

Analytical electrostatics for biomolecules: Beyond the generalized Born approximation

Grigori Sigalov, Andrew Fenley, and Alexey Onufriev^{a)}

Departments of Computer Science and Physics, Virginia Tech,^{b)} Blacksburg, Virginia 24061

(Received 27 December 2005; accepted 25 January 2006; published online 23 March 2006)

The modeling and simulation of macromolecules in solution often benefits from fast analytical approximations for the electrostatic interactions. In our previous work [G. Sigalov *et al.*, *J. Chem. Phys.* **122**, 094511 (2005)], we proposed a method based on an approximate analytical solution of the linearized Poisson-Boltzmann equation for a sphere. In the current work, we extend the method to biomolecules of arbitrary shape and provide computationally efficient algorithms for estimation of the parameters of the model. This approach, which we tentatively call ALPB here, is tested against the standard numerical Poisson-Boltzmann (NPB) treatment on a set of 579 representative proteins, nucleic acids, and small peptides. The tests are performed across a wide range of solvent/solute dielectrics and at biologically relevant salt concentrations. Over the range of the solvent and solute parameters tested, the systematic deviation (from the NPB reference) of solvation energies computed by ALPB is 0.5–3.5 kcal/mol, which is 5–50 times smaller than that of the conventional generalized Born approximation widely used in this context. At the same time, ALPB is equally computationally efficient. The new model is incorporated into the AMBER molecular modeling package and tested on small proteins. © 2006 American Institute of Physics. [DOI: 10.1063/1.2177251]

I. INTRODUCTION

In numerous problems of computational chemistry and structural biology, including but not limited to the modeling of structure, dynamics, and interactions of proteins and nucleic acids, protein folding, drug/ligand docking, and so on, the free energy of biomolecules in solution must be calculated. Since the electrostatic forces are the strongest long-range interactions on the atomic scale, these interactions are a major part of the solvation energy. Numerical Poisson-Boltzmann^{1–11} (NPB) methods are routinely used to calculate the electrostatic part of the solvation free energy and other electrostatic properties. They provide the most accurate estimates when the computation expense is not an issue. However, they often have to be replaced by analytical methods such as the generalized Born (GB) approximation^{1,12–23} when large or fast-changing systems are involved. Phenomenological by its nature but much more effective computationally than NPB, the GB approximation proves useful and has become widespread, especially in molecular dynamics (MD) applications^{24–36} where the high speed of calculations is a prerequisite.

In the conventional GB theory, a molecule is considered as a continuous region with dielectric constant ϵ_{in} surrounded by infinite solvent with dielectric ϵ_{out} . Charges q_i are located at positions \mathbf{r}_i inside the molecule. Their interaction in the presence of a polarized solvent contributes to the electrostatic part of the solvation energy ΔG_{el} , which is commonly calculated as follows:¹²

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

where R_i are the effective Born radii of the atoms. An established¹² form of f_{ij} is

$$f_{ij} = \sqrt{r_{ij}^2 + R_i R_j} \exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right). \quad (2)$$

While various forms of Eq. (2) were proposed over the years, the foundation of the GB model—Eq. (1)—remained unchanged; in what follows we call the models based on Eq. (1) “conventional GB” model. We have recently shown³⁷ that Eq. (1) misses some important physics. One of the manifestations of this shortcoming of the GB model is that it fails to correctly reproduce the dependence of ΔG_{el} on the solute and solvent dielectrics.³⁷ This deficiency of the approximation stems from the incorrect functional dependence of ΔG_{el} on ϵ_{in} and ϵ_{out} . It is easy to see from Eq. (1) that, in GB theory, swapping ϵ_{in} and ϵ_{out} only changes the sign of the solvation energy: $\Delta G_{\text{el}}(\epsilon_{\text{in}}, \epsilon_{\text{out}}) + \Delta G_{\text{el}}(\epsilon_{\text{out}}, \epsilon_{\text{in}}) = 0$. Meanwhile, such symmetry does not actually exist in nature, as can be shown by theoretical consideration or NPB calculations. Even at $\epsilon_{\text{in}}/\epsilon_{\text{out}} = 1/80$ (aqueous solvation), the GB model can be noticeably improved, as we will show below.

Practical considerations of computational ease and speed play a major role in theory development. In particular, it is often imperative for an algorithm that computes the energy within an MD application to be expressed by an analytical equation simple enough to be implemented effectively. To calculate the forces acting on atoms, one typically needs

^{a)}Electronic mail: alexey@cs.edu.vt.edu

^{b)}URL: <http://people.cs.vt.edu/onufriev/>

computationally facile forms of the derivatives of the pairwise atomic interaction terms with respect to atomic coordinates.

Even for a shape as trivial as a sphere, the exact closed-form solution of the linearized Poisson-Boltzmann (LPB) equation turns out to be cumbersome and, in its original form due to Kirkwood,³⁸ unsuitable for numerical implementation because of the slowly converging infinite series it contains.³⁷

However, as our recent study shows,³⁷ the infinite series solution can be regularized and summed up approximately, leading to the following equation:

$$\Delta G_{\text{el}} \approx -\frac{1}{2} \left(\frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \frac{1}{1 + \alpha\beta} \sum_{ij} q_i q_j \left(\frac{1}{f_{ij}} + \frac{\alpha\beta}{A} \right), \quad (3)$$

where f_{ij} is same as in Eq. (2), $\beta = \epsilon_{\text{in}} / \epsilon_{\text{out}}$, $\alpha \approx 0.571412$, and A is the *electrostatic size* of the molecule. The latter provides a relationship between the molecule's global shape and its electrostatic energy.³⁷

In spite of its functional simplicity and resemblance to the conventional GB theory, Eq. (3) it is not a GB theory, even in the most general sense: the solvation free energy depends, via f_{ij} , not only on local parameters of each pair of atoms such as their effective Born radii, but also, via A , on the over-all shape and size of the structure. Also, the prefactor in Eq. (3) is different from that in the GB Eq. (1). To distinguish the new approximation from the GB model, we tentatively call it the analytical linearized Poisson-Boltzmann (ALPB) approach. The ALPB is free from the deficiencies of Eq. (1) mentioned above, and, in particular, Eq. (3) has correct physical asymptotics in both limits $\beta \rightarrow 0$ and $\beta \rightarrow \infty$.

The goals of the present paper are as follows: (1) Provide a fast, efficient way to estimate the key new parameter of the ALPB model, A . We provide an analytical, unambiguous, parameter-free, and numerically effective method to calculate the electrostatic size A of an arbitrarily shaped molecule. This step will complete the construction of an approximate analytical solution to the linearized Poisson-Boltzmann equation for molecular applications. (2) Extensively test Eq. (3) on a large, representative set of realistic biomolecules. (3) Verify that the ALPB can be used in realistic molecular dynamics applications and produce stable trajectories.

The paper is organized as follows. We present an analytical algorithm to calculate the electrostatic size of a molecule in Sec. II. Methodological details and the description of the test sets are presented in Sec. III. In Sec. IV we establish the accuracy of the new approach by comparing ALPB with the NPB on a large set of representative molecular structures. The accuracy comparison of the ALPB with the GB is also included. We then introduce the salt dependence into the model and establish the algorithmic stability and efficiency of the ALPB in molecular dynamics simulation of a folded protein; possible applications to other scenarios such as the folding/refolding of proteins is discussed. Conclusions are presented in Sec. V.

II. THEORY

A. Definition of the electrostatic size of a molecule

The definition of the effective electrostatic size of the molecule follows directly from Eq. (3):

$$A = \lim_{\epsilon_{\text{in}} \rightarrow \infty} \frac{q^2}{2\Delta G_{\text{el}}} \Big|_{\epsilon_{\text{out}}=1}. \quad (4)$$

So, if ΔG_{el} in this limit is known (e.g., from the numerical solution of the PB equation), A is easily found from Eq. (4). The following considerations clarify the definition of the electrostatic size and suggest ways of computing it in practice. Imagine that the molecule is filled with an infinite dielectric, $\epsilon_{\text{in}} \rightarrow \infty$, while $\epsilon_{\text{out}} = 1$. Due to infinite polarizability, all buried charges are exactly compensated by polarization charges, and so only the induced surface charges remain uncompensated—their sum is equal to the net charge of the molecule, and their distribution is completely independent of the distribution of the original fixed charges q_i inside. This is because the electric field $\mathbf{E} = 0$ in the interior, and the system is therefore equivalent to a conductor. A conducting sphere of size (radius) A with a single charge q stores electrostatic energy $\Delta G_{\text{el}} = q^2/2A$ irrespective of the charge location. For a charged ($q \neq 0$) molecule of arbitrary shape, the same functional dependence $\Delta G_{\text{el}} \sim q^2$ holds. By analogy with a sphere, we can write it down as $\Delta G_{\text{el}} = q^2/2A$, only now A is the effective electrostatic size of the molecule. This formula is mathematically similar to the definition of the effective Born radius,

$$\Delta G_{ii}^{\text{el}} = -\frac{1}{2} \left(\frac{1}{\epsilon_{\text{in}}} - \frac{1}{\epsilon_{\text{out}}} \right) \frac{q_i^2}{R_i}. \quad (5)$$

However, since unlike R_i , A is independent of the positions of charges in the molecule, one can write

$$\lim_{\epsilon_{\text{in}} \rightarrow \infty} \frac{q_i^2}{2\Delta G_{ii}^{\text{el}}} \Big|_{\epsilon_{\text{out}}=1} = \lim_{\epsilon_{\text{in}} \rightarrow \infty} R_i \Big|_{\epsilon_{\text{out}}=1} = A \quad (6)$$

and so, in principle, one can put a single probe charge anywhere inside the structure and obtain A by computing $\Delta G_{ii}^{\text{el}}$ or R_i in the appropriate limits. We find, however, that due to numerical artifacts and finite-grid representation, R_i s obtained via common NPB solvers form a distribution of finite width even at $\epsilon_{\text{in}} \sim 10^3 - 10^6$. Therefore, a large enough number of R_i ($\beta \rightarrow \infty$) needs to be calculated to provide a statistically reasonable estimate of A . This procedure that relies on the NPB is similar to the one used to obtain the “perfect” effective Born radii and is arguably the best approach for a

theoretical analysis.³⁹ However, since even a single NPB run may take appreciable time, such procedure is extremely time consuming and therefore is not suitable for use in MD applications, where A might have to be recalculated often.

On the other hand, since A appears to be similar to the effective Born radii in the GB model, one can, in principle, compute it by integrating the electric field (displacement) density around the molecule. However, unlike in the GB theory, we find that the approaches such as the Coulomb field approximation (CFA) widely used to compute the effective Born radii are not applicable here due to the nature of the electric field in the $\epsilon_{\text{in}} \rightarrow \infty$ regime. The approach that we have developed is very different and is outlined below.

Note that the electrostatic solvation energy of a molecule in the limit $\epsilon_{\text{in}} \rightarrow \infty$ does not depend on the particular spacial arrangement of the charges inside but only on the total charge q and the shape and dimensions of the body through its electrostatic size A . This means that A is a characteristic on the geometry of the molecule. Therefore, it might be possible to find A using relatively simple geometric considerations. In an attempt to provide such considerations, and keeping in mind that our goal is to develop a fast analytical method that could be used in MD applications, we will approximate the molecule by a simple shape for which $\Delta G_{\text{el}}(\epsilon_{\text{in}} \rightarrow \infty, \epsilon_{\text{out}} = 1)$ is found exactly, and then use it to find A via Eq. (4). While the simplest shape—a sphere—is a possibility, we would like to consider the next step in complexity—an ellipsoid, which encompasses spherical molecules as a particular case, but goes beyond that to account for prolate and oblate shapes as well. Our strategy is as follows. For a given molecular structure, we first find the corresponding best fit ellipsoid, defined by its semiaxes. Next, we find the exact solvation energy of this ellipsoid, in the limits of Eq. (4); since this quantity is a unique function of the ellipsoid's geometry (semiaxes), we automatically obtain the molecule's approximate electrostatic size via Eq. (4).

B. Approximation of a molecular shape by an effective ellipsoid

Let us determine the semiaxes of an effective ellipsoid that best approximates the shape of an arbitrary molecule.

Consider the molecule as a complex body made of spherical atoms with their centers at \mathbf{r}_i and radii a_i . Actual atomic weights are ignored because A should not depend on the mass distribution inside the molecule. Instead, we consider atoms as hard balls of a constant specific weight. An atom's "mass" is defined as $m_i = a_i^3$; the mass of the molecule is then $M = \sum_i m_i$.

We start with finding the center of mass of the molecule, $\mathbf{r}_0 = M^{-1} \sum_i m_i \mathbf{r}_i$. Then we move the origin to the center of mass, $\mathbf{r}'_i = \mathbf{r}_i - \mathbf{r}_0$. Below we drop the prime in $\mathbf{r}'_i = (x'_i, y'_i, z'_i)$ for the sake of brevity. Now, the components of the molecule's inertia tensor I are found as follows:

$$\begin{aligned} I_{11} &= \sum_i m_i \left(y_i^2 + z_i^2 + \frac{2}{5} a_i^2 \right), \\ I_{22} &= \sum_i m_i \left(x_i^2 + z_i^2 + \frac{2}{5} a_i^2 \right), \\ I_{33} &= \sum_i m_i \left(x_i^2 + y_i^2 + \frac{2}{5} a_i^2 \right), \\ I_{12} &= I_{21} = - \sum_i m_i x_i y_i, \\ I_{13} &= I_{31} = - \sum_i m_i x_i z_i, \\ I_{23} &= I_{32} = - \sum_i m_i y_i z_i. \end{aligned} \quad (7)$$

The principal moments of inertia are equal to the eigenvalues λ_i of the tensor I and are found from the following equation:

$$\text{Det} \begin{pmatrix} I_{11} - \lambda & I_{12} & I_{13} \\ I_{21} & I_{22} - \lambda & I_{23} \\ I_{31} & I_{32} & I_{33} - \lambda \end{pmatrix} = k_3 \lambda^3 + k_2 \lambda^2 + k_1 \lambda + k_0 = 0, \quad (8)$$

where the coefficients k_i are written down as follows, due to the tensor's symmetry:

$$k_0 = \text{Det } I = I_{11} I_{22} I_{33} + 2 I_{12} I_{23} I_{13} - I_{11} I_{23}^2 - I_{22} I_{13}^2 - I_{33} I_{12}^2, \quad (9)$$

$$k_1 = -I_{11} I_{22} - I_{22} I_{33} - I_{11} I_{33} + I_{12}^2 + I_{23}^2 + I_{13}^2, \quad (10)$$

$$k_2 = \text{Tr } I = I_{11} + I_{22} + I_{33}, \quad (11)$$

$$k_3 = -1. \quad (12)$$

Because of its physical nature, the cubic equation [(8)–(12)] is guaranteed to have three real roots, therefore the trigonometric method of solution⁴⁰ is most appropriate. We reduce Eq. (8) to the incomplete cubic equation $\lambda^3 + p\lambda + q = 0$, where

$$p = -\frac{k_2^2}{3} - k_1, \quad q = -\frac{2k_2^3}{27} - \frac{k_1 k_2}{3} - k_0. \quad (13)$$

Two auxiliary parameters are then introduced,

$$s = -2 \sqrt{-\frac{p}{3}}, \quad \alpha = \arccos \left(-\frac{4q}{s^3} \right). \quad (14)$$

Finally, the eigenvalues of the tensor of inertia I are found as follows:

$$\lambda_1 = s \cos \frac{\alpha}{3} + \frac{k_2}{3}, \quad \lambda_{2,3} = -s \cos \left(\frac{\alpha \pm \pi}{3} \right) + \frac{k_2}{3}. \quad (15)$$

The principal moments of inertia of the molecule I_{xx} , I_{yy} , and I_{zz} are equal to the eigenvalues λ_i chosen so that $I_{xx} \leq I_{yy} \leq I_{zz}$.

Now let us find the semiaxes a , b , and c of a solid ellipsoid that has the same weight M and principal moments of inertia $I_{\alpha\alpha}$ as the molecule under consideration. The principal moments of inertia are expressed through the ellipsoid's semiaxes as follows:

$$\begin{aligned} I_{xx} &= \frac{1}{5}M(b^2 + c^2), & I_{yy} &= \frac{1}{5}M(a^2 + c^2), \\ I_{zz} &= \frac{1}{5}M(a^2 + b^2), \end{aligned} \quad (16)$$

therefore

$$\begin{aligned} a &= \sqrt{\frac{5}{2M}(-I_{xx} + I_{yy} + I_{zz})}, \\ b &= \sqrt{\frac{5}{2M}(I_{xx} - I_{yy} + I_{zz})}, \\ c &= \sqrt{\frac{5}{2M}(I_{xx} + I_{yy} - I_{zz})}. \end{aligned} \quad (17)$$

The above ellipsoid, with semiaxes a , b , and c , provides the simplest nontrivial approximation—beyond a spherical one—to the shape of a biomolecule under consideration.

C. Electrostatic size of an ellipsoid

A charged conducting ($\epsilon_{\text{in}} \rightarrow \infty$) ellipsoid with semiaxes a , b , and c is known⁴¹ to store electrostatic energy,

$$\Delta G_{\text{el}}^{\text{ell}} = \frac{q^2}{4} \int_0^\infty \frac{d\theta}{\sqrt{(a^2 + \theta)(b^2 + \theta)(c^2 + \theta)}}, \quad (18)$$

in vacuum ($\epsilon_{\text{out}} = 1$). Compare Eq. (18) to the definition of the effective electrostatic size, Eq. (4), to obtain for the electrostatic size of an ellipsoid,

$$A_{\text{ell}} = \frac{q^2}{2\Delta G_{\text{el}}^{\text{ell}}} = 2 \left[\int_0^\infty \frac{d\theta}{\sqrt{(a^2 + \theta)(b^2 + \theta)(c^2 + \theta)}} \right]^{-1}. \quad (19)$$

In the particular case of a sphere, $a=b=c=A$, the integral in Eq. (19) is equal to $2/a$, therefore $A_{\text{ell}}=a$, as expected.

Equation (19) reduces to an elliptic integral of the first kind, which can only be calculated numerically. To use an integration scheme such as Simpson's method, the upper infinite limit must be eliminated as shown in Appendix A.

To calculate A in numerical applications where computation speed is critical and numerical integration should be avoided, the integral in Eq. (19) is approximately expressed through elementary functions. Note that in the case of a prolate ellipsoid with rotational symmetry, $b=c$, Eq. (19) gives

$$\tilde{A}_{\text{ell}} = 2a\gamma \left(\log \frac{1+\gamma}{1-\gamma} \right)^{-1}, \quad (20)$$

where $\gamma = \sqrt{1 - b^2/a^2}$. In the general case of $b \neq c$, we only need to redefine γ , e.g., as

$$\gamma = \sqrt{1 - \frac{(b+c)^2}{4a^2}}. \quad (21)$$

Equations (20) and (21) provide a simple approximation that exactly coincides with Eq. (19) when $b=c$ and is accurate enough in most other cases, as shown below in Sec. IV A. A similar estimate is possible for the case $a \approx b > c$, which is rare among realistic biomolecular shapes. Given the complexity of Eq. (19), it is unlikely that a useful expression is derivable for shapes with even less symmetry than ellipsoidal.

D. Analytical differentiable estimate of the electrostatic size of a molecule

To use Eq. (3) in MD applications, one needs to find analytical derivatives of A with respect to atomic coordinates, $\nabla_i A = (\partial A / \partial x_i, \partial A / \partial y_i, \partial A / \partial z_i)$. In principle, this can be done for A given by Eq. (19) with semiaxes from Eq. (17). Note, however, that A is expressed through the effective ellipsoid's semiaxes a , b , and c via an elliptic integral, and the semiaxes are, in turn, functions of the roots of a cubic equation with cumbersome coefficients. Therefore, the exact derivatives $\nabla_i A$ would contain elliptic integrals and generally would have a complicated form that may not be suitable for fast calculations. An algorithm based on Eq. (19) may also be prone to numerical instabilities due to the possible degeneration and branching of the solutions of the cubic equation if the molecule's symmetry changes during conformational dynamics.

Therefore, one wonders if the final formula for A could be simplified by providing an approximate expression in lieu of the exact one based upon an elliptic integral (19). Since the calculation of A is already approximate by its nature and A itself is meant for use in another approximate method of solvation energy calculation, it is possible that simplifying Eq. (19) would not add much to the error already made or even compensate part of that error.

We use the inertia tensor's invariants to find an approximate electrostatic size. Comparison shows that the best results are provided by the tensor's determinant k_0 given by Eq. (9). We rewrite Eq. (9) using the principal moments of inertia and Eq. (16) to obtain

$$\text{Det } I = I_{xx}I_{yy}I_{zz} = \frac{M^3}{125}(a^2 + b^2)(b^2 + c^2)(c^2 + a^2). \quad (22)$$

For a spherical molecule of size $a=b=c=A$, Eq. (22) yields $\text{Det } I = 8a^6M^3/125$. Therefore, by dimensional analysis,

$$\begin{aligned} A_{\text{Det}} &= \sqrt{\frac{5}{2M}} \sqrt[6]{\text{Det } I} = \sqrt{\frac{5}{2M}} [I_{11}I_{22}I_{33} + 2I_{12}I_{23}I_{13} \\ &\quad - I_{11}I_{23}^2 - I_{22}I_{13}^2 - I_{33}I_{12}^2]^{1/6}. \end{aligned} \quad (23)$$

The derivatives of A_{Det} are found with the help of Eq. (7) as shown in Appendix B.

Thus, both A_{Det} and $\nabla_i A_{\text{Det}}$ are found directly from the coordinates of the atoms, via I_{jk} , in a computationally efficient manner. The number of machine operations required is $O(N)$, where N is the number of atoms. Through the use of

an invariant of the inertia tensor, the electrostatic size is approximated without the danger of numerical instabilities associated with the solving of the cubic equation, Eq. (8).

In what follows, we establish the accuracy of approximations (19), (20), and (23) against the NPB reference on a large set of biomolecular structures.

III. METHODS

A. Numerical Poisson-Boltzmann (NPB) solvers used

We have chosen the finite-difference NPB solver PEP (Ref. 42) to serve as a reference for all other electrostatic models discussed in this work. While PEP is extremely computationally expensive, it is arguably one of the more accurate among finite-difference NPB solvers of comparable degree of sophistication. The PEP package is particularly well suited for calculating individual pairwise interactions and effective Born radii. In all calculations, we use four levels of focusing with the finest grid size of 0.0625 Å. The solvent probe radius is 1.4 Å, the ion exclusion (Stern) radius is 2 Å. We have modified the original PEP package by changing the machine representation of some variables from single to double precision. These calculations are always referred to as “reference NPB” in this work.

We also use another established NPB solver, DELPHI-II,² to provide a way to access the inherent variation of the NPB reference due to algorithmic details such as approximations to the molecular surface. A cubic box of 251 grid points in each dimension and grid size 0.5 Å is used *except* in Sec. IV B where 0.25 Å grid spacing has been set. In this case, the consensus set had to be reduced to 395 structures: running DELPHI-II on the other, larger structures caused segmentation fault on our 2.4 GHz Pentium IV machine with 1 G of random access memory (RAM).

B. Structures: Consensus benchmark set

To test our approximate models against the NPB reference described above, we have constructed a benchmark set of representative biomolecular structures, that we have called *the consensus benchmark set*. The set is intended to minimize the numerical artifacts of the NPB procedures. We start with a set of ~600 representative structures used by Feig *et al.* in a similar context.⁴² We have expanded the set by several other structures, including *A* and *B* form DNA decamers, β hairpin, myoglobin, and lysozyme in standard protonation states. We then have calculated the electrostatic solvation energy ΔG_{el} for these biomolecules using two NPB solvers described above, PEP and DELPHI. The distribution of the relative difference $|\Delta G_{\text{el}}^{\text{DELPHI}} - \Delta G_{\text{el}}^{\text{PEP}}| / \Delta G_{\text{el}}^{\text{PEP}}$ is shown in Fig. 1 for $\epsilon_{\text{in}}=1$ and $\epsilon_{\text{out}}=1000$. The distributions computed with $\epsilon_{\text{out}}=8, 20, 40, 80,$ and 1000 are almost identical. The structures with the values of $|\Delta G_{\text{el}}^{\text{DELPHI}} - \Delta G_{\text{el}}^{\text{PEP}}| / \Delta G_{\text{el}}^{\text{PEP}} < 2.5\%$ form the consensus benchmark set. The above threshold leaves us with 579 structures with dimensions varying from 9 to 32 Å, or 247–8254 atoms per structure. The range of the absolute values of the total charge in the consensus set is from 0 to 28|*e*|. The consensus set along with the computed ΔG_{el} is available from the authors upon request.

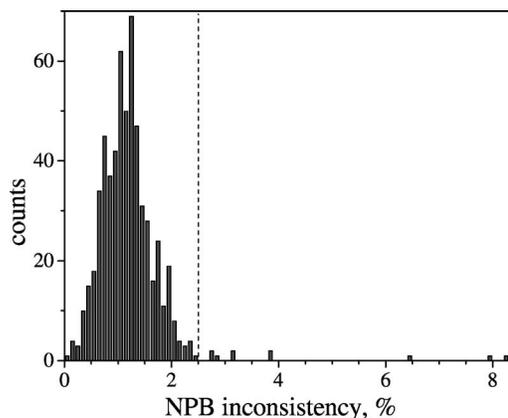


FIG. 1. Definition of the consensus set used in this work to test the approximate analytical models. The plot shows the distribution of the relative difference of the electrostatic solvation energy calculated by the NPB solver DELPHI relative to the reference NPB solver PEP, $|\Delta G_{\text{el}}^{\text{DELPHI}} - \Delta G_{\text{el}}^{\text{PEP}}| / \Delta G_{\text{el}}^{\text{PEP}}$. Energies are calculated for a set of 595 biomolecules by both solvers at $\epsilon_{\text{in}}=1$, $\epsilon_{\text{out}}=1000$. The main peak of the distribution is bounded by value of $\approx 2.5\%$ (shown by vertical dashed line) and encompasses 579 proteins and nucleic acids which form the consensus benchmark set.

C. Calculation of the reference electrostatic sizes by the NPB

The electrostatic size A has been calculated for each molecule of the consensus benchmark set using our reference NPB solver, see Sec. III B. These values are referred to as A_{NPB} .

Physical considerations discussed above dictate that all effective Born radii $R_i \rightarrow A$ as $\beta = \epsilon_{\text{in}} / \epsilon_{\text{out}} \rightarrow \infty$, irrespective of the location of charge i in the structure. Direct NPB calculation with various β corroborate this conclusion within acceptable error bounds. Indeed, we found that the width of the R_i distribution for a given molecule decreases as β increases in the range $\beta=100-10\,000$. However, even at $\beta=10\,000$ this width is not negligible, and its decrease in going from $\beta=1000-10\,000$ is very small to expect it to decrease substantially further with even higher values of β . To reduce these numerical artifacts of NPB in A_{NPB} , we have used the following averaging procedure. For each molecule, all effective Born radii R_i are calculated at $\epsilon_{\text{out}}=1$, $\epsilon_{\text{in}}=1000$. The average \bar{R} and standard deviation σ are found. Then we drop all R_i such that $|R_i - \bar{R}| > \sigma$ and calculate the average \bar{R}' and standard deviation σ' for the “trimmed” distribution. The trimmed average \bar{R}' is adopted as A_{NPB} . The distribution width σ' is normally about half of the original σ ; for most of the structures tested, $\sigma' \approx 0.2-0.4$ Å.

IV. RESULTS AND DISCUSSION

A. Accuracy of the analytical methods of calculation of the electrostatic size

We have calculated the electrostatic sizes A_{ell} [Eq. (19)], \tilde{A}_{ell} [Eqs. (20) and (21)], and A_{Det} [Eq. (23)] for each molecule from the benchmark set and compared them to the corresponding NPB reference values, Table I. Each of the three analytical methods considered here provide a reasonable approximation to A_{NPB} , with A_{Det} being the most accu-

TABLE I. Differences between the effective electrostatic sizes calculated by various approximate analytical methods and the NPB reference. Average difference over the consensus benchmark set and the corresponding standard deviation for each pair of methods are shown.

Method	Reference	Difference (Å)
A_{ell}	A_{NPB}	2.23 ± 2.56
\tilde{A}_{ell}	A_{NPB}	2.24 ± 2.58
A_{Det}	A_{NPB}	0.55 ± 1.71
\tilde{A}_{ell}	A_{ell}	0.011 ± 0.022

rate of them, on average. Note that the reference values A_{NPB} are themselves defined to within a standard deviation, normally 0.2–0.4 Å, due to the finite width of distribution of the Born radii computed by an NPB solver, see Sec. III. It is important to realize that A is ultimately used to calculate the electrostatic solvation energies and charge-charge interactions via Eq. (25). So it is the accuracy of these estimates and not the accuracy of the calculation of A itself that is key. A comparison of ΔG_{el} calculated by various methods is presented below. As we shall see, the ALPB equation, Eq. (3), is rather insensitive to reasonable variations in A , and so errors of an order of a few angstroms in the estimate of the electrostatic size are acceptable for most practical purposes.

As expected, the analytical methods for calculating A that are exact for ideal, highly symmetric shapes yield the largest errors for irregularly shaped molecules. The molecule that gives rise to the absolute largest error among all of the 579 structures in the consensus benchmark set is a protein shown in Fig. 2 (left). Clearly, this structure is anything but an ellipsoid; it is not even globular. So it is not surprising that the deviation of its electrostatic size from the NPB reference is about 50%. What is, perhaps, unexpected is that even for this structure the error of the ALPB solvation energy is only about 0.1% ($\Delta G_{\text{el}}^{\text{ALPB}} - \Delta G_{\text{el}}^{\text{NPB}} = 3.27$ kcal/mol, $\Delta G_{\text{el}}^{\text{NPB}} = -2643.22$ kcal/mol for $\epsilon_{\text{out}} = 80$). At the same time, the approximate methods are capable of providing estimates of the effective electrostatic sizes that are much closer to the corresponding NPB reference values for structures that are more or less globular but still far from being ellipsoidal. An example of such structures is a small DNA fragment, Fig. 2 (right) for which A is approximated within 5% of its NPB value. Not surprisingly, the error in $\Delta G_{\text{el}}^{\text{ALPB}}$ is smaller, 0.005% (–0.22 kcal/mol) relative to the NPB value. While the general trend is present, we find the correlation between the accuracy of the approximate effective electrostatic size estimates and the accuracy of ALPB itself is too weak to be of practical use, and we do not pursue its quantification. A thorough analysis of the ALPB accuracy will be presented below.

The analytical methods of estimating the electrostatic size presented above are computationally fast, especially the most straightforward A_{Det} estimate. For example, the code used in this work computes A for all 579 benchmark set molecules (total of ca. 10^6 atoms) in about 60 s on a standard 2 GHz personal computer. The time of each calculation is directly proportional to the number of atoms N in the structure and therefore is negligible compared to the typical cost

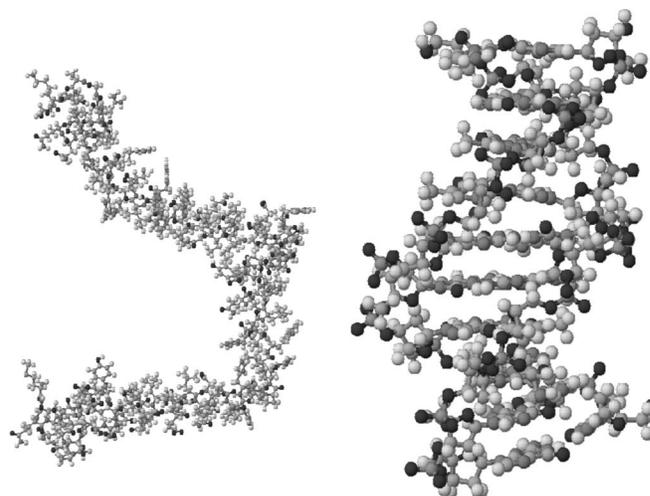


FIG. 2. (Left) The structure out of the consensus benchmark set for which the analytical methods of electrostatic size estimation are *least* accurate, a viral protein [Protein Data Bank (PDB) ID: 1esx]. For this irregular, clearly nonglobular structure $A_{\text{NPB}} = 20.49 \pm 0.36$ Å, $A_{\text{Det}} = 31.24$ Å. (Right) An example of a globular but nonellipsoidal molecule for which the analytical value A_{Det} is quite accurate, B-DNA (10 base-pair duplex); $A_{\text{NPB}} = 14.03 \pm 0.23$ Å, $A_{\text{Det}} = 14.68$ Å.

of many other applications, which scale as N^α , $\alpha > 1$. A C++ implementation of the algorithms described in this section is freely available from <http://people.cs.vt.edu/~onufriev/software.html>.

B. Accuracy of the solvation energy calculation

The goal of this section is to show that the ALPB approximation given by Eq. (3) provides a statistically significant improvement over the conventional GB approach [Eq. (1)] in calculating the electrostatic part of the solvation energy of realistic biomolecules. To evaluate the accuracy of the approximations the ALPB method is based upon, it is compared to the numerically exact Poisson-Boltzmann treatment; both models share the same underlying physics of the continuum solvation. To provide statistically significant estimates, we have generated reference NPB ΔG_{el} data for all 579 molecules from our consensus benchmark set (see Sec. III) using a range of values for the solvent dielectric, see Fig. 3 and Table II. The accuracy of various approximations to the effective electrostatic size discussed above is also tested in this context. For each method of ΔG_{el} estimation, relative error $\Delta \Delta G_{\text{el}} = (\Delta G_{\text{el}} - \Delta G_{\text{el}}^{\text{NPB}})$ is computed for each molecule, and then its average and standard deviation are calculated over the consensus benchmark set.

The main conclusion from Table II is that for all reasonable values of ϵ_{out} the accuracy of ALPB method remains equally high. Moreover, when looking at the range of error compared to the NPB treatment, $(\Delta \Delta G_{\text{el}} - \sigma_{\Delta \Delta G})$ to $(\Delta \Delta G_{\text{el}} + \sigma_{\Delta \Delta G})$, where $\sigma_{\Delta \Delta G}$ is the standard deviation of $\Delta \Delta G_{\text{el}}$ value, the ALPB range always contains the zero error point, while this is mostly not the case for the conventional GB method. The mean error of the GB method at $\epsilon_{\text{out}} = 80$ is almost five times larger than that of ALPB and increases with decreasing ϵ_{out} . The absence of a systematic ALPB error and its presence in the GB model is particularly well illustrated

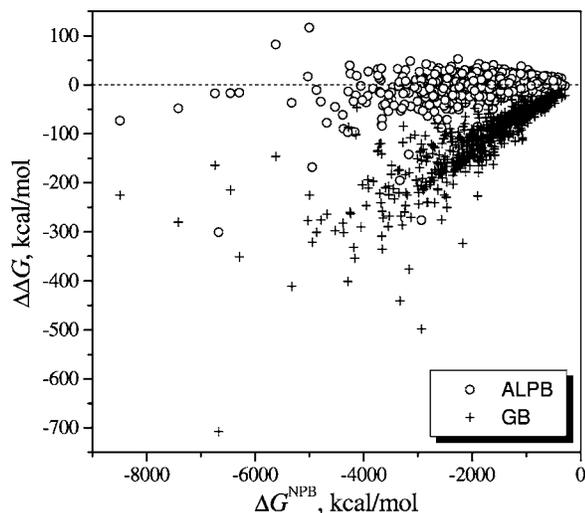


FIG. 3. Relative errors $\Delta\Delta G_{el} = (\Delta G_{el} - \Delta G_{el}^{NPB})$ in electrostatic solvation energies of ALPB [using A_{Det} , Eq. (23)], and conventional GB methods for the consensus set of 579 biomolecules computed at $\epsilon_{in}=1$, $\epsilon_{out}=8$. Each data point represents the difference in energies for a particular molecule. The ALPB data points (circles) are uniformly distributed around zero error line, while the GB data points (crosses) exhibit a systematic error roughly proportional to the corresponding solvation energy.

in Fig. 3. Since in this analysis we use the same perfect effective Born radii in both the GB and ALPB, it is clear that the systematic error of the conventional GB model ultimately stems from its functional form, Eq. (1), which obviously misses some important physics. This point is further illustrated in Fig. 4, where the electrostatic part of the free energy of *transfer* between water and a low dielectric medium is computed and compared to the NPB reference for a typical medium-size protein in the biologically most relevant range of solvent dielectrics.

It is also noteworthy that the difference in solvation energies produced by the ALPB and our reference NPB solver PEP is comparable to that between two different NPB solvers, DELPHI and PEP. Namely, at $\epsilon_{in}=80$ the average disagreement between DELPHI and PEP is 5.99 ± 6.94 kcal/mol. So, the DELPHI solvation energies (0.25 Å grid resolution is used here, see Sec. III) are more shifted off the PEP reference, on average, than the ALPB energies (average error 2.21 ± 17.12 kcal/mol, see Table II), though the standard de-

TABLE II. Accuracy of the ALPB Eq. (3) is compared to that of the conventional GB method, Eq. (1), on a consensus benchmark set of 579 proteins, nucleic acid, and small peptide structures. Each entry is the difference, in kcal/mol, between the approximate electrostatic solvation free energy and the corresponding NPB reference, averaged over the consensus benchmark set. The error margin is computed as the standard deviation. In all of the calculations $\epsilon_{in}=1$.

ϵ_{out}	ALPB with different A			GB
	A_{NPB}	A_{ell}	A_{Det}	
8	-3.45 ± 31.40	-1.07 ± 31.35	-2.79 ± 31.62	-113.73 ± 76.73
20	-3.81 ± 23.38	-2.73 ± 23.23	-3.51 ± 23.43	-53.70 ± 41.68
40	-0.67 ± 19.04	-0.11 ± 18.94	-0.51 ± 19.06	-26.63 ± 26.33
80	2.13 ± 17.11	2.42 ± 17.06	2.21 ± 17.12	-11.11 ± 19.17

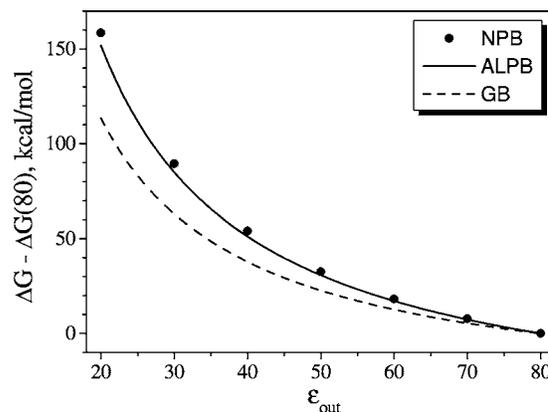


FIG. 4. Electrostatic part of the transfer energy between water and a medium of variable dielectric ϵ_{out} (at constant $\epsilon_{in}=1$) computed for the protein myoglobin using ALPB [with A_{Det} , Eq. (23)], conventional GB, and the reference NPB methods.

viation of the ALPB error is larger than that of DELPHI (but still smaller than that of the GB method). The disagreement between the two NPB solvers most likely reflects the subtle differences in their respective approximations of molecular surface as well as a different size of the finest grid, due to different computer memory requirements.

We emphasize that in these tests of the ALPB and GB we use perfect (NPB-based) effective Born radii. This choice eliminates the uncertainty associated with computing the effective radii and tests the quality of the ALPB formalism [Eq. (3)] directly. A great number of different ways to estimate effective Born radii have been developed over the past decade and a half, ranging from the very approximate to high quality, near perfect estimates.¹⁹ As with the GB model, the choice of a particular approximation for the effective radii to be used in ALPB is dictated by practical considerations and trade-offs between accuracy and speed: the point in discussed in some length in Ref. 43.

Finally, from Table II one can see that the ALPB method is equally accurate whether A_{NPB} , A_{ell} , or A_{Det} is used in Eq. (3). The numbers for \tilde{A}_{ell} (not shown in Table II) are the same as those for A_{ell} within 0.01 kcal/mol. Note that the approach that leads to A_{Det} is the simplest and provides straightforward analytical derivatives, Appendix B. In Appendix C, we present a procedure used to optimize the computation of data shown in Table II.

C. Incorporating the effects of salt

A typical biological environment contains electrolytes, whose electrostatic screening effects are often important to consider. A rigorous derivation of an ALPB equation that takes into account these effects is beyond the scope of this paper and will be published elsewhere. For immediate practical purposes it is possible to modify Eq. (3) so that it retains its simplicity and yet accounts for the screening effects of monovalent salt in the same manner and at the same level of accuracy as in the GB model.⁴⁴ Note that the GB formula Eq. (1) is the limiting case of the ALPB Eq. (3) with $\alpha \rightarrow 0$. Now consider the salt-dependent GB equation,

$$\Delta G_{\text{el}}^{\text{GB}}(\kappa) = -\frac{1}{2} \sum_{ij} q_i q_j \left(\frac{1}{\epsilon_{\text{in}}} - \frac{e^{-\kappa f_{ij}}}{\epsilon_{\text{out}}} \right) \frac{1}{f_{ij}}, \quad (24)$$

where $\kappa = 0.3163 \sqrt{[\text{salt}]}$ is the Debye-Hückel screening parameter of dimension \AA^{-1} when salt concentration is expressed in (mol/l). Clearly, a natural, albeit heuristic, candidate for a salt-dependent extension of the ALPB is

$$\Delta G_{\text{el}}^{\text{ALPB}}(\kappa) = -\frac{1}{2(1 + \alpha\beta)} \sum_{ij} q_i q_j \left(\frac{1}{\epsilon_{\text{in}}} - \frac{e^{-\kappa f_{ij}}}{\epsilon_{\text{out}}} \right) \times \left(\frac{1}{f_{ij}} + \frac{\alpha\beta}{A} \right). \quad (25)$$

The average errors of the salt-dependent ALPB and GB methods relative to the NPB reference, $\Delta \Delta G_{\text{el}}^{\text{ALPB}}(\kappa) = (\Delta G_{\text{el}}^{\text{ALPB}}(\kappa) - \Delta G_{\text{el}}^{\text{NPB}}(\kappa))$ and $\Delta \Delta G_{\text{el}}^{\text{GB}}(\kappa) = (\Delta G_{\text{el}}^{\text{GB}}(\kappa) - \Delta G_{\text{el}}^{\text{NPB}}(\kappa))$, are shown in Table III. Similar to the no-salt case considered above, the average error of ALPB amended with the salt prescription is smaller than the GB error by a factor of 3–4 at $\epsilon_{\text{out}} = 80$ and by a factor of 6–7 at $\epsilon_{\text{out}} = 20$. These differences are close to the ones we have seen when comparing the ALPB and GB without salt: the similarity indicates that the GB prescription for handling the salt effect works in the ALPB at the same level of accuracy as in the GB itself, and the accuracy improvement seen in Table III is due to Eq. (3) being more physically rigorous than Eq. (1) (Fig. 5).

D. Accuracy of per atom energy contributions

In many applications, such as MD simulations or pK estimates, it is essential that not only the total solvation energy ΔG_{el} but also the individual terms that sum up to it are calculated correctly. It may happen that the errors in $\Delta G_{ij}^{\text{el}}$ terms cancel upon summation, yielding the total energy that would appear quite accurate despite substantial errors in the individual $\Delta G_{ij}^{\text{el}}$ —a problem that is known to occur in earlier GB models.³⁹ Quantities such as the contribution to the force acting on i th atom ($-\nabla \sum_j \Delta G_{ij}^{\text{el}}$) may therefore contain an uncompensated error. To assess the accuracy of ALPB from this angle and to see if such deceptive error cancellation is actually taking place, we compute the per atom contributions to the solvation energy $\mathcal{E}_i = \sum_j \Delta G_{ij}^{\text{el}}$ by each of the approximate methods, ALPB and conventional GB. For each given molecule from the consensus benchmark set, the errors of the approximate methods relative to the NPB reference $\Delta \mathcal{E}_i = (\mathcal{E}_i - \mathcal{E}_i^{\text{NPB}})$ are computed for each atom. Then they are averaged over all atoms to produce a set of average $\langle \Delta \mathcal{E} \rangle$ for each molecule. These error values form fairly symmetric distributions; their averages over the consensus benchmark set represent systematic deviations of the corresponding models from the NPB reference, see Table IV. The systematic error of the ALPB method is three to a few hundred times lower than that of the GB method, depending on the solvent dielectric value.

E. Using ALPB in MD simulations

Given the mathematical similarity of the GB and ALPB models, it is hard to expect that the latter will prove to be

TABLE III. Accuracy of the ALPB with salt dependence added via Eq. (25) is compared to that of the conventional GB method with salt, Eq. (24), on the consensus benchmark set of 579 structures. The monovalent salt concentration is set to 0.1M. In both ALPB and GB calculations we follow the scaling prescription $\kappa \rightarrow 0.73\kappa$ due to Srinivasan *et al.*⁴⁴ that is intended to mimic the effects of nonzero ion exclusion radius. Each entry is the difference, in kcal/mol, between the approximate electrostatic solvation free energy and the corresponding NPB reference, averaged over the consensus benchmark set. The error margin is computed as the standard deviation. In all of the calculations $\epsilon_{\text{in}} = 1$. Here we do not consider media with lower dielectric values as salt is unlikely to dissolve in them to any appreciable extent.

ALPB with different A				
ϵ_{out}	A_{NPB}	A_{el}	A_{Det}	GB
20	6.47 ± 21.33	7.57 ± 21.63	6.77 ± 21.57	-43.68 ± 36.49
80	2.52 ± 17.05	2.80 ± 17.00	2.59 ± 17.06	-10.74 ± 18.97

unstable in molecular dynamics simulations where the GB model has been widely and successfully used for this purpose. Still, tests are needed. Since ALPB can use the effective Born radii already computed by many MD packages, the implementation of ALPB is particularly straightforward. We have incorporated ALPB into a prerelease version of AMBER9.⁴⁵ In the course of an MD simulation, the electrostatic solvation energy [Eq. (3)] is typically calculated every time step. A natural question arises, should A be calculated at every time step, too, or could one save computer resources and reduce the algorithmic complexity even further by recalculating A only once in a while? To address this question and test the stability of ALPB in a typical MD simulation, we have produced ~7 ns long MD simulation of protein ubiquitin, with Eq. (3) replacing the conventional GB model to describe implicit solvation. This protein was used before for testing the GB models.³⁰ The MD protocol used here is the same as in the above reference, except for the value of the solvent dielectric, see below, and the use of ALPB instead of the GB. Electrostatic size A_{Det} is calculated using Eq. (23). For select equidistant time points, reference A_{NPB} values are also computed. To explore the regime of maximum difference between the ALPB and GB, we set $\epsilon_{\text{out}} = 20$ which may correspond to solvation in concentrated water/alcohol mixtures; lower values of ϵ_{out} , although technically possible, are probably unreasonable for the case of protein solvation. Although we do not know if ubiquitin remains in its native state under the corresponding experimental conditions, it is unlikely that any global unfolding occurs over the course of 7 ns. It is therefore reassuring that the ALPB based simulation produced a stable trajectory, with the maximum backbone rmsd from the x-ray structure not exceeding 1.6 \AA (excluding three C-terminal residues). Very similar numbers for this protein were obtained with one of the latest GB models in AMBER (Ref. 46) under the conditions of aqueous solvation,³⁰ $\epsilon_{\text{in}} = 80$.

The evolution of the electrostatic size of ubiquitin in the course of the MD simulation is shown in Fig. 6. It is easy to see that A_{el} only slightly overestimates the reference NPB-based values: $(A_{\text{el}} - A_{\text{NPB}}) = 0.32 \pm 0.07 \text{ \AA}$, whereas A_{Det} underestimates the reference NPB values by $(A_{\text{NPB}} - A_{\text{Det}})$

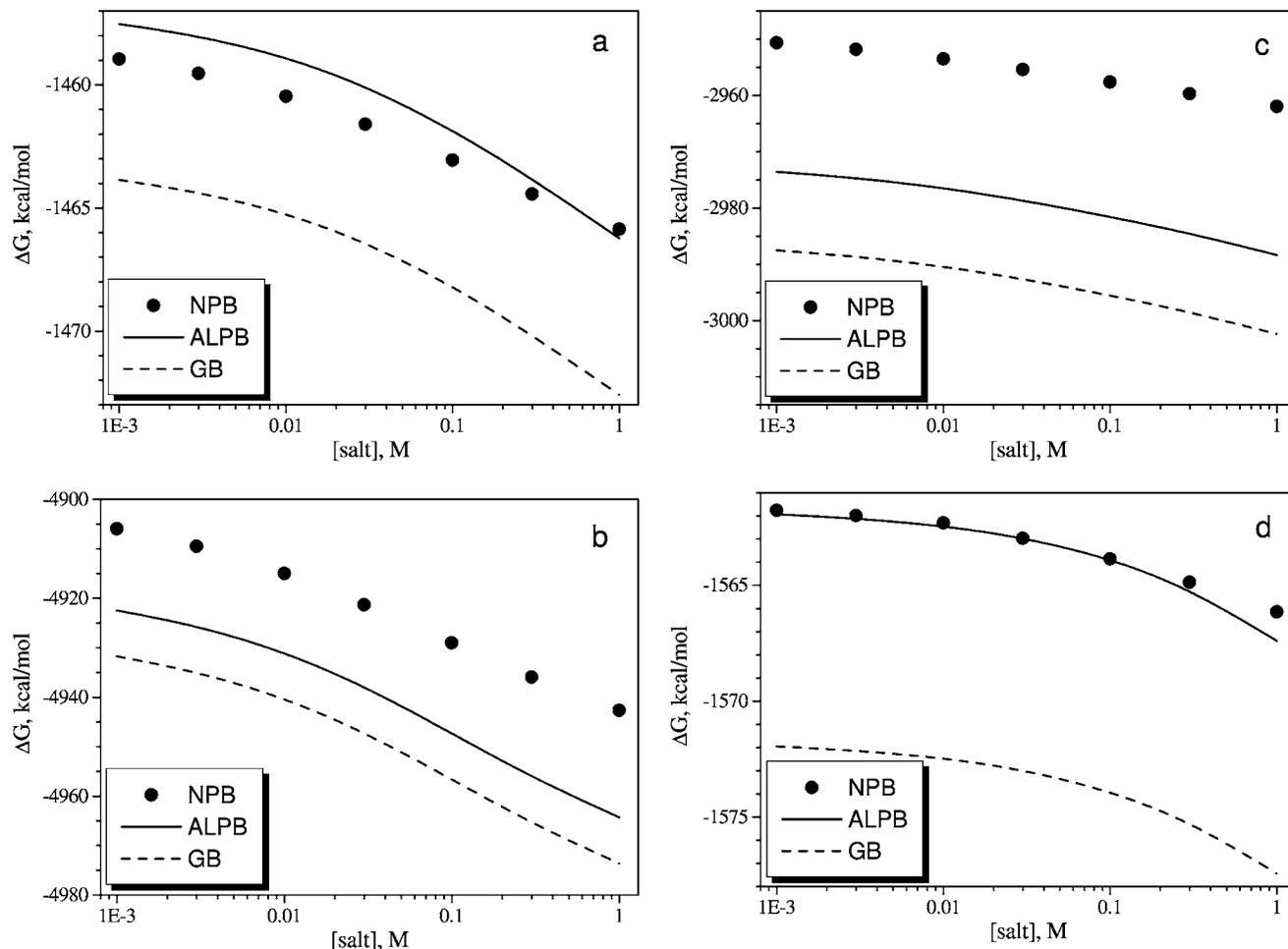


FIG. 5. The dependence of the electrostatic part of the free energy of solvation on the salt concentration within the ALPB [using A_{Det} , Eq. (23)] and the GB models is exemplified on four structures from the consensus benchmark set: (a) a small protein, leucine zipper acidic chain (PDB ID: 1fmh), (b) A-DNA (10 base-pair duplexes), (c) native myoglobin (PDB ID: 2mb5), and (d) thioredoxin (PDB ID: 2trx). The NPB reference values are also shown for comparison; in all of the calculations $\epsilon_{\text{in}}=1$, $\epsilon_{\text{out}}=80$.

$=0.62 \pm 0.05 \text{ \AA}$ (the differences in A are averaged over the trajectory). The differences are of the order of one to two standard deviations of A_{NPB} values themselves. Note that the variations in all three values of A are well correlated to each other over the time course of the simulation. Given the insensitivity of ALPB to variations in A discussed above, we conclude that the approximate methods provide useful estimates of A in this case. Another practically important observation is that in the course of the above simulation A changes by only a few tens of an angstrom; the amplitude of fluctuations of the electrostatic size is smaller than the error margin of the reference NPB values A_{NPB} . This conclusion is likely

TABLE IV. Systematic errors in the calculation of the on-site potentials (analytical methods vs NPB). Each number is the difference ($\mathcal{E}_i - \mathcal{E}_i^{\text{NPB}}$), where $\mathcal{E}_i = \sum_j \Delta G_{ij}^{\text{el}}$ (kcal/mol), averaged over all atoms of each molecule and then averaged again over all molecules from the consensus benchmark set. ALPB is using A_{Det} , Eq. (23), for the effective electrostatic size.

ϵ_{out}	ALPB	GB
20	-0.000 16	-0.032 16
40	0.000 58	-0.016 08
80	0.001 53	-0.006 96

to hold true for any MD simulation of proteins in or near their native states: since the electrostatic size characterizes the molecule's global shape, it will unlikely be changing appreciably in these types of simulations. Therefore, one does not need to recalculate A often; it is even acceptable to find A at the beginning of the simulation and then keep it constant.

Large structural transformations of a biomolecule such as protein folding or unfolding constitute a challenge to any analytical theory. To test our analytical approximations under these conditions, we use snapshots from an unfolding-refolding trajectory of protein-A. This trajectory was generated earlier³⁰ and samples a wide range of states, from the native, folded state (marked by F in Fig. 7) to the completely unfolded state at 450 K (U in Fig. 7). At each time step, the number of residue-residue contacts, which serves as a measure of protein's proximity to its folded state, and the electrostatic size A_{Det} are calculated; A_{NPB} is computed for select snapshots using the reference NPB solver. As expected, the electrostatic size of the protein correlates with the number of residue-residue contacts, i.e., with the degree of its compactness, see Fig. 7. One can also see that the solvation energy calculated by ALPB [Eq. (3)] correlates with NPB better than conventional GB values [Eq. (1)] at all times, even dur-

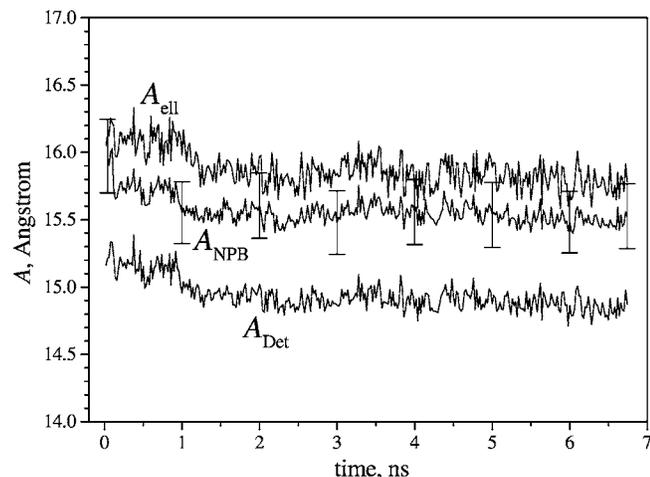


FIG. 6. Evolution of the electrostatic size of protein ubiquitin during MD simulation under the following conditions: $T=300$ K, $\epsilon_{in}=1$, and $\epsilon_{out}=20$. Solid lines show A_{ell} , A_{Det} , and A_{NPB} found at select points using procedure described in Sec. III C. Error bars (shown sparsely) correspond to the standard deviation σ' of NPB-based values of the electrostatic size of ubiquitin molecule.

ing drastic structural changes. The average (along dynamic trajectory) differences in the solvation energy values are as follows: $\overline{\Delta\Delta G_{el}^{GB}} = \langle \Delta G_{el}^{GB} - \Delta G_{el}^{NPB} \rangle = -44.98 \pm 5.67$ kcal, $\overline{\Delta\Delta G_{el}^{ALPB}} = \langle \Delta G_{el}^{ALPB} - \Delta G_{el}^{NPB} \rangle = 11.57 \pm 4.97$ kcal. Thus, the systematic error of the analytical approximation based on the ALPB approach is four times smaller than the error of the conventional GB method for this protein in very different structural states.

While an offset in energy that remained a constant throughout the entire trajectory would obviously not be a problem for MD applications, a bias towards certain structural states would. We have computed this bias towards the folded states of protein A within both the ALPB and the GB models, $\delta = \overline{\Delta\Delta G_{el}|_F} - \overline{\Delta\Delta G_{el}|_U}$. The definitions of the states are shown in Fig. 7, and the results are summarized in Table V. The artificial bias of the ALPB method is 1 kcal/mol and is almost an order of magnitude smaller than that produced by the GB model.

While the approximate analytical estimate of the electrostatic size deviates from the NPB reference values by as much as 30% over some parts of the protein A unfolding/refolding trajectory, Fig. 7, the variation of A_{NPB} itself is only about 15%. This suggests that using a constant $A=A(t=0)$ may not be completely unreasonable even if one expects substantial conformational changes to occur, as in this example.

Macromolecular complex formation and receptor-ligand docking are another important class of problems, especially in practical applications such as rational drug design. Implicit solvation approaches, such as the widely used molecular mechanics PB (GB)/solvent accessibility, [MMPB(GB)/SA] scheme,⁴⁷ are particularly advantageous here—they allow one to estimate free energies of many conformations in a computationally facile manner. In a typical scenario one estimates the free energy of complex formation as the difference between free energies of the complex and the state in which the ligand and the receptor are completely separated.^{48,49} The ALPB approach can serve to represent the

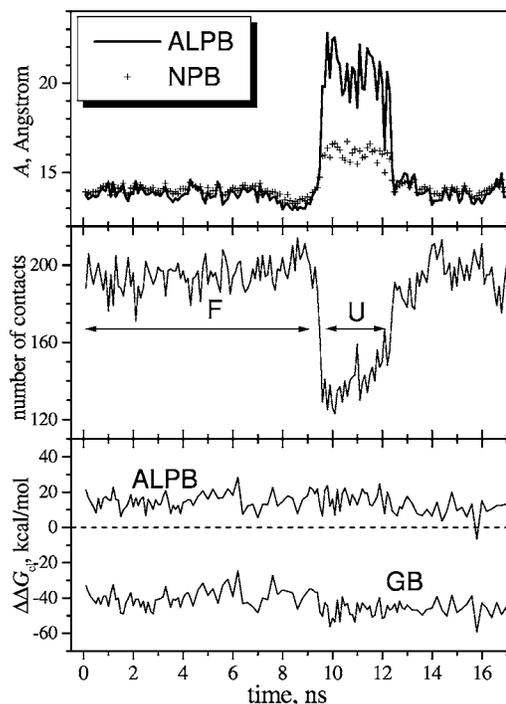


FIG. 7. (Top panel) Variation of electrostatic size A computed via the approximate method of effective ellipsoid [A_{Det} , Eq. (23)] is compared to the NPB reference (A_{NPB}) during the unfolding and refolding of protein A (PDB ID: lbdd) at $\epsilon_{in}=1$, $\epsilon_{out}=8$. Details of the unfolding protocol are found in Ref. 30. (Middle panel) Proximity to the folded states of the protein during the simulation is characterized by the number of residue-residue contacts. Ranges of the folded state (F , $0 < t < 9.3$ ns) and unfolded state (U , $9.6 < t < 12.3$ ns) are shown by arrows. (Bottom panel) The difference between electrostatic solvation energies computed by ALPB and NPB, $\Delta\Delta G_{el}^{ALPB} = \Delta G_{el}^{ALPB} - \Delta G_{el}^{NPB}$, and between conventional GB and NPB, $\Delta\Delta G_{el}^{GB} = \Delta G_{el}^{GB} - \Delta G_{el}^{NPB}$. Electrostatic size $A=A_{Det}$ is used in ALPB Eq. (3).

electrostatic solvation effects in this case. In the “docked” state, the components of the complex form one continuous structures, and so the ALPB can be used as described above to provide an increased accuracy over the GB treatment. In the other state, each molecule is treated by the ALPB separately, each characterized by its own value of the effective electrostatic size. Since the ligand-receptor separation distance is assumed to be infinite in this state, there are no ligand-receptor interactions to account for.

F. Note on parameter α of the ALPB equation

As was mentioned earlier, the ALPB model is parameter-free. We would like to stress that the coefficient α used in ALPB calculations in this work, Eq. (3), was *not* obtained by

TABLE V. The artificial bias towards the folded states in the folding/unfolding of protein A produced by the two approximations: ALPB [using A_{Det} , Eq. (23)] and GB. Average (along dynamic trajectory, see Fig. 7) values of the energy offset relative to the NPB reference $\overline{\Delta\Delta G_{el}^{GB}} = \langle \Delta G_{el}^{GB} - \Delta G_{el}^{NPB} \rangle$ and $\overline{\Delta\Delta G_{el}^{ALPB}} = \langle \Delta G_{el}^{ALPB} - \Delta G_{el}^{NPB} \rangle$ are calculated separately for the folded state, $0 < t < 9.3$ ns, and unfolded state, $9.6 < t < 12.3$ ns. The bias is defined as $\delta = \overline{\Delta\Delta G_{el}|_F} - \overline{\Delta\Delta G_{el}|_U}$.

Method	Bias δ (kcal/mol)
ALPB	-0.97
GB	7.21

fitting to a massive set of NPB solvation energies. Instead, the value

$$\alpha = \frac{4(3\pi^2 - 28)}{42\zeta(3) + 3\pi^2 - 76} - 1 \approx 0.571412, \quad (26)$$

where $\zeta(n)$ is the Riemann zeta function, originated from “first principles,”³⁷ through the minimization of the rms error between the *exact*³⁸ and the ALPB solutions of the Poisson equation on a large spherical molecule with a random distribution of charges. A natural question arises: Can this *ab initio* α be improved through fitting to the NPB reference solvation energies on the consensus set? To this end, we have computed an average error, relative to the NPB, of electrostatic solvation energy as a function of α : $\langle \Delta G_{\text{el}}^{\text{ALPB}}(\alpha) - \Delta G_{\text{el}}^{\text{NPB}} \rangle$. The results are shown in Fig. 8 for the consensus benchmark set of 579 molecules that we consider here. It can be seen that the *ab initio* α is virtually the same as the best fit value of ≈ 0.5898 . Note that the ALPB with $\alpha=0$ exactly coincides with conventional GB, compare Eqs. (1) and (3). The point corresponding to the difference between two NPB solvers, DELPHI and our reference solver PEP, is also shown on the graph. Note that DELPHI does not contain parameter α and is therefore shown out of the horizontal scale. Calculations show that not only the average error $\overline{\Delta \Delta G_{\text{el}}^{\text{ALPB}}}$ is vanishing near the *ab initio* $\alpha \approx 0.571412$ but also the standard deviation of the $\Delta \Delta G_{\text{el}}^{\text{ALPB}}$ value, as function of α , has its minimum at the same point. This means that the ALPB theory may not be improved by minor adjustments that would only shift the average error $\overline{\Delta \Delta G_{\text{el}}^{\text{ALPB}}}$, such as varying the parameter α . Since α enters Eq. (3) only in the combination $\alpha\beta$, where $\beta = \epsilon_{\text{in}}/\epsilon_{\text{out}}$, the electrostatic solvation energy ΔG_{el} is not very sensitive to α for small β such as, e.g., $\beta = 1/80$. For this reason $\alpha = 0.571412$ could be replaced, for the sake of simplicity, by an easy-to-remember value $\frac{4}{7} \approx 0.571429$, which would not change the final results to a noticeable extent.

V. CONCLUSIONS

The implicit solvent methodology is widely used to provide a computationally efficient representation of an aqueous environment in molecular modeling. Electrostatic interactions are often hardest to account for due to irregular shapes of realistic molecular structures. Considerable effort has been spent by the community to develop fast, analytical approaches to compute these interactions; the most widely used is arguably the generalized Born (GB) approximation. In this work we have continued to develop and test a new approach, the analytical linearized Poisson-Boltzmann (ALPB) that is introduced in a previous publication by the authors. The approach goes beyond the GB approximation in its accuracy and range of applicability but is as computationally efficient as the GB model. The main goal of the current work has been the development of effective approximations for the input parameters of the new model, extensive testing on a large representative set of biomolecular structures, and implementation and testing of the model in molecular dynamics (MD).

The main challenge proved to be the development of a

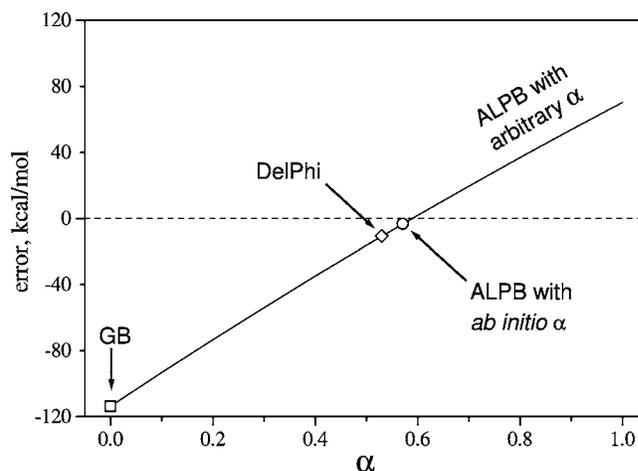


FIG. 8. Average error, relative to the NPB reference, of electrostatic solvation energies of 579 biomolecules from the consensus benchmark set, calculated at $\epsilon_{\text{in}}=1$, $\epsilon_{\text{out}}=8$ by various methods. Dependence of the ALPB (using A_{Det} , Eq. (23)) error on parameter α is shown. DELPHI and conventional GB energies shown for comparison do not depend on α .

computationally efficient approximation for the *effective electrostatic size* of the molecule—a new physical parameter that quantifies the relationship of the over-all size and shape of the molecular structure to its electrostatic contribution to the solvation free energy. As shown earlier, the effective electrostatic size reintroduces key physics into the approximate formalism, which is missing in the conventional GB theory. Similar to the effective Born radii in the GB model, the electrostatic size can, in principle, be computed by integrating the electric field (displacement) density around the molecule. But unlike the GB theory, approaches such as the Coulomb field approximation (CFA) widely used to compute the effective Born radii turn out to be not applicable. In this work we have proposed a different scheme: instead of approximating the complex electric field produced by an irregularly shaped molecule, we approximate its molecular surface by that of an ellipsoid and then compute the field (and the electrostatic size) exactly. While this approximation may seem like a drastic one for realistic biomolecules, the extensive tests against the standard numerical Poisson-Boltzmann (NPB) treatment have shown that it is quite adequate if used in ALPB to compute charge-charge interactions and solvation energies. Namely, the electrostatic part of the solvation free energy has been computed by ALPB and compared to NPB for almost 600 realistic biomolecules representing various structural classes. This statistically significant test reveals that ALPB virtually eliminates the systematic error relative to the NPB: the average differences are within 3.5 kcal/mol. Depending on the solvent dielectric, these systematic deviations are 5–50 times smaller than those of the conventional GB theory and, in contrast to the GB model, remain small over the entire relevant range of solvent dielectrics. Of course, this does not mean that for an arbitrary biomolecular structure the ALPB is guaranteed to be more accurate than the GB, but only that the ALPB is more likely to be more accurate.

In the aqueous solvation regime, the average deviation of the ALPB-computed solvation free energy from the NPB

reference is of the same magnitude as the difference between these energies computed by two popular NPB solvers. Since approximate models such as the ALPB or GB share the same physical foundations with the numerically exact Poisson-Boltzmann approach, the latter has been used as a standard here. Agreement with this standard is an absolutely necessary condition for any approximate physical model within the continuum solvent framework. Moreover, the evolution of the GB approach over the past decade and a half has demonstrated that at least in this particular case of approximate electrostatic models, a closer agreement with the PB generally translated into the better accuracy of the method over all. It can not be overemphasized, however, that agreement with the PB is only a necessary condition: tests against explicit solvent methods and ultimately against experiment will eventually decide the worthiness of our model and its place among others.

To lay the foundation for these types of tests, and ultimately for the use of ALPB in molecular modeling applications, we have tested the approximation in the context of MD simulations. This is where ALPB may be expected to make most impact due to its computational efficiency and improved accuracy relative to the GB model. Given the methodological focus of this work, we have refrained from applying the ALPB to full-fledged research problems and concentrated on testing its stability and computational performance in realistic biomolecular scenarios. To this end we have obtained the derivatives of the ALPB configurational energy with respect to atomic coordinates: the expressions turned out to be simple and well suited for numerical calculations. The electrostatic screening effects of monovalent salt have also been introduced into the new formalism, using the same philosophy as in the GB theory. We have tested the resulting model in a 7 ns MD simulation of a small protein ubiquitin, which produced a stable trajectory with virtually no additional computational overhead relative to an equivalent GB-based simulation. We have also tested the ALPB on a series of structures representing a temperature unfolding/refolding trajectory of a small protein. The ALPB provided consistently better accuracy than GB over the entire trajectory and virtually eliminated the unphysical bias (in solvation energy) towards the folded state that was present in the GB formalism. Interestingly, while the conformational changes involved are very large, the changes in the associated effective electrostatic size are only about 15%. For the native state simulation of protein ubiquitin these changes are considerably smaller. In the biologically relevant parameter regimes, the solvation energy is found to be not very sensitive to changes in the effective electrostatic size. Therefore, recomputing this quantity only occasionally, or even only once at the beginning of an MD simulation, appears to be a decent approximation. The last option is definitely recommended for simulations of native or near-native states and is probably useful in general: it simplifies the expressions for the corresponding component of the atomic force and virtually eliminates any possible overhead associated with the need to recompute the electrostatic size at every MD step.

Contrary to what one might expect from a theory derived on the basis of very simple shapes such as sphere or ellip-

soid, we find that the ALPB provides a reasonable approximation to the exact (or numerical) Poisson-Boltzmann formalism for many molecules of irregular shape, including those with pockets and cavities, and rodlike molecules such as DNA. Perhaps this is not so surprising given that most biological structures are globular, topologically equivalent to a sphere. While it is hard to make this argument precise, we have noticed that ALPB accuracy is worst (but still better than that of the GB) on structures that can be loosely described as having multiple distinct, distant, but connected domains of appreciable size. It should also be noted that our claim of higher accuracy of the ALPB relative to the GB implies the use of reasonably accurate effective Born radii in both models. While the “second generation” GB models do provide quite accurate sets of radii, we do not know how the ALPB may compare to the GB if older routines were used to compute these. Some of the original methods of computing the effective Born radii were notorious for underestimating them for buried atoms.

As mentioned before, the ALPB is free from parameters obtained by fitting. That is, all the parameters have a clear physical meaning and are derived on the basis of rigorous electrostatics applied to simple shapes. An attempt to optimize one of these parameters [α in Eq. (3)], using the common practice of fitting to a large set of molecular structures and their corresponding NPB solvation energies has failed, in the sense that it has produced a value of the optimized parameter virtually equal to the original “first principles” value. This result provides further support for the use of the simple shapes in rigorous derivations of approximate analytical models for biomolecular applications. The advantage of the first principles derivations of the type that led to the ALPB is that they provide further guidance and physical insights that are often impossible to deduce from massive fits.

The ALPB theory presented in this paper is noticeably more accurate but as computationally effective as the GB model in describing electrostatic effects of macromolecular solvation within the continuum solvent framework. Molecular dynamics simulations based on the ALPB model appear to be stable. While clearly many more tests are needed for a definitive conclusion, the new model might have the potential to replace the GB formalism in many of its current applications. Perhaps even more importantly, it provides a rigorous foundation for future development of physics-based analytical electrostatic models.

APPENDIX A: TRANSFORMATION OF EQ. (19) FOR NUMERICAL IMPLEMENTATION

To simplify the use of standard numerical integration techniques such as Simpson’s method, the elliptic integral found in Eq. (19) can be transformed into a sum of two integrals with finite integration limits, e.g., as follows:

$$\begin{aligned}
& \int_0^\infty \frac{d\theta}{\sqrt{(a^2 + \theta)(b^2 + \theta)(c^2 + \theta)}} \\
&= \int_0^d \frac{d\theta}{\sqrt{(a^2 + \theta)(b^2 + \theta)(c^2 + \theta)}} \\
&+ \int_d^\infty \frac{d\theta}{\sqrt{(a^2 + \theta)(b^2 + \theta)(c^2 + \theta)}} \\
&= \int_0^d \frac{d\theta}{\sqrt{(a^2 + \theta)(b^2 + \theta)(c^2 + \theta)}} \\
&+ \int_0^{1/\sqrt{d}} \frac{2d\xi}{\xi^3 \sqrt{(a^2 + \xi^{-2})(b^2 + \xi^{-2})(c^2 + \xi^{-2})}}, \quad (\text{A1})
\end{aligned}$$

where the substitution $\theta = \xi^{-2}$ is made in the second integral. A natural choice for parameter d , which breaks down the original infinite integration interval, is $d = a^2$,

$$\begin{aligned}
& \int_0^\infty \frac{d\theta}{\sqrt{(a^2 + \theta)(b^2 + \theta)(c^2 + \theta)}} \\
&= \int_0^{a^2} \frac{d\theta}{\sqrt{(a^2 + \theta)(b^2 + \theta)(c^2 + \theta)}} \\
&+ \int_0^{1/a} \frac{2d\xi}{\sqrt{(1 + a^2\xi^2)(1 + b^2\xi^2)(1 + c^2\xi^2)}}. \quad (\text{A2})
\end{aligned}$$

The functions in Eq. (A2) have no singularities within their limits of integration and are therefore well suited for numerical integration. In this work, the 1-4-1 Simpson's method is used to compute A_{el} .

APPENDIX B: ANALYTICAL DERIVATIVES OF THE SOLVATION FREE ENERGY

In ALPB theory (with the heuristic salt prescription), the electrostatic part of solvation energy ΔG_{el} is expressed as follows:

$$\begin{aligned}
\Delta G_{\text{el}} &\approx \sum_{ij} \Delta G_{ij}^{\text{ALPB}} \\
&= -\frac{1}{2(1 + \alpha\beta)} \sum_{ij} q_i q_j \left(\frac{1}{\epsilon_{\text{in}}} - \frac{e^{-\kappa f_{ij}}}{\epsilon_{\text{out}}} \right) \left(\frac{1}{f_{ij}} + \frac{\alpha\beta}{A} \right) \\
&= -\frac{1}{2\epsilon_{\text{in}}(1 + \alpha\beta)} \\
&\quad \times \sum_{ij} q_i q_j (1 - \beta e^{-\kappa f_{ij}}) \left(\frac{1}{f_{ij}} + \frac{\alpha\beta}{A} \right), \quad (\text{B1})
\end{aligned}$$

$$f_{ij} = \sqrt{r_{ij}^2 + R_i R_j} \exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right), \quad (\text{B2})$$

where $\alpha \approx 0.571412$, $\beta = \epsilon_{\text{in}}/\epsilon_{\text{out}}$, and A is the electrostatic size of the molecule. Under the usual molecular simulation conditions, α , β , ϵ_{in} , and κ are constants. The effective Born radii R_i depend on positions of all atoms. Generally speaking, A also depends upon atomic coordinates.

For the purpose of MD implementation, analytical derivatives of the free energy ΔG_{el} over atomic coordinates should be supplied, and are given below.

$$\frac{\partial \Delta G_{ij}^{\text{ALPB}}}{\partial x_k} = \frac{\partial \Delta G_{ij}^{\text{ALPB}}}{\partial f_{ij}} \frac{\partial f_{ij}}{\partial x_k} + \frac{\partial \Delta G_{ij}^{\text{ALPB}}}{\partial A} \frac{\partial A}{\partial x_k}; \quad (\text{B3})$$

$$\begin{aligned}
\frac{\partial \Delta G_{ij}^{\text{ALPB}}}{\partial f_{ij}} &= \frac{q_i q_j}{2\epsilon_{\text{in}}(1 + \alpha\beta)} \left[(1 - \beta e^{-\kappa f_{ij}}) \frac{1}{f_{ij}^2} \right. \\
&\quad \left. - \beta \kappa e^{-\kappa f_{ij}} \left(\frac{1}{f_{ij}} + \frac{\alpha\beta}{A} \right) \right], \quad (\text{B4})
\end{aligned}$$

$$\frac{\partial \Delta G_{ij}^{\text{ALPB}}}{\partial A} = \frac{q_i q_j}{2\epsilon_{\text{in}}(1 + \alpha\beta)} (1 - \beta e^{-\kappa f_{ij}}) \frac{\alpha\beta}{A^2}; \quad (\text{B5})$$

$$\begin{aligned}
\frac{\partial f_{ij}}{\partial x_k} &= \frac{1}{2f_{ij}} \frac{\partial}{\partial x_k} \left[r_{ij}^2 + R_i R_j \exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right) \right] \\
&= \frac{1}{2f_{ij}} \left[\left(1 + \frac{r_{ij}^2}{16R_i R_j} \exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right) \right) \frac{\partial r_{ij}^2}{\partial x_k} \right. \\
&\quad \left. + \left(1 - \frac{r_{ij}^4}{16R_i^2 R_j^2} \right) \exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right) \left(R_i \frac{\partial R_j}{\partial x_k} + R_j \frac{\partial R_i}{\partial x_k} \right) \right]; \quad (\text{B6})
\end{aligned}$$

$$\frac{\partial r_{ij}^2}{\partial x_k} = 2(\delta_{ik} - \delta_{jk})(x_i - x_j), \quad (\text{B7})$$

where δ_{ik} is the Kronecker delta. Derivatives over y_k (or z_k) are obtained by formal substitution of x by y (or z). Derivatives of R_i depend on the specific algorithm used to calculate the effective Born radii. Conventional GB formulas can be easily obtained from the above equations by substitution $\alpha = 0$.

For $A = A_{\text{Det}}$ given by Eq. (7), derivatives over atomic coordinates can be written down as follows:

$$\begin{aligned}
\frac{\partial A_{\text{Det}}}{\partial x_i} &= \sqrt{\frac{5}{2M}} \frac{1}{6(\text{Det } I)^{5/6}} \frac{\partial}{\partial x_i} \text{Det } I = \frac{125m_i}{48M^3 A_{\text{Det}}^5} \\
&\quad \times [x_i(I_{11}(I_{22} + I_{33}) - I_{12}^2 - I_{13}^2) + y_i(I_{12}I_{33} - I_{13}I_{23}) \\
&\quad + z_i(I_{13}I_{22} - I_{12}I_{32})]. \quad (\text{B8})
\end{aligned}$$

By analogy,

$$\begin{aligned}
\frac{\partial A_{\text{Det}}}{\partial y_i} &= \frac{125m_i}{48M^3 A_{\text{Det}}^5} \times [x_i(I_{12}I_{33} - I_{13}I_{23}) + y_i(I_{22}(I_{11} + I_{33}) \\
&\quad - I_{12}^2 - I_{23}^2) + z_i(I_{11}I_{23} - I_{12}I_{13})], \quad (\text{B9})
\end{aligned}$$

$$\begin{aligned}
\frac{\partial A_{\text{Det}}}{\partial z_i} &= \frac{125m_i}{48M^3 A_{\text{Det}}^5} \times [x_i(I_{13}I_{22} - I_{12}I_{32}) + y_i(I_{11}I_{23} \\
&\quad - I_{12}I_{13}) + z_i(I_{33}(I_{11} + I_{22}) - I_{13}^2 - I_{23}^2)]. \quad (\text{B10})
\end{aligned}$$

Substituting Eqs. (B8)–(B10) into Eq. (B3) completes the task of differentiating ΔG_{el} over atomic coordinates.

APPENDIX C: ADDITIONAL COMPUTATIONAL OPTIMIZATIONS

Finally, we provide the following optimizations for the calculation of $\Delta G_{\text{el}}^{\text{ALPB}}$ in a variable dielectric environment: for a molecule with given coordinates and charges of atoms, parameters

$$\mu = \sum_{ij} \frac{q_i q_j}{f_{ij}}, \quad \nu = \sum_{ij} q_i q_j e^{-\kappa f_{ij}}, \quad \omega = \sum_{ij} \frac{q_i q_j e^{-\kappa f_{ij}}}{f_{ij}} \quad (\text{C1})$$

can be calculated and used as invariant characteristics of the molecule. Each of these parameters is independent of the values of the internal and external dielectrics, and can be reused when the dielectric values change. Rewriting Eq. (25) with the new parameters yields a simple formula,

$$\Delta G_{\text{el}}^{\text{ALPB}} = - \frac{\mu - \beta\omega + \frac{\alpha\beta}{A}(q^2 - \beta\nu)}{2\epsilon_{\text{in}}(1 + \alpha\beta)}, \quad (\text{C2})$$

where $q = \sum_i q_i$ is the total charge of the molecule and $\beta = \epsilon_{\text{in}}/\epsilon_{\text{out}}$. Note that ν and ω depend on κ and that $\nu(\kappa=0) = q^2$, while $\omega(\kappa=0) = \mu$.

- ¹C. J. Cramer and D. G. Truhlar, Chem. Rev. (Washington, D.C.) **99**, 2161 (1999).
- ²B. Honig and A. Nicholls, Science **268**, 1144 (1995).
- ³P. Beroza and D. A. Case, Methods Enzymol. **295**, 170 (1998).
- ⁴J. D. Madura, M. E. Davis, M. K. Gilson, R. C. Wade, B. A. Luty, and J. A. McCammon, Rev. Comput. Chem. **5**, 229 (1994).
- ⁵M. K. Gilson, Curr. Opin. Struct. Biol. **5**, 216 (1995).
- ⁶M. Scarsi, J. Apostolakis, and A. Caflisch, J. Phys. Chem. A **101**, 8098 (1997).
- ⁷K. Chin, K. Sharp, B. Honig, and A. Pyle, Nat. Struct. Biol. **6**, 1055 (1999).
- ⁸R. Luo, L. David, and M. K. Gilson, J. Comput. Chem. **23**, 1244 (2002).
- ⁹M. Feig and C. L. Brooks III, Curr. Opin. Struct. Biol. **14**, 217 (2004).
- ¹⁰A. D. MacKerell, J. Comput. Chem. **25**, 1584 (2004).
- ¹¹N. Baker, Curr. Opin. Struct. Biol. **15**, 137 (2005).
- ¹²W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, J. Am. Chem. Soc. **112**, 6127 (1990).
- ¹³M. Schaefer and M. Karplus, J. Phys. Chem. **100**, 1578 (1996).
- ¹⁴S. Edinger, C. Cortis, P. Shenkin, and R. Friesner, J. Phys. Chem. B **101**, 1190 (1997).
- ¹⁵B. Jayaram, Y. Liu, and D. J. Beveridge, J. Chem. Phys. **109**, 1465 (1998).
- ¹⁶A. Aghosh, C. S. Rapp, and R. A. Friesner, J. Phys. Chem. **102**, 10983 (1998).
- ¹⁷D. Bashford and D. Case, Annu. Rev. Phys. Chem. **51**, 129 (2000).

- ¹⁸A. Onufriev, D. Bashford, and D. Case, J. Phys. Chem. B **104**, 3712 (2000).
- ¹⁹M. S. Lee, J. F. R. Salsbury, and C. L. Brooks III, J. Chem. Phys. **116**, 10606 (2002).
- ²⁰G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, Chem. Phys. Lett. **246**, 122 (1995).
- ²¹G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, J. Phys. Chem. **100**, 19824 (1996).
- ²²D. Qiu, P. Shenkin, F. Hollinger, and W. C. Still, J. Phys. Chem. A **101**, 3005 (1997).
- ²³A. K. Felts, Y. Harano, E. Gallicchio, and R. M. Levy, Proteins **56**, 310 (2004).
- ²⁴B. N. Dominy and C. L. Brooks, J. Phys. Chem. B **103**, 3765 (1999).
- ²⁵L. David, R. Luo, and M. K. Gilson, J. Comput. Chem. **21**, 295 (2000).
- ²⁶V. Z. Spassov, L. Yan, and S. Szalma, J. Phys. Chem. B **106**, 8726 (2002).
- ²⁷N. Calimet, M. Schaefer, and T. Simonson, Proteins **45**, 144 (2001).
- ²⁸V. Tsui and D. Case, J. Am. Chem. Soc. **122**, 2489 (2000).
- ²⁹T. Wang and R. Wade, Proteins **50**, 158 (2003).
- ³⁰A. Onufriev, D. Bashford, and D. A. Case, Proteins **55**, 383 (2004).
- ³¹E. Gallicchio and R. M. Levy, J. Comput. Chem. **25**, 479 (2004).
- ³²C. Simmerling, B. Strockbine, and A. E. Roitberg, J. Am. Chem. Soc. **124**, 11258 (2002).
- ³³H. Nymeyer and A. E. Garcia, Proc. Natl. Acad. Sci. U.S.A. **100**, 13934 (2003).
- ³⁴M. C. Lee and Y. Duan, Proteins **55**, 620 (2004).
- ³⁵K. Bernacki, C. Kalyanaraman, and M. P. Jacobson, J. Biomol. Screening **10**, 675 (2005).
- ³⁶P. Bonnet and R. A. Bryce, J. Mol. Graphics Modell. **24**, 147 (2005).
- ³⁷G. Sigalov, P. Scheffel, and A. Onufriev, J. Chem. Phys. **122**, 094511 (2005).
- ³⁸J. G. Kirkwood, J. Chem. Phys. **2**, 351 (1934).
- ³⁹A. Onufriev, D. Case, and D. Bashford, J. Comput. Chem. **23**, 1297 (2002).
- ⁴⁰G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers; Definitions, Theorems, and Formulas for Reference and Review* (McGraw-Hill, New York, 1968).
- ⁴¹W. R. Smythe, *Static and Dynamic Electricity* (Hemisphere, New York, 1989).
- ⁴²Freely available from <http://www.scripps.edu/mb/case/berozal/>
- ⁴³M. Feig, A. Onufriev, M. S. Lee, W. Im, D. A. Case, and C. L. Brooks III, J. Comput. Chem. **25**, 265 (2004).
- ⁴⁴J. Srinivasan, M. Trevathan, P. Beroza, and D. Case, Theor. Chem. Acc. **101**, 426 (1999).
- ⁴⁵Available from <http://amber.scripps.edu>
- ⁴⁶D. 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).
- ⁴⁷P. A. Kollman, I. Massova, C. Reyes *et al.*, Acc. Chem. Res. **33**, 889 (2000).
- ⁴⁸P. Ferrara, H. Gohlke, D. Price, G. Klebe, and C. Brooks, J. Med. Chem. **47**, 3032 (2004).
- ⁴⁹H. Gohlke and D. Case, J. Comput. Chem. **25**, 238 (2004).