(1-t)$ and $\mathrm{Li}_1(t)=-\ln(1-t)$. For the pair interaction term $\Delta G_{ij}^{\mathrm{el}}$ we substitute Eq. (B3) into Eq. (B1) to obtain

$$\sum_{l=0}^{\infty}\frac{t^lP_l(x)}{1+\frac{l}{l+1}\beta}=\frac{1}{(1+\beta)t}\sum_{n=0}^{\infty}\sum_{l=0}^{\infty}\left(\frac{\beta}{1+\beta}\right)^n\frac{t^{l+1}P_l(x)}{(l+1)^n}$$

$$=\frac{1}{(1+\beta)t}\sum_{n=0}^{\infty}\left(\frac{\beta}{1+\beta}\right)^nQ_n(t,x). \tag{B6}$$

Here, for the sake of brevity, we have introduced a function

$$Q_n(t,x)=\sum_{l=0}^{\infty}\frac{t^{l+1}P_l(x)}{(l+1)^n}. \tag{B7}$$

Though there is no general simplification for this series, $Q_0(t,x)$ and $Q_1(t,x)$ may be found as follows:

$$\frac{1}{t}Q_0(t,x)=\sum_{l=0}^{\infty}t^lP_l(x)=\frac{1}{\sqrt{1-2xt+t^2}}, \tag{B8}$$

which is a well-known relation for Legendre's polynomials. Now, note that

$$\frac{d}{dt}Q_n(t,x)=\sum_{l=0}^{\infty}\frac{t^lP_l(x)}{(l+1)^{n-1}}=\frac{1}{t}Q_{n-1}(t,x), \tag{B9}$$

therefore

$$Q_n(t,x)=\int_0^t\frac{1}{\tau}Q_{n-1}(\tau,x)d\tau. \tag{B10}$$

In particular,

$$Q_1(t,x)=\int_0^t\frac{d\tau}{\sqrt{1-2x\tau+\tau^2}}=\ln\frac{t-x+\sqrt{1-2xt+t^2}}{1-x}. \tag{B11}$$

Consecutive use of Eq. (B10) allows one to express $Q_n(t,x)$ via elementary functions and $\mathrm{Li}_k(z)$ of the order up to $k=n$.

To summarize, the following equations are obtained:

$$\sum_{l=0}^{\infty}\frac{t^l}{1+\frac{l}{l+1}\beta}=\frac{1}{(1+\beta)t}\sum_{n=0}^{\infty}\left(\frac{\beta}{1+\beta}\right)^n\mathrm{Li}_n(t) \tag{B12}$$

$$=\frac{1}{(1+\beta)t}\left[\frac{t}{1-t}-\frac{\beta}{1+\beta}\ln(1-t)+\sum_{n=2}^{\infty}\left(\frac{\beta}{1+\beta}\right)^n\mathrm{Li}_n(t)\right], \tag{B13}$$

$$\sum_{l=0}^{\infty}\frac{t^lP_l(x)}{1+\frac{l}{l+1}\beta}=\frac{1}{(1+\beta)t}\sum_{n=0}^{\infty}\left(\frac{\beta}{1+\beta}\right)^nQ_n(t,x) \tag{B14}$$

$$=\frac{1}{(1+\beta)t}\left[\frac{t}{\sqrt{1-2xt+t^2}}+\frac{\beta}{1+\beta}\ln\frac{t-x+\sqrt{1-2xt+t^2}}{1-x}+\sum_{n=2}^{\infty}\left(\frac{\beta}{1+\beta}\right)^nQ_n(t,x)\right]. \tag{B15}$$

Equations (B12) and (B14) are valid and converge for any β. They are especially computationally efficient for small β when the last term proportional to β^2 is negligible, and the first two terms alone may provide the desirable accuracy.

2. Case $\epsilon_{\mathrm{in}}>\epsilon_{\mathrm{out}}$

Consider now the case $\beta>1$. One can obtain the following series for $l>0$:

$$\frac{1}{1+\frac{l}{l+1}\beta}=\frac{l+1}{l(1+\beta)+1}$$

$$=1-\frac{l\beta}{l(1+\beta)+1}$$

$$=1-\frac{\beta}{1+\beta}\frac{l}{l+\frac{1}{1+\beta}}$$

$$=1-\frac{\beta}{1+\beta}\frac{1}{1+\frac{1}{l(1+\beta)}}$$

$$=1-\frac{\beta}{1+\beta}\sum_{n=0}^{\infty}\frac{(-1)^n}{l^n(1+\beta)^n}. \tag{B16}$$

Substitution of Eq. (B16) into Eq. (B2) yields

$$\sum_{l=0}^{\infty}\frac{t^l}{1+\frac{l}{l+1}\beta}=1+\sum_{l=1}^{\infty}t^l\left[1-\frac{\beta}{1+\beta}\sum_{n=0}^{\infty}\frac{(-1)^n}{l^n(1+\beta)^n}\right]$$

$$=1+\frac{t}{1-t}-\frac{\beta}{1+\beta}\sum_{n=0}^{\infty}\frac{(-1)^n}{(1+\beta)^n}\sum_{l=0}^{\infty}\frac{t^l}{l^n}$$

$$=\frac{1}{1-t}-\frac{\beta}{1+\beta}\sum_{n=0}^{\infty}\frac{(-1)^n}{(1+\beta)^n}\mathrm{Li}_n(t). \quad \text{(B17)}$$

By analogy, we obtain for the pair interaction term

$$\sum_{l=0}^{\infty}\frac{t^lP_l(x)}{1+\frac{l}{l+1}\beta}=1+\sum_{l=1}^{\infty}t^lP_l(x)\left[1-\frac{\beta}{1+\beta}\sum_{n=0}^{\infty}\frac{(-1)^n}{l^n(1+\beta)^n}\right]$$

$$=1+\left(\frac{1}{\sqrt{1-2xt+t^2}}-1\right)$$

$$-\frac{\beta}{1+\beta}\sum_{n=0}^{\infty}\frac{(-1)^n}{(1+\beta)^n}\sum_{l=1}^{\infty}\frac{t^lP_l(x)}{l^n}$$

$$=\frac{1}{\sqrt{1-2xt+t^2}}$$

$$-\frac{\beta}{1+\beta}\sum_{n=0}^{\infty}\frac{(-1)^n}{(1+\beta)^n}Q_n^{\dagger}(t,x), \quad \text{(B18)}$$

where another auxiliary function was introduced,

$$Q_n^{\dagger}(t,x)=\sum_{l=1}^{\infty}\frac{t^lP_l(x)}{l^n}. \quad \text{(B19)}$$

Note that for $0\leqslant t\leqslant 1$ and $-1\leqslant x\leqslant 1$ the Legendre polynomial is also bounded, $-1\leqslant P_l(x)\leqslant 1$, and the following inequalities hold:

$$|Q_n^{\dagger}(t,x)|\leqslant\sum_{l=1}^{\infty}\frac{t^l}{l^n}=\mathrm{Li}_n(t)\leqslant\mathrm{Li}_n(1)=\zeta(n). \quad \text{(B20)}$$

The same estimate holds for $Q_n(t,x)$.

Comparing definition (B19) with Eq. (B7), it can be easily shown by analogy that

$$Q_0^{\dagger}(t,x)=-1+\frac{1}{\sqrt{1-2xt+t^2}}, \quad \text{(B21)}$$

$$Q_1^{\dagger}(t,x)=-\ln\left[\frac{1}{2}(1-xt+\sqrt{1-2xt+t^2})\right], \quad \text{(B22)}$$

$$Q_n^{\dagger}(t,x)=\int_0^t\frac{1}{\tau}Q_{n-1}^{\dagger}(\tau,x)d\tau. \quad \text{(B23)}$$

Using Eq. (B23) recursively, $Q_n^{\dagger}(t,x)$ may be expressed via elementary functions and $\mathrm{Li}_k(z)$ of the order up to $k=n$.

Finally, the following equations are obtained:

$$\sum_{l=0}^{\infty}\frac{t^l}{1+\frac{l}{l+1}\beta}=\frac{1}{1-t}-\frac{\beta}{1+\beta}\sum_{n=0}^{\infty}\frac{(-1)^n}{(1+\beta)^n}\mathrm{Li}_n(t) \quad \text{(B24)}$$

$$=\frac{\beta}{1+\beta}+\frac{1}{1+\beta}\frac{1}{1-t}-\frac{\beta}{(1+\beta)^2}\ln(1-t)$$

$$-\frac{\beta}{1+\beta}\sum_{n=2}^{\infty}\frac{(-1)^n}{(1+\beta)^n}\mathrm{Li}_n(t), \quad \text{(B25)}$$

$$\sum_{l=0}^{\infty}\frac{t^lP_l(x)}{1+\frac{l}{l+1}\beta}=\frac{1}{\sqrt{1-2xt+t^2}}-\frac{\beta}{1+\beta}\sum_{n=0}^{\infty}\frac{(-1)^n}{(1+\beta)^n}Q_n^{\dagger}(t,x) \quad \text{(B26)}$$

$$=\frac{\beta}{1+\beta}+\frac{1}{1+\beta}\frac{1}{\sqrt{1-2xt+t^2}}$$

$$-\frac{\beta}{(1+\beta)^2}\ln\frac{1-xt+\sqrt{1-2xt+t^2}}{2}$$

$$-\frac{\beta}{1+\beta}\sum_{n=2}^{\infty}\frac{(-1)^n}{(1+\beta)^n}Q_n^{\dagger}(t,x). \quad \text{(B27)}$$

It is noteworthy that for a pair of charges such that $\theta=0$, $P_l(\cos\theta)\equiv 1$, and the Kirkwood equations (8) and (9) coincide. It is easy to see that the optimized series also coincide in the limit $\theta\to 0$: Eq. (B14) reduces to Eq. (B12), and Eq. (B26) reduces to Eq. (B24), as expected.

Note that the terms that contain $1/(1-t)$ and $\ln(1-t)$ in Eqs. (B13) and (B25) have a singularity at $t\to 1$. Terms that contain $\sqrt{1-2t\cos\theta+t^2}$ in Eqs. (B15) and (B27) only have a singularity when $t\to 1$ and $\cos\theta\to 1$ simultaneously, which can never happen because atoms do not overlap. Functions $\mathrm{Li}_n(t)$, $Q_n(t,x)$, and $Q_n^{\dagger}(t,x)$ are all bounded from above by the Riemann zeta function $\zeta(n)$ [see Eq. (B20)]. Since $\zeta(2)=\pi^2/6$ and $\zeta(n)$ decreases with n, the terms proportional to $\mathrm{Li}_n(t)$, $Q_n(t,x)$, and $Q_n^{\dagger}(t,x)$ do not have any singularities for $n\geqslant 2$.

[1] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, J. Am. Chem. Soc. **112**, 6127 (1990).
[2] M. Schaefer and M. Karplus, J. Phys. Chem. **100**, 1578 (1996).
[3] S. Edinger, C. Cortis, P. Shenkin, and R. Friesner, J. Phys. Chem. B **101**, 1190 (1997).
[4] B. Jayaram, Y. Liu, and D. J. Beveridge, J. Chem. Phys. **109**, 1465 (1998).
[5] A. Aghosh, C. S. Rapp, and R. A. Friesner, J. Phys. Chem. **102**, 10983 (1998).
[6] C. J. Cramer and D. G. Truhlar, Chem. Rev. (Washington, D.C.) **99**, 2161 (1999).
[7] D. Bashford and D. Case, Annu. Rev. Phys. Chem. **51**, 129 (2000).
[8] A. Onufriev, D. Bashford, and D. Case, J. Phys. Chem. B **104**, 3712 (2000).
[9] M. S. Lee, J. F. R. Salsbury, and C. L. Brooks III, J. Chem. Phys. **116**, 10606 (2002).
[10] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, Chem. Phys. Lett. **246**, 122 (1995).
[11] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, J. Phys. Chem. **100**, 19824 (1996).
[12] D. Qiu, P. Shenkin, F. Hollinger, and W. C. Still, J. Phys. Chem. A **101**, 3005 (1997).

[13]A. K. Felts, Y. Harano, E. Gallicchio, and R. M. Levy, Proteins **56**, 310 (2004).
[14]B. N. Dominy and C. L. Brooks, J. Phys. Chem. B **103**, 3765 (1999).
[15]L. David, R. Luo, and M. K. Gilson, J. Comput. Chem. **21**, 295 (2000).
[16]V. Z. Spassov, L. Yan, and S. Szalma, J. Phys. Chem. B **106**, 8726 (2002).
[17]N. Calimet, M. Schaefer, and T. Simonson, Proteins **45**, 144 (2001).
[18]V. Tsui and D. Case, J. Am. Chem. Soc. **122**, 2489 (2000).
[19]T. Wang and R. Wade, Proteins **50**, 158 (2003).
[20]A. Onufriev, D. Bashford, and D. A. Case, Proteins **55**, 383 (2004).
[21]E. Gallicchio and R. M. Levy, J. Comput. Chem. **25**, 479 (2004).
[22]C. Simmerling, B. Strockbine, and A. E. Roitberg, J. Am. Chem. Soc. **124**, 11258 (2002).
[23]H. Nymeyer and A. E. Garcia, Proc. Natl. Acad. Sci. U.S.A. **100**, 13934 (2003).
[24]M. C. Lee and Y. Duan, Proteins **55**, 620 (2004).
[25]B. Honig and A. Nicholls, Science **268**, 1144 (1995).
[26]P. Beroza and D. A. Case, Methods Enzymol. **295**, 170 (1998).
[27]J. D. Madura, M. E. Davis, M. K. Gilson *et al.*, Rev. Comput. Chem. **5**, 229 (1994).
[28]M. K. Gilson, Curr. Opin. Struct. Biol. **5**, 216 (1995).
[29]M. Scarsi, J. Apostolakis, and A. Caflisch, J. Phys. Chem. A **101**, 8098 (1997).
[30]R. Luo, L. David, and M. K. Gilson, J. Comput. Chem. **23**, 1244 (2002).
[31]A. Onufriev, D. Case, and D. Bashford, J. Comput. Chem. **23**, 1297 (2002).
[32]D. Bashford and M. Karplus, Biochemistry **29**, 10219 (1990).
[33]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975).
[34]J. G. Kirkwood, J. Chem. Phys. **2**, 351 (1934).
[35]T. Grycuk, J. Chem. Phys. **119**, 4817 (2003).
[36]D. Bashford, in *Scientific Computing in Object-Oriented Parallel Environments*, Lecture Notes in Computer Science, Vol. 1343 edited by Y. Ishikawa, R. R. Oldehoeft, J. V. W. Reynders, and M. Tholburn (Springer, Berlin, 1997), pp. 233–240.
[37]A. Nicholls and B. Honig, J. Comput. Chem. **12**, 435 (1991).
[38]A. Bondi, J. Phys. Chem. **68**, 441 (1964).
[39]M. L. Connolly, Science **221**, 709 (1983).
[40]W. Im, M. Feig, and C. L. Brooks III, Biophys. J. **85**, 2900 (2003).