

Generalized Born Model with a Simple Smoothing Function

WONPIL IM, MICHAEL S. LEE, CHARLES L. BROOKS, III

Department of Molecular Biology (TPC6), The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received 3 January 2003; Accepted 10 April 2003

Abstract: Based on recent developments in generalized Born (GB) theory that employ rapid volume integration schemes (M. S. Lee, F. R. Salabury, Jr., and C. L. Brooks III, *J Chem Phys* 2002, 116, 10606) we have recast the calculation of the self-electrostatic solvation energy to utilize a simple smoothing function at the dielectric boundary. The present GB model is formulated in this manner to provide consistency with the Poisson–Boltzmann (PB) theory previously developed to yield numerically stable electrostatic solvation forces based on finite-difference methods (W. Im, D. Beglov, and B. Roux, *Comp Phys Commun* 1998, 111, 59). Our comparisons show that the present GB model is indeed an efficient and accurate approach to reproduce corresponding PB solvation energies and forces. With only two adjustable parameters— a_0 to modulate the Coulomb field term, and a_1 to include a correction term beyond Coulomb field—the PB solvation energies are reproduced within 1% error on average for a variety of proteins. Detailed analysis shows that the PB energy can be reproduced within 2% absolute error with a confidence of about 95%. In addition, the solvent-exposed surface area of a biomolecule, as commonly used in calculations of the nonpolar solvation energy, can be calculated accurately and efficiently using the simple smoothing function and the volume integration method. Our implicit solvent GB calculations are about 4.5 times slower than the corresponding vacuum calculations. Using the simple smoothing function makes the present GB model roughly three times faster than GB models, which attempt to mimic the Lee–Richards molecular volume.

© 2003 Wiley Periodicals, Inc. *J Comput Chem* 24: 1691–1702, 2003

Key words: Born model; smoothing function; Poisson–Boltzmann

Introduction

Arguably, molecular simulations in which all solvent molecules are treated explicitly represent the most detailed approach to molecular modeling, protein folding, and the dynamics of biomolecules.¹ However, mainly due to the increased time demand of such simulations as the system size increases, considerable effort has been focused on the development of various implicit solvent models in which the average influence of the solvent on a solute is approximated implicitly.² In general, continuum electrostatics can be used to define the electrostatic potential and the electrostatic solvation energy of a solute with arbitrary shape by solving the Poisson–Boltzmann (PB) equation using finite-difference methods.^{3–5} There are many successful applications of the PB theory to complex biomolecular problems.^{6–8} Nonetheless, the computational cost of solving the PB equation has been a bottleneck in the application of PB theory to protein folding and dynamics of biomolecules.^{9,10}

Alternatively, inspired by the Born equation for solvation energies of ions,¹¹ the generalized Born (GB) model has been used quite successfully to estimate the electrostatic solvation energy

ΔG_{elec} .^{12–24} Because the GB model is intrinsically based on the same underlying continuum approximation as used in the PB theory, its accuracy is naturally assessed by comparison with the PB results. The most reliable GB formula was first proposed by Still et al.,¹²

$$\Delta G_{\text{elec}} = -\frac{1}{2} \tau \sum_{\alpha\beta} \frac{q_{\alpha} q_{\beta}}{\sqrt{r_{\alpha\beta}^2 + R_{\alpha}^{\text{GB}} R_{\beta}^{\text{GB}} \exp(-r_{\alpha\beta}^2/4R_{\alpha}^{\text{GB}} R_{\beta}^{\text{GB}})}} \quad (1)$$

where R_{α}^{GB} is the “effective Born radius” of atom α , and $\tau = 1/\epsilon_p - 1/\epsilon_s$, where ϵ_p represents the (low) dielectric response of the interior of the solute and ϵ_s the (high) solvent dielectric constant. ΔG_{elec} in Eq. (1) corresponds to the electrostatic free energy of transferring a solute in a medium of ϵ_p to a medium of

Correspondence to: C. L. Brooks, III; e-mail: brooks@scripps.edu

Contract/grant sponsor: National Institute of Health; contract/grant numbers: GM 37554 and RR 12255

Contract/grant sponsor: NSF (through the CTBP)

ϵ_s . In the present study, ϵ_p is simply set to one to be consistent with the molecular mechanics force field. In principle, the “exact” effective Born radii can be calculated by performing PB calculations for one atom at a time by setting all other charges to zero, and then by inserting the calculated self (or atomic) electrostatic solvation energy into the Born equation. The radius obtained from the Born equation represents the distance between a particular atom and the “effective” spherical dielectric boundary. The prime assumption in the GB method is that the solvent-shielded charge–charge interactions in PB can be reproduced by the cross-term in Eq. (1) using the same effective Born radii. Indeed, the GB model of Eq. (1) has been shown to reproduce ΔG_{elec} from PB calculations excellently provided that the effective Born radii are accurate.^{21,23} Thus, the majority of literature on extensions of the GB theory has been focused on the efficient and accurate evaluation of the Born radii.^{12–24}

In continuum electrostatics, the self-electrostatic solvation energy can be expressed rigorously in terms of an integral of a spatially dependent electrostatic field density.^{16,17} Most GB models have approximated the electrostatic field as the Coulomb field, neglecting the reaction field, which is generated by the charge density arising from solvent polarization at the dielectric boundary. This is the so-called “Coulomb field approximation.” To obtain the effective Born radii efficiently, the integral is then evaluated by numerical surface/volume integration methods,^{12,16,17,21,24} which are more rigorous than conventional pairwise summation approximations.^{13,15,18} It is well known that the Coulomb field approximation itself underestimates the self-solvation energy, and thus overestimates the effective Born radii compared to the exact Born radii calculated by PB.²¹ Recently, using a numerical volume integration method with an empirical correction term beyond Coulomb field, Lee et al. showed a significant improvement over the Coulomb field approximation: 1% error on average was achieved for absolute electrostatic solvation energies of a large set of proteins and for relative solvation energies of protein conformations, compared to the corresponding PB solvation energies.^{21,24}

In the surface/volume integration approach, one has to define the dielectric boundary by which the interior (low dielectric region) and the exterior (high dielectric region) of a solute are divided. In continuum electrostatics, the energy of the system strongly depends on the definition of the dielectric boundary. Conventionally, either the van der Waals surface (overlapping atomic spheres) or molecular surface (including the reentrant surface) have been used to define the discontinuous dielectric boundary.^{25,26} However, it is well known that the discontinuity at the dielectric boundary causes numerical instabilities in the calculation of solvation forces using both PB^{27,28} and GB with the integration approach.^{21,24} This problem may be avoided by using a continuous and smooth function at the dielectric boundary. Which smoothing function is best to use is closely related to which dielectric boundary is used in PB. Most GB models have been optimized to reproduce the PB results with molecular surface, which might be the best representation in continuum electrostatics because of its tendency to reduce artificial high dielectric constant regions in small voids and crevices in biomolecules. However, there is no way to calculate the molecular surface analytically. Indeed, Lee et al. have developed an elaborate scheme using a Gaussian-type

smoothing function to mimic the molecular surface and to obtain numerically stable solvation forces in their GB model.^{21,24}

The present study is motivated by the fact that there is no analytical GB model that uses the same definition of the dielectric boundary as used in PB. To obtain numerically stable solvation forces in PB, Im et al. employed a simple polynomial smoothing function at the dielectric boundary, which continuously changes the dielectric constant from ϵ_p in the interior of the solute to ϵ_s in the solvent region.²⁸ Here, we develop an equivalent GB theory using the volume integration method and the correction term beyond Coulomb field introduced by Lee et al.^{21,24} Although the dielectric boundary defined by the simple smoothing function does not correspond to the molecular surface representation, the present GB model has several advantages. First, using the simple smoothing function speeds up the GB calculation over GB models that try to mimic the molecular surface.^{21,24} Second, by using the same type of smoothing function as already employed in PB solvers, we can compare the electrostatic solvation energy as well as the solvation forces between PB and GB. Last, as shown by Im et al.,²⁸ it is relatively easy to calculate the surface of a solute approximately using the simple smoothing function.

The rest of the article is organized as follows. In the next section the background and theoretical developments are given in detail with the numerical implementation of the method. This is followed by numerical tests that compare the electrostatic solvation energy and forces from PB to those from GB. The calculated surface area is also compared with the exact surface. Finally, the computational cost of the present GB model is examined. The article concludes with a brief summary of the main results.

Theoretical Development

The solvation free energy corresponding to the work required to insert a solute of fixed conformation into a polar solvent can be expressed as the sum of nonpolar (np) and electrostatic (elec) contributions, i.e., $\Delta G_{\text{solv}} = \Delta G_{\text{elec}} + \Delta G_{\text{np}}$.² The nonpolar solvation energy ΔG_{np} includes the formation of a cavity in the solvent as well as solvent–solute van der Waals interactions. This energy term is often expressed as the product of the solvent-exposed surface area, S , of the solute and a phenomenological surface tension coefficient γ .^{27,29,30}

$$\Delta G_{\text{np}} = \gamma S. \quad (2)$$

The electrostatic solvation energy ΔG_{elec} is the work required to assemble the charges $\{q_\alpha\}$ of the solute in the solvent, and can be expressed in terms of the reaction field potential $\phi_{\text{rf}}(\mathbf{r})$ or the reaction field Green’s function $G_{\text{rf}}(\mathbf{r}, \mathbf{r}')$,^{3,4,31}

$$\Delta G_{\text{elec}} = \frac{1}{2} \sum_{\alpha} q_{\alpha} \phi_{\text{rf}}(\mathbf{r}_{\alpha}) = \frac{1}{2} \sum_{\alpha\beta} q_{\alpha} G_{\text{rf}}(\mathbf{r}_{\alpha}, \mathbf{r}_{\beta}) q_{\beta}, \quad (3)$$

where $G_{\text{rf}}(\mathbf{r}, \mathbf{r}')$ corresponds to the reaction field potential at \mathbf{r} due to a point charge at \mathbf{r}' . Based on continuum electrostatics, in which the solvent is represented as a featureless high dielectric medium,

the reaction field potential $\phi_{\text{rf}}(\mathbf{r})$ can be computed by solving the PB equation numerically using finite-difference methods,^{3,4,5,28}

$$\nabla \cdot [\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] - \bar{\kappa}^2(\mathbf{r}) \phi(\mathbf{r}) = -4\pi \rho(\mathbf{r}), \quad (4)$$

where $\varepsilon(\mathbf{r})$, $\bar{\kappa}(\mathbf{r})$, and $\rho(\mathbf{r})$ are the dielectric constant, the modified Debye–Hückel screening factor, and the fixed charge density of the solute, respectively. Even though it is possible to obtain (numerically) stable electrostatic solvation forces in the context of the finite-difference method,²⁸ the calculations are too expensive to perform long molecular dynamics (MD) simulations of biomolecules. An alternative and efficient approximation is given by the GB model of Eq. (1), which uses a simple analytical formula for the reaction field Green's function $G_{\text{rf}}(\mathbf{r}, \mathbf{r}')$ in Eq. (3).¹² It can be shown that Eq. (1) is exact for a pair of atoms, α and β , in the limiting cases of $r_{\alpha\beta} \rightarrow 0$ and $r_{\alpha\beta} \rightarrow \infty$.^{12,20}

Effective Born Radii Evaluation

The quantitative agreement between PB ΔG_{elec} and GB ΔG_{elec} strongly depends on the effective Born radii $\{R_{\alpha}^{\text{GB}}\}$.^{21,23} Using the Born equation,¹¹ one can extract the exact Born radius, R_{α}^{GB} , of atom α in a solute from its self-electrostatic free energy, $\Delta G_{\text{elec},\alpha}$, calculated by solving the PB equation [Eq. (4)] after setting all other charges to zero,

$$\Delta G_{\text{elec},\alpha} = -\frac{1}{2} \tau \frac{q_{\alpha}^2}{R_{\alpha}^{\text{GB}}}. \quad (5)$$

Thus, $\Delta G_{\text{elec},\alpha}$ or R_{α}^{GB} from PB calculations serve as a benchmark to assess the quality of the effective Born radii calculated by various approximate GB theories.

Based on the Coulomb field approximation,^{12,16,20} one can express the self-electrostatic solvation energy, $\Delta G_{\text{elec},\alpha}$, as a volume integration,

$$\Delta G_{\text{elec},\alpha}^0 = -\frac{1}{2} \tau q_{\alpha}^2 \left(\frac{1}{\eta_{\alpha}} - \frac{1}{4\pi} \int_{r>\eta_{\alpha}} d\mathbf{r} \frac{\nu(\mathbf{r}; \{\mathbf{r}_{\alpha}\})}{|\mathbf{r} - \mathbf{r}_{\alpha}|^4} \right), \quad (6)$$

where η_{α} is an arbitrarily defined integration starting point, necessary to avoid the singularity at $|\mathbf{r} - \mathbf{r}_{\alpha}| = 0$, and $\nu(\mathbf{r})$ is a solute volume function that is one in the interior of a solute and zero in the solvent region. Because the Coulomb field approximation neglects the reaction field, it is well known that this approximation underestimates the self-solvation energy and thus overestimates the effective atomic Born radii compared to the exact ones calculated from Eq. (4).²¹ In principle, one can express the exact atomic solvation energy, $\Delta G_{\text{elec},\alpha}$, as the sum of a series of correction terms beyond the Coulomb field approximation. Recently, Lee et al. have used an empirical correction term, $\Delta G_{\text{elec},\alpha}^1$, to the Coulomb field term, $\Delta G_{\text{elec},\alpha}^0$, and demonstrated great improvement over the Coulomb field approximation for the calculated effective Born radii.^{21,24} Using the correction term one can approximate $\Delta G_{\text{elec},\alpha}$ as

$$\Delta G_{\text{elec},\alpha} \approx a_0 \Delta G_{\text{elec},\alpha}^0 + a_1 \Delta G_{\text{elec},\alpha}^1, \quad (7)$$

where a_0 and a_1 are empirical coefficients, and $\Delta G_{\text{elec},\alpha}^1$ is given by,²⁴

$$\Delta G_{\text{elec},\alpha}^1 = -\frac{1}{2} \tau q_{\alpha}^2 \left(\frac{1}{4\eta_{\alpha}^4} - \frac{1}{4\pi} \int_{r>\eta_{\alpha}} d\mathbf{r} \frac{\nu(\mathbf{r}; \{\mathbf{r}_{\alpha}\})}{|\mathbf{r} - \mathbf{r}_{\alpha}|^7} \right)^{1/4}. \quad (8)$$

In continuum electrostatics the definition of the dielectric boundary is not unique. Moreover, it is well known that a discontinuity at the dielectric boundary causes numerical instability in calculations of solvation forces using finite-difference PB equations.^{27,28} To avoid this problem, a continuous and smooth dielectric boundary must be employed. It is also clear from Eqs. (6) and (8) that the volume function $\nu(\mathbf{r})$ should be continuous and smooth to obtain stable solvation forces in the GB model. For this reason the spatially dependent dielectric constant $\varepsilon(\mathbf{r})$ in Eq. (4) was defined as a (smooth) volume exclusion function, $\mathcal{H}(\mathbf{r})$, going from zero in the interior of a solute to one in the solvent region, in calculations of solvation forces based on finite-difference PB calculations,²⁸

$$\varepsilon(\mathbf{r}) = 1 + (\varepsilon_s - 1) \mathcal{H}(\mathbf{r}; \{\mathbf{r}_{\alpha}\}). \quad (9)$$

$\mathcal{H}(\mathbf{r})$ is a function of all atomic positions $\{\mathbf{r}_{\alpha}\}$ in the system, and can be expressed as a product of a simple polynomial atomic volume exclusion function $H_{\alpha}(r)$,²⁸

$$\mathcal{H}(\mathbf{r}; \{\mathbf{r}_{\alpha}\}) = \prod_{\alpha} H_{\alpha}(|\mathbf{r} - \mathbf{r}_{\alpha}|), \quad (10)$$

where

$$H_{\alpha}(r) = \begin{cases} 0, & r \leq R_{\alpha}^{\text{PB}} - w \\ \frac{1}{2} + \frac{3}{4w} (r - R_{\alpha}^{\text{PB}}) - \frac{1}{4w^3} (r - R_{\alpha}^{\text{PB}})^3, & R_{\alpha}^{\text{PB}} - w < r < R_{\alpha}^{\text{PB}} + w \\ 1, & r \geq R_{\alpha}^{\text{PB}} + w \end{cases} \quad (11)$$

where r is the distance between a spatial point and atom α , R_{α}^{PB} is the atomic radius to define the dielectric boundary in PB calculations, and $2w$ is a smoothing length that confines the region where the smoothing function is applied. The first derivative of the smoothing function is zero at $R_{\alpha}^{\text{PB}} - w$ and $R_{\alpha}^{\text{PB}} + w$. It should be noted that the smoothed dielectric boundary does not correspond exactly to the van der Waals surface representation, even though one can recover the van der Waals surface by setting w to zero. Thus, one can expect some difference between PB results calculated with the smoothed dielectric boundary or the boundary defined by van der Waals surface if the same PB radii $\{R_{\alpha}^{\text{PB}}\}$ are used. Indeed, simple analysis based on the solvation of an ion in water showed that the smoothed boundary causes a slight shift of the induced solvent charge

density toward the ionic center from the PB radius, i.e., the solvent charge is mostly localized over the distance between $R^{\text{PB}} - w$ and R^{PB} .³² This shift results in an overestimation of the PB solvation energy compared to the PB energy with a discontinuous van der Waals surface. Considering the influence of the smoothed boundary on the PB energy, Nina et al. used an empirical formula to modify the optimal protein PB radii $\{R_{\alpha}^{\text{PB0}}\}$ to closely reproduce the PB energies calculated with the van der Waals surface as well as the charging free energies calculated by MD free energy simulations (MD/FES) for 20 standard amino acids.³³

$$R_{\alpha}^{\text{PB}} = s(R_{\alpha}^{\text{PB0}} + w), \quad (12)$$

where s is a scaling factor with a value close to 1 and is shown in Table 2 in the article by Nina et al. (1999).³² Interestingly, PB ΔG_{elec} calculated for proteins increases as w increases and the extent of its augmentation also increases as the protein size increases. For example, it varies from -893.7 kcal/mol ($w = 0.1$ Å) to -856.9 ($w = 0.5$ Å) for protein 1AJJ (PDB code: 37 amino acids), and from -3670 kcal/mol ($w = 0.1$ Å) to -3443 ($w = 0.5$ Å) for protein 1AK6 (PDB code: 174 amino acids). Our interpretation of this result is that the overlapped smoothed dielectric boundary region that was assigned a high dielectric constant in the van der Waals surface definition is now assigned a low dielectric constant. This suggests that the smoothed dielectric boundary has some ability to prevent high dielectric constants from being assigned in small voids and crevices in biomolecules, even though this surface is rather closely related to the van der Waals surface representation.

It is straightforward to link the volume exclusion function $\mathcal{H}(\mathbf{r})$ in PB to the volume function $\nu(\mathbf{r})$ in GB, i.e.,

$$\nu(\mathbf{r}; \{\mathbf{r}_{\alpha}\}) = 1 - \mathcal{H}(\mathbf{r}; \{\mathbf{r}_{\alpha}\}). \quad (13)$$

There are both advantages and disadvantages in using such a simple smoothing function. One disadvantage is that most GB models have been optimized to reproduce the PB results with the molecular surface,^{21,24} and thus, the energies are not directly comparable between these models. Nonetheless, the present GB model also has a number of advantages regarding the speedup in GB calculations and the direct comparison between PB and GB solvation forces (see Numerical Tests and Discussion). Furthermore, as proposed by Im et al., it is relatively easy to calculate the solvent-exposed surface area approximately using the simple smoothing function,²⁸

$$\begin{aligned} S &\approx \int d\mathbf{r} \|\nabla \mathcal{H}(\mathbf{r}; \{\mathbf{r}_{\alpha}\})\| \\ &= \sum_{\alpha} \int d\mathbf{r} \|\nabla H_{\alpha}(|\mathbf{r} - \mathbf{r}_{\alpha}|)\| \prod_{\beta \neq \alpha} H_{\beta}(|\mathbf{r} - \mathbf{r}_{\beta}|). \end{aligned} \quad (14)$$

Numerical Implementation

Following recent developments of Lee et al.,²¹ we used numerical quadrature techniques for the integration of Eqs. (6) and (8). In this context, Eqs. (6) and (8) can be approximated as

$$\Delta G_{\text{elec},\alpha}^0 \approx -\frac{1}{2} \tau q_{\alpha}^2 \left(\frac{1}{\eta_{\alpha}} - \frac{1}{4\pi} \sum_m \sum_n w_m w_n \frac{\nu(\mathbf{r}_{\alpha} + \mathbf{r}_{mn}; \{\mathbf{r}_{\alpha}\})}{r_{mn}^2} \right) \quad (15)$$

and

$$\Delta G_{\text{elec},\alpha}^1 \approx -\frac{1}{2} \tau q_{\alpha}^2 \left(\frac{1}{4\eta_{\alpha}^4} - \frac{1}{4\pi} \sum_m \sum_n w_m w_n \frac{\nu(\mathbf{r}_{\alpha} + \mathbf{r}_{mn}; \{\mathbf{r}_{\alpha}\})}{r_{mn}^5} \right)^{1/4}, \quad (16)$$

where indices m and n are for radial and angular integration points, w_m and w_n are weights for radial and angular integration components, and \mathbf{r}_{mn} is the integration point centered at the origin. In the present study, the integration points and weights for the radial component are generated by the Gaussian–Legendre quadrature³⁴ and those for the angular component by the Lebedev quadrature.³⁵ Therefore, the integration can be viewed as a mapping of those integration points onto each atom followed by the summation over spherically arranged integration points around each atom. In terms of Eqs. (5) and (7), the effective atomic Born radii are given by

$$R_{\alpha}^{\text{GB}} = -\frac{1}{2} \tau q_{\alpha}^2 \frac{1}{a_0 \Delta G_{\text{elec},\alpha}^0 + a_1 \Delta G_{\text{elec},\alpha}^1}. \quad (17)$$

It should be noted that the effective Born radii do not depend on atomic charges but only on the conformation of a solute as well as atomic PB radii, as seen in Eq. (17).

The calculation of solvation forces becomes complicated because the effective Born radius of atom α is a function of all other atoms as well as atom α through the fact that the volume function $\nu(\mathbf{r}; \{\mathbf{r}_{\alpha}\})$ is a function of both a position in the system and all atomic positions. Nonetheless, one can express the electrostatic solvation force acting on atom α as

$$\frac{\partial \Delta G_{\text{elec}}}{\partial \mathbf{r}_{\alpha}} = \frac{\partial \Delta G_{\text{elec}}}{\partial r_{\alpha\beta}} \frac{\partial r_{\alpha\beta}}{\partial \mathbf{r}_{\alpha}} + \sum_{\beta} \frac{\partial \Delta G_{\text{elec}}}{\partial R_{\beta}^{\text{GB}}} \frac{\partial R_{\beta}^{\text{GB}}}{\partial \mathbf{r}_{\alpha}}. \quad (18)$$

The first term becomes

$$\frac{\partial \Delta G_{\text{elec}}}{\partial r_{\alpha\beta}} \frac{\partial r_{\alpha\beta}}{\partial \mathbf{r}_{\alpha}} = -\frac{\tau}{4} \sum_{\beta \neq \alpha} \frac{q_{\alpha} q_{\beta} [4 - \exp(-D_{\alpha\beta})]}{[r_{\alpha\beta}^2 + R_{\alpha}^{\text{GB}} R_{\beta}^{\text{GB}} \exp(-D_{\alpha\beta})]^{3/2}} (\mathbf{r}_{\beta} - \mathbf{r}_{\alpha}), \quad (19)$$

where

$$D_{\alpha\beta} = \frac{r_{\alpha\beta}^2}{4R_{\alpha}^{\text{GB}}R_{\beta}^{\text{GB}}}. \quad (20)$$

In the second term of Eq. (18), the derivatives of ΔG_{elec} relative to effective atomic Born radii simply become

$$\frac{\partial \Delta G_{\text{elec}}}{\partial R_{\beta}^{\text{GB}}} = \frac{\tau}{2} \sum_{\gamma} \frac{q_{\beta} q_{\gamma} \exp(-D_{\beta\gamma})}{[r_{\beta\gamma}^2 + R_{\beta}^{\text{GB}} R_{\gamma}^{\text{GB}} \exp(-D_{\beta\gamma})]^{3/2}} \left(R_{\gamma}^{\text{GB}} + \frac{r_{\beta\gamma}^2}{4R_{\beta}^{\text{GB}}} \right). \quad (21)$$

However, the derivatives of effective atomic Born radii relative to atom α are more involved, such as

$$\begin{aligned} \frac{\partial R_{\beta}^{\text{GB}}}{\partial \mathbf{r}_{\alpha}} &= \frac{(R_{\beta}^{\text{GB}})^2}{4\pi} \sum_m \sum_n w_m w_n \frac{\partial \nu(\mathbf{r}_{\beta} + \mathbf{r}_{mn})}{\partial \mathbf{r}_{\alpha}} \\ &\times \left[\frac{a_0}{r_{mn}^2} - \frac{1}{4} \left(\frac{1}{2} \frac{\tau q_{\beta}^2}{\Delta G_{\text{elec},\beta}^1} \right)^3 \frac{a_1}{r_{mn}^5} \right], \end{aligned} \quad (22)$$

where the derivatives of the volume functions $\nu(\mathbf{r})$ relative to atom α become, when, $\beta = \alpha$,

$$\frac{\partial \nu(\mathbf{r}_{\alpha} + \mathbf{r}_{mn})}{\partial \mathbf{r}_{\alpha}} = - \sum_{\gamma \neq \alpha} \frac{\partial H_{\gamma}(|\mathbf{r}_{\alpha} + \mathbf{r}_{mn} - \mathbf{r}_{\gamma}|)}{\partial \mathbf{r}_{\alpha}} \frac{\mathcal{H}(|\mathbf{r}_{\alpha} + \mathbf{r}_{mn}|)}{H_{\gamma}(|\mathbf{r}_{\alpha} + \mathbf{r}_{mn} - \mathbf{r}_{\gamma}|)}, \quad (23)$$

and when $\beta \neq \alpha$,

$$\frac{\partial \nu(\mathbf{r}_{\beta} + \mathbf{r}_{mn})}{\partial \mathbf{r}_{\alpha}} = - \frac{\partial H_{\alpha}(|\mathbf{r}_{\beta} + \mathbf{r}_{mn} - \mathbf{r}_{\alpha}|)}{\partial \mathbf{r}_{\alpha}} \frac{\mathcal{H}(|\mathbf{r}_{\beta} + \mathbf{r}_{mn}|)}{H_{\alpha}(|\mathbf{r}_{\beta} + \mathbf{r}_{mn} - \mathbf{r}_{\alpha}|)}. \quad (24)$$

The derivatives of the atomic volume function $H(\mathbf{r})$ in Eqs. (23) and (24) are, respectively,

$$\frac{\partial H_{\gamma}(|\mathbf{r} - \mathbf{r}_{\gamma}|)}{\partial \mathbf{r}} = \left(\frac{3}{4w} - \frac{3}{4w^3} (|\mathbf{r} - \mathbf{r}_{\gamma}| - R_{\gamma}^{\text{PB}})^2 \right) \frac{\mathbf{r} - \mathbf{r}_{\gamma}}{|\mathbf{r} - \mathbf{r}_{\gamma}|} \quad (25)$$

and

$$\frac{\partial H_{\alpha}(|\mathbf{r} - \mathbf{r}_{\alpha}|)}{\partial \mathbf{r}_{\alpha}} = - \left(\frac{3}{4w} - \frac{3}{4w^3} (|\mathbf{r} - \mathbf{r}_{\alpha}| - R_{\alpha}^{\text{PB}})^2 \right) \frac{\mathbf{r} - \mathbf{r}_{\alpha}}{|\mathbf{r} - \mathbf{r}_{\alpha}|}. \quad (26)$$

Similarly, the surface of a solute can be also easily calculated using Eq. (14) and the numerical quadrature techniques,

$$S \approx \sum_{\alpha} \sum_m \sum_n w_m w_n \|\nabla H_{\alpha}(r_{mn})\| \prod_{\beta \neq \alpha} H_{\beta}(|\mathbf{r}_{\alpha} + \mathbf{r}_{mn} - \mathbf{r}_{\beta}|). \quad (27)$$

It should be noted that the numerical integration needs to be performed only in the smoothed boundary region because the gradient of atomic volume function $H(\mathbf{r})$ is zero elsewhere. Ac-

cording to Eq. (2), the nonpolar solvation force acting on atom α is proportional to the derivatives of the surface relative to atom α ,

$$\begin{aligned} \frac{\partial S}{\partial \mathbf{r}_{\alpha}} &= \sum_m \sum_n w_m w_n \|\nabla H_{\alpha}(r_{mn})\| \\ &\times \left(\sum_{\beta \neq \alpha} \frac{\partial H_{\beta}(|\mathbf{r}_{\alpha} + \mathbf{r}_{mn} - \mathbf{r}_{\beta}|)}{\partial \mathbf{r}_{\alpha}} \frac{\Pi_{\gamma \neq \alpha} H_{\gamma}(|\mathbf{r}_{\alpha} + \mathbf{r}_{mn} - \mathbf{r}_{\gamma}|)}{H_{\beta}(|\mathbf{r}_{\alpha} + \mathbf{r}_{mn} - \mathbf{r}_{\beta}|)} \right) \\ &+ \sum_{\beta \neq \alpha} \sum_m \sum_n w_m w_n \|\nabla H_{\beta}(r_{mn})\| \\ &\times \left(\frac{\partial H_{\alpha}(|\mathbf{r}_{\beta} + \mathbf{r}_{mn} - \mathbf{r}_{\alpha}|)}{\partial \mathbf{r}_{\alpha}} \frac{\Pi_{\gamma \neq \beta} H_{\gamma}(|\mathbf{r}_{\beta} + \mathbf{r}_{mn} - \mathbf{r}_{\gamma}|)}{H_{\alpha}(|\mathbf{r}_{\beta} + \mathbf{r}_{mn} - \mathbf{r}_{\alpha}|)} \right). \end{aligned} \quad (28)$$

As the number of atoms (N_{atom}) in the system increases, the integration of Eqs. (15), (16), and (27) becomes prohibitively costly because the volume function $\nu(\mathbf{r}_{\alpha} + \mathbf{r}_{mn}; \{\mathbf{r}_{\alpha}\})$ is a function of both an integration point of atom α , $\mathbf{r}_{\alpha} + \mathbf{r}_{mn}$, and all atomic positions $\{\mathbf{r}_{\alpha}\}$. In a nutshell, the integration requires $N_{\text{int}} \times N_{\text{atom}}^2$ operations, where N_{int} is the total number of integration points. However, this operation is significantly reduced to $N_{\text{int}} \times N_{\text{atom}} \times N_{\text{list}}$, where N_{list} is the small number of atoms that contribute to the volume function in the vicinity of an integration point of an atom. Such a significant reduction is made possible by the fact that the smoothing length is typically less than 1 Å and that most atoms have $H_{\beta}(\mathbf{r}_{\alpha} + \mathbf{r}_{mn}) = 1$ for $\nu(\mathbf{r}_{\alpha} + \mathbf{r}_{mn}; \{\mathbf{r}_{\alpha}\})$ according to Eq. (11). Thus, it is essential to build a lookup table that quickly establishes which atom contributes to the volume function in the vicinity of an integration point.²¹ For efficiency, the lookup table is built on a grid with a mesh size of d , i.e., each grid point contains a list of atoms. During the integration, each integration point is assigned into the closest grid point, and only the atoms associated with the grid point is used for the integration. To make sure that all atoms necessary for the correct integration at an integration point are listed in the grid point assigned by the integration point, atom α is included in the list of the grid point if the distance between atom α and the grid point is less than R_{cut} where

$$R_{\text{cut}} = R_{\alpha}^{\text{PB}} + w + \frac{\sqrt{3}}{2} d + r_{\text{buffer}}. \quad (29)$$

The third term is necessary to recover atoms which lie on the grid interface. The lookup table is updated during the computer simulations only if any atom in the system moves more than a buffer length r_{buffer} . If necessary, r_{buffer} can be specified to update the lookup table less frequently. Because the choice of d , w , and r_{buffer} affects the calculation speed, they should be optimized for each system before long computer simulations (see Numerical Tests and Discussion).

Because we do not know initially how many atoms, N_{list} , are associated with each grid point, we have to determine N_{list} according to Eq. (29) and allocate the corresponding memory, i.e., $\sum_i N_{\text{grid}}^i N_{\text{list},i}$, where N_{grid} is the total number of grids having a neighbor-

Table 1. Coefficients a_0 and a_1 in Eq. (7) for Different Smoothing Lengths ($2w$).

	w (Å)					
	0.1	0.2	0.3	0.4	0.5	0.6
a_0	-0.081	-0.148	-0.180	-0.168	-0.154	-0.173
a_1	1.600	1.729	1.817	1.856	1.886	1.945

ing atom list and $N_{\text{list},i}$ is the number of atoms listed in the i th grid point. This precalculation is necessary to reduce memory requirements. Once the lookup table is filled with atom lists, the integration of Eqs. (15) and (16) is relatively straightforward. One can now assign the integration point to a particular grid point closest to it, and calculate the volume function at the integration point by evaluating the atomic volume exclusion function according to Eqs. (11) and (13). After the calculation of effective Born radii is done, the evaluations of GB ΔG_{elec} in Eq. (1) as well as $\partial \Delta G_{\text{elec}} / \partial R_{\alpha}^{\text{GB}}$ for forces in Eq. (18) can be performed using the preestablished nonbonded atom list. For completion of force calculations, the evaluation of $\partial r_{\alpha\beta} / \partial \mathbf{r}_{\alpha}$ in Eq. (22) can be done using the lookup table, similar to the integration of Eqs. (15) and (16). Evaluations of the surface Eqs. (27) and (28) are simpler than the electrostatic part. The present development has been implemented into the GBSW module in the CHARMM biomolecular simulation program.³⁸

Parametrization

The performance of the present GB model largely depends on two parameters a_0 and a_1 in Eq. (7). Assuming that a_0 and a_1 are insensitive to conformational changes of proteins, they were optimized to obtain the least error between GB $\Delta G_{\text{elec},\alpha}$ and PB $\Delta G_{\text{elec},\alpha}$ for each atom of protein 1AJJ (PDB code: 37 amino acids). For this procedure, all PB calculations were performed with a grid spacing of 0.21 Å using the PBEQ module^{28,36,37} of CHARMM.³⁸ The modified optimal PB radii for proteins^{32,33} according to Eq. (12) and the charges from the all-atom parameter set PARAM22³⁹ were used for both PB and GB. For GB calculations, 50 radial integration points up to 20 Å and 38 angular integration points were used in the integration of Eqs. (15) and (16) for each atom. The integration points and weights for the radial component are generated by Gaussian–Legendre quadrature,³⁴ and those for the angular component by the Lebedev quadrature,³⁵ as noted above. PB and GB calculations were performed for several different smoothing lengths. Table 1 shows the optimized a_0 and a_1 for each smoothing length.

Numerical Tests and Discussions

To assess the performance of the present GB model, the GB electrostatic solvation energies are first compared with corresponding PB energies for two different sets of proteins. The first set (SET1) is composed of 611 single-chain proteins obtained from the Brookhaven Protein Data Bank (PDB)⁴⁰ based on the criteria

of very low-sequence identity and high resolution. The second set (SET2) is made up of 251 unfolded and folded conformations of protein L.⁴¹ The electrostatic forces calculated from GB are then compared with those from PB for two proteins. The accuracy of the calculated surface using Eq. (27) is also assessed by comparison with the exact surface.

All PB calculations were performed in the same manner as described above. Because the cost of the GB calculation scales linearly as the number of integration points, we empirically changed the number of integration points to reduce this cost without sacrificing the GB performance too much. As shown in Figure 1, optimal for GB calculations were 38 angular integration points and 24 radial integration points up to 20 Å (5 points between 0.5 and 1 Å, and 19 points between 1 and 20 Å). Because the surface calculation requires the integration only within the smoothed boundary region according to Eq. (27), only three radial integration points were used with a fixed smoothing length of 0.2 Å ($w = 0.1$ Å) for each atom, i.e., between $(R_{\alpha}^{\text{PB}} - 0.1)$ and $(R_{\alpha}^{\text{PB}} + 0.1)$ Å. Different smoothing lengths can be used for electrostatic

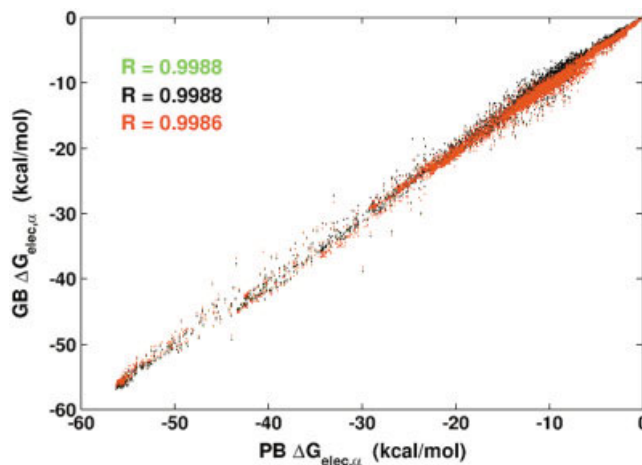


Figure 1. Comparison between PB and GB self-electrostatic solvation free energies $\Delta G_{\text{elec},\alpha}$ for all atoms of 22 relatively small proteins. The number of amino acids in the protein dataset ranges from 37 to 98. GB calculations were performed with 38 angular integration points and 100 (green), 50 (black), and 24 (red) radial integration points. The correlation coefficient R is also given in the figure. It is hard to see the green points because those are all overlapped with black and red points. The figure clearly shows that using 24 radial integration points is fine enough to obtain accurate $\Delta G_{\text{elec},\alpha}$, i.e., effective Born radii.

Table 2. Summary of Error Statistics (in kcal/mol)^a between GB and PB Electrostatic Solvation Energies with Different Smoothing Lengths ($2w$).

	w (Å)					
	0.1	0.2	0.3	0.4	0.5	0.6
SET1: 611 proteins						
ave. error	55.6 ± 51.8	22.8 ± 19.5	21.9 ± 21.5	24.2 ± 23.8	43.5 ± 38.9	52.5 ± 43.4
max error	429.7	118.3	153.0	146.3	323.6	367.0
SET2: 215 conformations of protein L.						
ave. error	39.1 ± 5.7	7.5 ± 4.7	6.7 ± 3.9	12.6 ± 5.1	34.8 ± 6.0	48.0 ± 6.9
max error	54.1	19.4	16.4	26.2	51.9	68.1

^aThe error is defined as $|\Delta G_{PB} - \Delta G_{GB}|$ for each protein or each conformation.

solvation energy and surface calculations because they are not coupled. It was shown before that the surface can be calculated more accurately with smaller w .²⁸ Unless specified explicitly, there is no cutoff for the calculation of ΔG_{elec} in Eq. (1).

Electrostatic Solvation Energies and Forces

The GB performance is mostly assessed by comparing GB ΔG_{elec} with PB ΔG_{elec} . Table 2 provides a summary of error statistics for the first set of 611 proteins (SET1) as the relative energy differences between PB and GB for different smoothing lengths. The electrostatic solvation energies in this set range from -250 to $-13,000$ kcal/mol; the proteins range from 19 to 839 residues. In Figure 2a, all GB ΔG_{elec} are compared with the corresponding PB ΔG_{elec} for the case of $w = 0.3$ Å. Even though the agreement looks perfect, it is not easy to understand detailed statistics from the figure because the range of solvation energies is so large. For example, the same kind of plot with $w = 0.5$ Å also shows perfect agreement with the correlation coefficient of 0.9997 (data not shown), even though Table 2 shows that the error with $w = 0.5$ Å is at least twice larger than that with $w = 0.3$ Å. As shown in Figure 2b, a better representation of the statistical error can be obtained by plotting the number of proteins and the average error percentages as a function of the energy difference between PB and GB. There are 360 proteins (about 55%) within 20 kcal/mol difference between PB and GB, and 575 proteins (about 95%) within 60 kcal/mol difference. In SET1, 20 kcal/mol difference roughly corresponds to 1% error on average and 60 kcal/mol to 2% error. Indeed, the average error percentages are 2.38 ± 1.51 ($w = 0.1$), 1.12 ± 0.86 ($w = 0.2$), 1.07 ± 0.89 ($w = 0.3$), 1.22 ± 1.06 ($w = 0.4$), 2.08 ± 1.47 ($w = 0.5$), and $2.57 \pm 1.60\%$ ($w = 0.6$), respectively (see also Table 2). The agreement between PB and GB is very impressive, considering that a_0 and a_1 were optimized with only one small protein. In the present GB model, it is possible to reproduce the PB energy within 2% with confidence of about 95% if $w = 0.2$, 0.3, or 0.4 are used.

Table 2 shows the relative errors for 251 unfolded and folded conformations of protein L. (SET2) with different smoothing lengths. The PB energies range approximately from -1050 kcal/mol to -1550 kcal/mol. The average error percentages are 3.01 ± 0.45 ($w = 0.1$), 0.58 ± 0.37 ($w = 0.2$), 0.52 ± 0.31 ($w =$

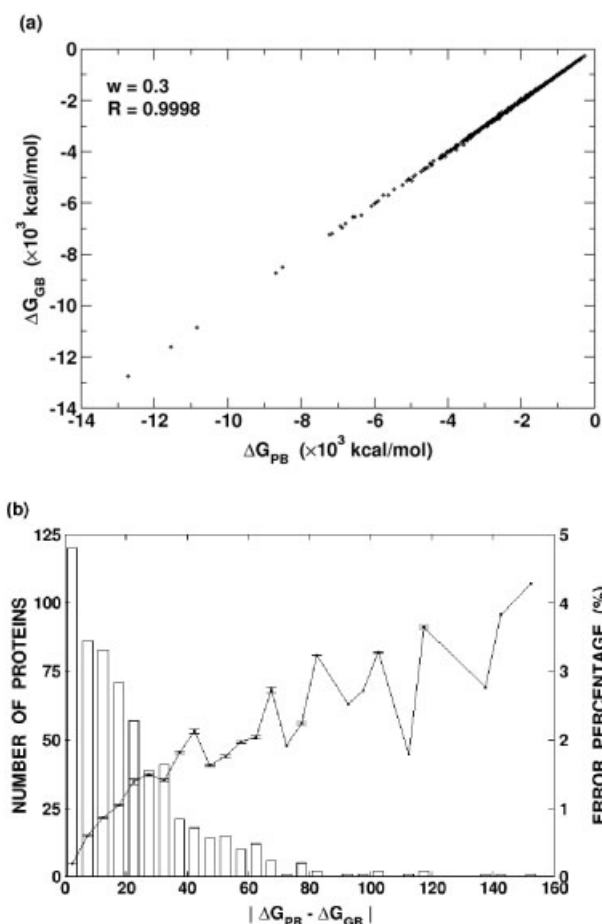


Figure 2. (a) Comparison between ΔG_{PB} and ΔG_{GB} for SET1 with a smoothing length of 0.6 Å ($w = 0.3$ Å). The correlation coefficient R is 0.9998. (b) Number of proteins (bar) and average error percentages (solid line) as a function of error between ΔG_{PB} and ΔG_{GB} for SET1 with $w = 0.3$ Å. A 5.0 kcal/mol bin was used in calculations of number of proteins and average error percentages with its standard deviation. The error percentage is defined as $|\Delta G_{PB} - \Delta G_{GB}|/|\Delta G_{PB}| \times 100$.

0.3), 0.99 ± 0.40 ($w = 0.4$), 2.78 ± 0.48 ($w = 0.5$), and $3.89 \pm 0.53\%$ ($w = 0.6$), respectively. Again, the agreement between PB and GB is excellent particularly for $w = 0.2, 0.3$, or 0.4 .

It is conceivable that the agreement between PB ΔG_{elec} and GB ΔG_{elec} for various proteins and conformations may indicate agreement in electrostatic solvation forces between PB and GB. Nonetheless, it is interesting to see how well GB reproduces PB solvation forces. It is well known that numerically stable forces cannot be obtained in finite-difference PB calculations without smoothness of the dielectric boundary.²⁸ Thus, no attempt has been made so far to compare electrostatic solvation forces between PB and GB using the van der Waals surface or the molecular surface. In this study, however, we are able to compare GB solvation forces directly with PB solvation forces because the same smoothing function is used for both PB and GB. Two proteins are chosen based on the error percentage between PB and GB energies; 1A63 (PDB code; 120 amino acids) with 1.2% error and 1AK6 (PDB code; 174 amino acids) with 4.3% error when $w = 0.3$ Å. In Figure 3a and b, all X-, Y-, and Z-components of the GB solvation forces of heavy atoms are compared with the corresponding PB forces. The average error between PB and GB forces are $0.78 \pm$

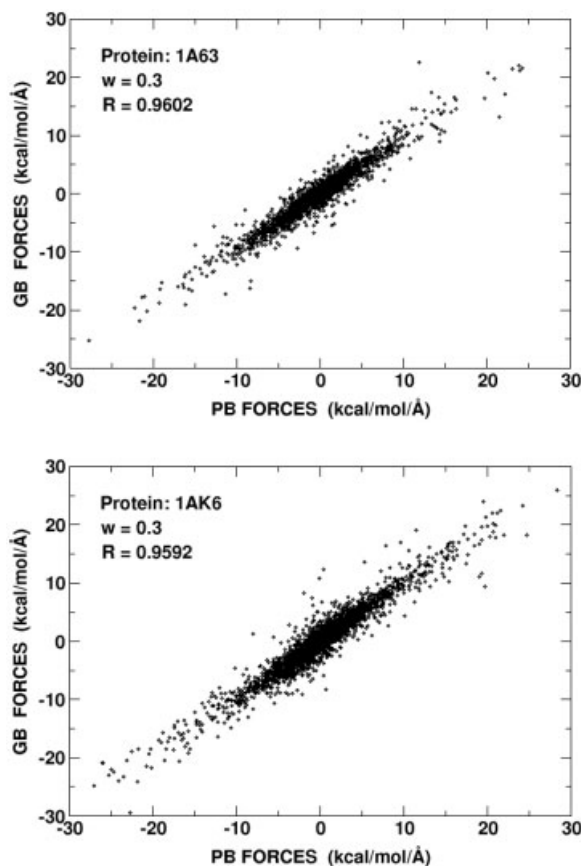


Figure 3. Comparison between GB and PB electrostatic solvation forces for two proteins 1A63 (PDB code: 120 amino acids) and 1AK6 (PDB code: 174 amino acids) with a smoothing length of 0.6 Å ($w = 0.3$ Å).

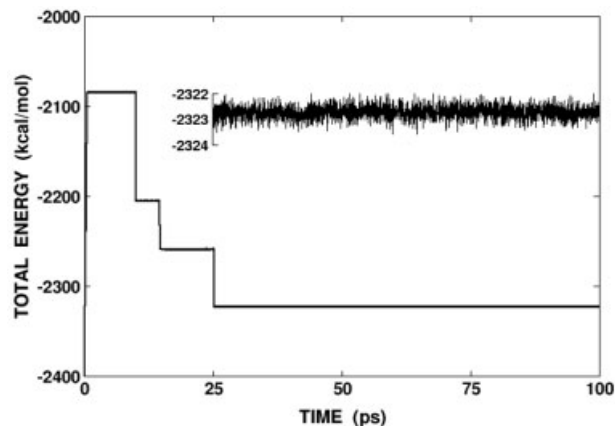


Figure 4. Variation of total energy during a 100-ps MD trajectory of the protein 1A63 at 300 K. In the inset, its variation after about 25 ps is plotted to clearly show the extent of the energy fluctuations. The lookup table as well as the (electrostatic plus nonpolar) solvation forces were updated every step of dynamics (1 femtosecond). For the GB calculations, we used 38 angular integration points and 24 radial integration points up to 20 Å. For the nonpolar solvation energy, the surface tension coefficient γ was set to 0.03 kcal/(mol \cdot Å²).

0.90 kcal/(mol \cdot Å) with a maximum error of 10.62 kcal/(mol \cdot Å) (1A63), and 0.89 ± 1.04 kcal/(mol \cdot Å) with a maximum error of 11.79 kcal/(mol \cdot Å), respectively. The GB forces clearly show reasonably good agreement with the PB forces even though the latter has some oscillation (or uncertainty) due to the discreteness of the grid.²⁸ To illustrate the oscillation in PB forces, the PB forces for 1AK6 were recalculated after adding a quarter of the grid spacing (0.0525 Å) to the coordinates of all atoms. The average error between PB forces before and after translation was 0.13 ± 0.26 kcal/(mol \cdot Å), with a maximum error of 4.14 kcal/(mol \cdot Å). Interestingly, the GB forces of 1AK6 are as reliable as those of 1A63 even though 1AK6 shows higher error than 1A63 in comparison of the solvation energy. It should be stressed that the PB force calculation for 1AK6 took nearly 3 h of CPU time on a 360 MHz SGI R10K with a grid of $417 \times 327 \times 303$, whereas the GB forces took only about 3 s.

Another measure to assess the accuracy of solvation forces calculated from the present continuum model is to monitor the conservation of total energy during MD simulations. We generated 100 ps trajectories of the protein 1A63 at 300 K with the Verlet algorithm after minimization of 200 steps. To prevent abrupt distortions, a positional harmonic restraint with a force constant of 0.1 kcal/(mol \cdot Å²) was applied to all heavy atoms during the first 10 ps of the simulation. During 30 ps of equilibration, the velocities of each atom were reassigned to maintain proper temperature ranges. As shown in Figure 4, the total energy is very well conserved after 25 ps, demonstrating the accuracy of calculated solvation forces.

Surface Area

Because the nonpolar solvation energy of a solute is proportional to its solvent-exposed surface area in the present continuum model,

Table 3. Summary of Error Statistics in Comparison between Exact Surface and Numerically Calculated Surface for 611 Proteins with Different Smoothing Lengths ($2w$).

	w (Å)					
	0.1	0.2	0.3	0.4	0.5	0.6
Ave. error (%)	0.43 ± 0.34	0.45 ± 0.38	0.47 ± 0.39	0.48 ± 0.40	0.51 ± 0.43	0.52 ± 0.43
Max error (%)	2.09	2.63	2.50	2.20	2.43	2.43

it is important to calculate the surface area efficiently and accurately. Previously, Im et al. showed that the surface area can be calculated accurately using Eq. (14) on a grid.²⁸ In this study, we are able to calculate it efficiently and accurately using Eq. (27) by taking advantage of numerical quadrature techniques. Table 3 shows the average error percentages between the exact surface and the numerically calculated surface of the 611 proteins in SET1 with different smoothing lengths. It is noted that the calculated surface area slightly depends on the smoothing length because the PB radii are slightly changed with w according to Eq. (12). The surface area in this set ranges from 1480 to 47550 Å². In Fig. 5a, all numerically calculated surfaces in SET1 are compared with the corresponding exact surfaces in the case of $w = 0.3$. In Figure 5b, all numerically calculated atomic surfaces are compared with the corresponding exact atomic surfaces for individual heavy atoms of the protein 1A63. The error is 1.0 ± 1.1 Å², with a maximum error of 8.2 Å². All the results presented here indicate that the numerically calculated surface area reproduces the exact solvent-exposed surface area within 0.5% error on average.

Computational Timing

In the results presented so far, no cutoff has been used in calculations of GB ΔG_{elec} in Eq. (1), which is indeed necessary to get good agreement between PB and GB. Because the electrostatic solvation energy and the (vacuum) Coulombic interaction energy should be balanced in computer simulations, both must be treated with the same cutoff. However, it is not practical to use no cutoff in computer simulations of biomolecules because the calculations become prohibitively costly as the number of atoms in the system increases. This problem can be circumvented by taking advantage of the well-known fact that the Coulombic interaction energy is strongly anticorrelated with the electrostatic solvation energy, i.e., the total (Coulomb plus solvation) electrostatic energy and forces are converged rapidly as a cutoff value increases.¹⁸ Table 4 shows errors of individual electrostatic energy terms and total electrostatic forces as a function of the cutoff value relative to those with no cutoff for protein 1AK6. Calculations with other proteins show similar trends (data not shown). It is observed that the sum of Coulomb and GB terms, needed in computer simulations, indeed shows rapid convergence, even though both terms themselves are not converged even with a large cutoff of 22 Å. The total energy appears to be converged even at a cutoff of 8 Å, but the energy itself is not a reliable measure to see the trend of convergence because of its fluctuation. The errors in total electrostatic forces turn out to be a more reliable measure of the convergence as a

function of cutoff. The results without a switching function show better convergence than those with a switching function beyond a cutoff of 12 Å. This shows the general trend of convergence of the total electrostatic energy and forces as a function of cutoff, and which cutoff value is good to use will depend on the system being studied.

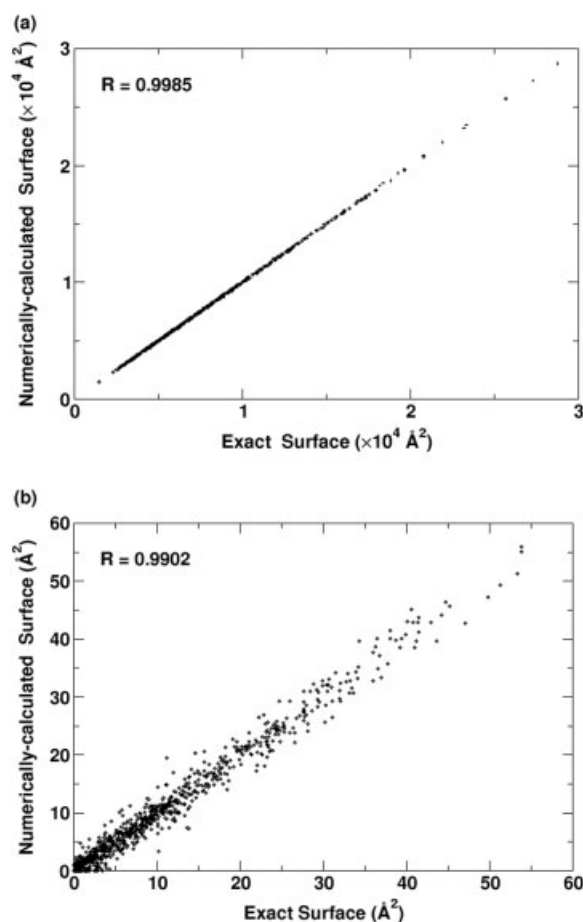


Figure 5. (a) Comparison between the exact surfaces and the numerically calculated surfaces using Eq. (27) for SET1. The correlation coefficient R is 0.9985. (b) Comparison between the exact surfaces and the numerically calculated surfaces for individual heavy atoms of protein 1A63 of 120 amino acids. The correlation coefficient R is 0.9902.

Table 4. Errors in Individual Electrostatic Energy Terms and Total Electrostatic Forces Relative to Those with No Cutoff for Protein 1AK6.

Cutoff (Å)	Coulomb (%)	GB (%)	Coulomb + GB (%)	Forces ^a (kcal/mol/Å)
With a switching function ^b				
8	32.73	7.11	0.41	0.232 ± 0.326 (2.975)
10	21.26	4.86	0.07	0.114 ± 0.162 (1.711)
12	43.33	9.92	0.12	0.070 ± 0.093 (0.877)
14	69.92	16.03	0.18	0.050 ± 0.063 (0.548)
16	38.94	8.93	0.09	0.041 ± 0.058 (0.381)
18	35.69	8.19	0.08	0.035 ± 0.046 (0.480)
20	7.44	1.71	0.17	0.034 ± 0.044 (0.371)
22	23.36	5.36	0.06	0.031 ± 0.039 (0.381)
Without a switching function				
8	34.39	9.18	0.96	0.347 ± 0.457 (4.468)
10	43.28	9.40	0.54	0.162 ± 0.251 (2.582)
12	48.29	11.21	0.01	0.068 ± 0.111 (1.372)
14	25.46	5.80	0.09	0.030 ± 0.051 (0.660)
16	55.89	12.82	0.14	0.013 ± 0.021 (0.292)
18	5.65	1.30	0.01	0.007 ± 0.009 (0.101)
20	14.80	3.40	0.03	0.004 ± 0.005 (0.086)
22	8.11	1.86	0.02	0.003 ± 0.004 (0.033)

^aThe average force error $|\mathbf{F}_{\text{cut}}| - |\mathbf{F}_{\text{nocut}}|$ and the maximum error in parenthesis for all heavy atoms.^bA switching function is applied from $r_{\text{cut}} - 2$ to r_{cut} .

It is of interest to examine the computational time required by the present continuum GB/SA model relative to the vacuum calculation as a function of the system size, different cutoffs, or the grid spacing d for the lookup table. For this purpose, three proteins were chosen based on their size. According to their PDB code, these proteins are 1AJJ (37 residues and 519 atoms), 1AK6 (174 residues and 2738 atoms), and 1YGE (839 residues and 13,265 atoms). The computation time was measured for 1000 steps of molecular dynamics simulation on a 360 MHz SGI R10K. The lookup table as well as the (electrostatic plus nonpolar) solvation forces were updated every step of dynamics. The results are summarized in Table 5. First, the cost of calculating the solvation

terms relative to the vacuum calculation decreases as the cutoff value increases. This result arises from the fact that the cost of the nonbonded interactions increases as the cutoff value increases, whereas that of calculating the effective Born radii remains the same regardless of cutoff. Second, the GB/SA performance is improved as the number of atoms in the system increases, demonstrating its favorable scaling with respect to the system size. Third, the grid spacing used for the lookup table affects the GB/SA performance. In general, a smaller grid spacing than 1.5 Å yields a deterioration of performance because the number of atoms listed in the lookup table increases as the grid spacing is decreased. Overall, it is expected that the present GB/SA model will be well

Table 5. Computation Time (in Minutes) for 1000 Steps of Dynamics.^a

Cutoff (Å)	d^b (Å)	r_{buffer}^c (Å)	1AJJ		1AK6		1YGE	
			Vacuum	GB/SA	Vacuum	GB/SA	Vacuum	GB/SA
12	1.5	0.0	0.44	2.87 (6.5×)	3.8	22.3 (5.8×)	27.7	134 (4.9×)
14	1.5	0.0	0.52	3.07 (5.9×)	5.1	24.7 (4.8×)	36.0	159 (4.4×)
16	1.5	0.0	0.60	3.28 (5.5×)	6.8	28.9 (4.2×)	50.4	194 (3.8×)
16	1.0	0.0	0.60	3.40 (5.7×)	6.8	29.4 (4.3×)	50.4	238 (4.7×)
16	2.0	0.0	0.60	3.29 (5.5×)	6.8	28.9 (4.2×)	50.4	184 (3.7×)
16	1.5	0.5	0.60	3.27 (5.4×)	6.8	29.0 (4.3×)	50.4	196 (3.9×)

^aFor the nonpolar solvation energy, the surface tension coefficient γ was set to 0.03 kcal/(mol · Å²). The value in parenthesis represents a relative time to the corresponding vacuum dynamics.^b d is the grid spacing used for the lookup table in Eq. (29).^c r_{buffer} is the buffer size used for the lookup table in Eq. (29). Setting r_{buffer} to zero means that the lookup table is updated every step of dynamics. The Born radii were updated at every time step.

suited for simulations of large biomolecules because using a large grid spacing reduces the overhead required for atom lists in the lookup table. For example, when the cutoff value is set to 16 Å, the vacuum simulation of protein 1YGE requires 60 Mbytes of memory, whereas the GB/SA simulations require 87 Mbytes with a grid spacing of 2.0 Å and 104 Mbytes with a grid spacing of 1.0 Å.

To examine the effect of having a buffer in Eq. (29), i.e., the effect of the update frequency for the lookup table, additional simulations were performed with a buffer size of 0.5 Å. The results are shown in Table 5. As the system size increases, using a buffer slightly worsens the GB/SA performance because the number of atoms in the lookup table increases as the buffer size increases, according to Eq. (29). This demonstrates that the present scheme for generating the lookup table is very efficient. Detailed analysis with protein 1AK6 showed that calculations of the lookup table, GB ΔG_{elec} and its forces, and ΔG_{np} and its forces took 4.3, 92.9, and 2.8% of the total time for the GB/SA terms, respectively. The calculation of ΔG_{np} (i.e., the surface area) appears to be quite efficient because its calculation also utilizes the preestablished lookup table.

Last, we performed the same MD simulations with the GBMV module (volume integration method for molecular surface representation)^{21,24} in CHARMM.³⁸ A cutoff of 16 Å was used without a switching function. With GBMV the simulations took 12.4 (1AJJ), 90.0 (1AK6), and 556.7 min (1YGE) for 1000 steps of dynamics, respectively, showing that GBMV is roughly three times slower than the present GB model. The speedup of the present GB model relative to GBMV mostly comes from the fact that GBMV uses a longer tail in the volume function, which is necessary to fit the molecular surface definition, and thus requires a significantly larger number of atoms in the lookup table than the present GB.

Concluding Discussion

Based on the recent GB theory of Lee et al.,^{21,24} we reformulated the calculation of the self-electrostatic solvation energy through the use of a simple smoothing function to represent the dielectric boundary, and a volume integration method analogous to that of the earlier study. The present GB model is consistent with a previously developed PB theory, which yields numerically stable electrostatic solvation forces using finite-difference methods.²⁸ Our evaluations of this new GB model showed that it is indeed an efficient and accurate approach to reproduce corresponding PB solvation energies and forces. Having only two adjustable parameters, a_0 for the Coulomb field term and a_1 for the correction term beyond Coulomb field, the PB solvation energies can be reproduced within 1% error on average for a variety of proteins. Detailed analysis shows that it is possible to reproduce the PB energy to within 2% with confidence of about 95% if $w = 0.2$, 0.3, or 0.4 are used. It is also shown that the solvent-exposed surface area of a biomolecule for the nonpolar solvation energy can be calculated very accurately and efficiently using the simple smoothing function.

The relative computational time required for the present GB/SA model depends on the system size, different cutoffs, and the grid spacing for the lookup table. The calculations including

GB/SA are usually 4.5 times slower than the corresponding vacuum calculations. In general, the GB/SA performance (relative to a vacuum calculation) increases as the size of the system or the cutoff value increases due to the scaling behavior of our algorithm. It is expected that this continuum model will be well suited for computer simulations of large biomolecules because using a large grid spacing reduces the overhead required for atom lists in the lookup table.

As expected, using the simple smoothing function makes the GB calculation roughly three times faster than the GB calculation which mimics the molecular surface definition. As discussed, the simple smoothed dielectric boundary introduced here is rather closely related to the van der Waals surface, even though it has some ability to prevent interior voids with high dielectric constant values from being assigned in biomolecules. Which surface representation is the best to use in the implicit solvent model is an open question and is still being debated.⁴² Having the limitations of the model in mind,² one could check this issue carefully in the context of dynamics of a macromolecule with different GB models or by comparing the charging free energies with explicit solvent model with those from different surface representations.^{33,43} Efforts in these directions are currently in progress.

Last, the GB model can be extended to include the influence of a membrane as PB does.²² The volume integration method is naturally suited for having a low-dielectric planar membrane in which the solvent is not accessible. Extension of the present GB model to include the influence of a membrane, and its applications, are currently in progress.

Acknowledgments

Useful discussions with Jianhan Chen, Michael Feig, and Freddie R. Salsbury, Jr., are gratefully acknowledged. We are particularly grateful to Michael Feig for preparation of a set of 611 proteins and for his efforts to develop the MMTSB toolset (<http://mmtsb.scripps.edu>). Financial support from the NIH (through RR12255) and the NSF funded center for Theoretical Biological Physics (CTBP) is appreciated.

References

- Brooks, C. L., III; Karplus, M.; Pettitt, B. M. In *Advances in Chemical Physics*; Prigogine, I., Rice, S. A., Eds.; John Wiley & Sons: New York, 1988, vol. LXXI.
- Roux, B.; Simonson, T. *Biophys Chem* 1999, 78, 1.
- Warwicker, J.; Watson, H. C. *J Mol Biol* 1982, 157, 671.
- Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins* 1986, 1, 47.
- Nicholls, A.; Honig, B. *J Comput Chem* 1991, 12, 435.
- Honig, B.; Nicholls, A. *Science* 1995, 268, 1144.
- Roux, B.; Bernèche, S.; Im, W. *Biochemistry* 2000, 39, 13295.
- Murray, D.; Honig, B. *Mol Cell* 2002, 9, 145.
- David, L.; Luo, R.; Gilson, M. *J Comput Chem* 2000, 21, 295.
- Luo, R.; David, L.; Gilson, M. *J Comput Chem* 2002, 23, 1244.
- Born, M. *Z Phys* 1920, 1, 45.
- Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J Am Chem Soc* 1990, 112, 6127.

13. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J Phys Chem* 1996, 100, 19824.
14. Schaefer, M.; Karplus, M. *J Phys Chem* 1996, 100, 1578.
15. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J Phys Chem A* 1997, 101, 3005.
16. Scarsi, M.; Apostolakis, J.; Caffisch, A. *J Phys Chem A* 1997, 101, 8098.
17. Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J Phys Chem B* 1998, 102, 10983.
18. Dominy, B. N.; Brooks, C. L., III. *J Phys Chem B* 1999, 103, 3765.
19. Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theor Chem Acc* 1999, 101, 426.
20. Onufriev, A.; Bashford, D.; Case, D. A. *J Phys Chem B* 2000, 104, 3712.
21. Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. *J Chem Phys* 2002, 116, 10606.
22. Spassov, V. Z.; Yan, L.; Szalma, S. *J Phys Chem B* 2002, 106, 8726.
23. Onufriev, A.; Bashford, D.; Case, D. A. *J Comput Chem* 2002, 23, 1297.
24. Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. *J Comput Chem*, 2003, 24, 1348.
25. Lee, B.; Richards, F. M. *J Mol Biol* 1971, 55, 379.
26. Richards, F. M. *Annu Rev Biophys Bioeng* 1977, 6, 151.
27. Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J Phys Chem* 1993, 97, 3591.
28. Im, W.; Beglov, D.; Roux, B. *Comput Phys Commun* 1998, 111, 59.
29. Hermann, R. B. *J Phys Chem* 1972, 76, 2754.
30. Simonson, T.; Brunger, A. *J Phys Chem* 1994, 98, 4683.
31. Sharp, K. A.; Honig, B. *Annu Rev Biophys Biophys Chem* 1990, 19, 301.
32. Nina, M.; Im, W.; Roux, B. *Biophys Chem* 1999, 78, 89.
33. Nina, M.; Beglov, D.; Roux, B. *J Phys Chem B* 1997, 101, 5239.
34. Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes: The Art of Scientific Computing*; Cambridge University Press: Cambridge, MA, 1989.
35. Lebedev, V. I.; Laikov, D. N. *Doklady Math* 1999, 59, 477.
36. Roux, B. *Biophys J* 1997, 73, 2980.
37. Im, W.; Bernèche, S.; Roux, B. *J Chem Phys* 2001, 114, 2924.
38. Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J Comput Chem* 1983, 4, 187.
39. MacKerell, A. D., Jr.; Bashford, D.; Bellot, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; S. Ha, D. J.-M.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J Phys Chem B* 1998, 102, 3586.
40. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res* 2000, 28, 235.
41. Wikstrom, M.; Drakenberg, T.; Forsen, S.; Sjobring, U.; Bjorck, L. *Biochemistry* 1994, 33, 14011.
42. Banavali, N. K.; Im, W.; Roux, B. *J Chem Phys* 2002, 117, 7381.
43. Banavali, N. K.; Roux, B. *J Phys Chem B* 2002, 106, 11026.