Annual Review of Biophysics

Generalized Born Implicit Solvent Models for Biomolecules

Alexey V. Onufriev[1] and David A. Case[2]

[1]Departments of Computer Science and Physics, Center for Soft Matter and Biological Physics, Virginia Tech, Blacksburg, Virginia 24060, USA; email: alexey@cs.vt.edu

[2]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, USA; email: david.case@rutgers.edu

Annu. Rev. Biophys. 2019. 48:275–96

First published as a Review in Advance on March 11, 2019

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

https://doi.org/10.1146/annurev-biophys-052118-115325

Keywords

electrostatics, dielectric, generalized Born, implicit solvation, biomolecular simulations

Abstract

It would often be useful in computer simulations to use an implicit description of solvation effects, instead of explicitly representing the individual solvent molecules. Continuum dielectric models often work well in describing the thermodynamic aspects of aqueous solvation and can be very efficient compared to the explicit treatment of the solvent. Here, we review a particular class of so-called fast implicit solvent models, generalized Born (GB) models, which are widely used for molecular dynamics (MD) simulations of proteins and nucleic acids. These approaches model hydration effects and provide solvent-dependent forces with efficiencies comparable to molecular-mechanics calculations on the solute alone; as such, they can be incorporated into MD or other conformational searching strategies in a straightforward manner. The foundations of the GB model are reviewed, followed by examples of newer, emerging models and examples of important applications. We discuss their strengths and weaknesses, both for fidelity to the underlying continuum model and for the ability to replace explicit consideration of solvent molecules in macromolecular simulations.

Contents

1. INTRODUCTION

Biochemical processes often depend strongly on their environment, often consisting of water plus mobile ions, whose main thermodynamic impact is to stabilize charges or polar groups and to screen charge–charge interactions. Computer simulations, based on MD or Monte Carlo sampling techniques, can represent these and other solvation effects by including explicit water molecules and ions as a part of the simulation setup (87). This approach is generally straightforward but can be both expensive and inconvenient for studies where the behavior of the biomolecular system (the solute) is of primary interest. The basic idea of implicit solvation is to integrate out explicit solvent degrees of freedom, incorporating their thermodynamic effects into a solvation free energy ΔG_{solv}. This leads to a simulation setup where the only explicit degrees of freedom are the coordinates of the solute and where the effects of the solvent environment adjust instantaneously as the solute configuration changes.

By its very nature, implicit solvent models describe only a thermodynamic equilibrium state; sampling of the equilibrium states can be greatly enhanced but at the expense of disregarding the kinetics. Of greater concern is a seemingly inevitable loss of accuracy and generality of fast implicit solvation: The complex ways in which collections of water molecules interact with biomolecules are greatly simplified to achieve computational efficiency, to the detriment of physical realism. Furthermore, explicit solvent simulations can be very general: It is relatively straightforward to explore changes in temperature and pressure, to incorporate heterogeneous environments (such as lipid bilayers), to deal with multivalent ions or mixtures of monovalent and divalent ions, and so on.

In contrast, the implicit solvent models discussed here are mostly specific to a fairly homogeneous water/ion environment in a narrow temperature range. Given these limitations, it may be useful to sum up reasons that implicit solvent models are still of wide interest:

1. There is no need for the lengthy equilibration of water that is typically necessary in explicit water simulations; implicit solvent models correspond to instantaneous solvent dielectric response.
2. Continuum simulations generally give improved sampling, owing to the absence of viscosity associated with the explicit water environment; hence, the macromolecule can more quickly explore the available conformational space.
3. There are no artifacts of periodic boundary conditions; the continuum model corresponds to solvation in an infinite volume of solvent.
4. New (and simpler) ways to estimate free energies become feasible; since solvent degrees of freedom are taken into account implicitly, estimating free energies of solvated structures is much more straightforward than with explicit water models (57, 104).
5. Implicit models provide a high degree of algorithm flexibility. For instance, a Monte-Carlo move involving a solvent-exposed side chain would require nontrivial rearrangement of the nearby water molecules if they were treated explicitly. With an implicit solvent model, this complication does not arise. Similar considerations (discussed in Section 4.6) arise when protonation changes are sampled in constant pH simulations.
6. Last but not least, the implicit solvation approach is very useful at guiding physical reasoning (122).

In this review, we focus our discussion on the generalized Born (GB) implicit solvent model, whose computational speed is roughly comparable to that of molecular-mechanics (force field) calculations for biomolecules in the absence of solvent and that provides both solvation energies and the solvent contribution to forces on the solute atoms. The forces enable MD simulations, and the speed makes possible the sorts of extensive explorations of conformational space that are needed for computational design, docking calculations, and free energy estimates. We build upon earlier reviews (8, 9, 82, 87, 120) but try to cover some of the most recent developments while also keeping this account reasonably self-contained.

2. POLAR SOLVATION: THE GENERALIZED BORN MODEL

One of the conceptually simplest models treats the solute as a low-dielectric region ϵ_{in}, embedded in a continuous medium characterized by a macroscopic dielectric constant ϵ_{out}, applying macroscopic concepts at a level of atomic detail (10, 98). There are many reasons to be suspicious of this model, not least of which is the difficulty of assigning a single dielectric constant to the structurally diverse interior of a protein or other macromolecule (97). Still, this represents a well-specified physical model, whose properties can be systematically explored and whose defects might be ameliorated through empirical parameterization or adjustment. Early explorations, before the development of digital computers, treated the solute molecule as a sphere; this model still provides useful insights, and we begin there.

2.1. Electrostatic Interactions in Spherical Geometries

The Born equation (12) describes the transfer free energy of a single spherical ion (with a single charge at its center) from the gas phase to a (water) environment characterized by a continuum

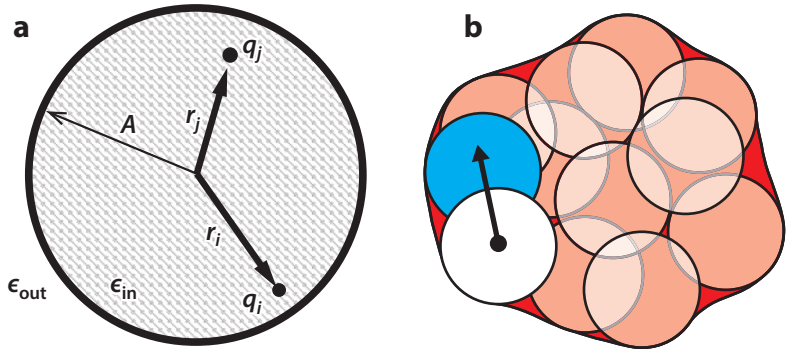

Figure 1

(*a*) A spherical solute of radius A. Two point charges, q_i and q_j, are placed at distances r_i and r_j, respectively, from the sphere center. (*b*) Representation of a molecule as a set of overlapping spheres. The integral needed in Equation 12 can be approximately written as a sum over all of the spheres except for the white one. Pairwise overlap of spheres, as with the white and blue spheres, and missing interstitial volume (*red*) are handled analytically.

dielectric ϵ_{out}:

$$\Delta G_{solv}(R_i) = -\left(1 - \frac{1}{\epsilon_{out}}\right)\frac{q^2}{2A}, \qquad 1.$$

where A is the ion radius, q is its charge, and Gaussian units are used; a derivation is given below in Equation 10. The complex effects of water–ion and water–water interactions across multiple water shells surrounding the ion, and including both the entropic and enthalpic contributions, are distilled into the functional dependence of ΔG_{solv} on the key parameters of the model: in this case, the solvent dielectric and the ion charge and size.

The Born formula (12) (Equation 1) is the simplest of several models (54, 55, 109) that share the same underling physics—they are exact solutions of the Poisson equation of continuum electrostatics for a spherical solute. These so-called spherical cow models still serve to illustrate some of the key features of electrostatic interactions in biomolecules. The Tanford–Kirkwood model, for example, was used for many years to rationalize the pH behavior of proteins (39, 109, 110).

Some simple and interesting limits arise when ϵ_{out} becomes very large (56). In particular, one can show that for an arbitrary charge distribution inside a spherical solute (**Figure 1*a***), the exact solution of the Poisson equation (in the limit $\epsilon_{out} \to \infty$) gives the solvation free energy of the following simple form (38, 100):

$$\Delta G_{solv} = -\frac{1}{2}\left(\frac{1}{\epsilon_{in}} - \frac{1}{\epsilon_{out}}\right)\sum_{i,j}\frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j}}, \qquad 1.$$

$$R_i = A - r_i^2/A. \qquad 2.$$

This result is very similar to the GB idea introduced in the next section; note that the double summation in Equation 2 includes self-energy terms, where $i = j$ (representing stabilization of charges via polarization of the external medium), as well as cross-terms that represent charge–charge interactions screened by the surrounding high-dielectric solvent.

2.2. Treating Molecules as Collections of Atoms

Spherical models are appealing owing to their simplicity and clear physical foundation but have obvious limitations for realistic molecules, which have multiple charge centers and are nonspherical. Attempts to generalize Equation 1 to the multi-atom case are at least 70 years old (42) and began to be applied in earnest to biomolecular problems about three decades ago (9, 105). If we imagine a molecule consisting of charges $q_1 \ldots q_N$ embedded in spheres of radii $a_1 \ldots a_N$, and if the separation r_{ij} between any two spheres is sufficiently large in comparison to the radii, then the solvation free energy can be given by a sum of individual Born terms and pairwise Coulombic terms:

$$\Delta G_{\text{solv}} \simeq \sum_{i}^{N} -\frac{q_i^2}{2a_i}\left(1-\frac{1}{\epsilon_{\text{out}}}\right) + \frac{1}{2}\sum_{i}^{N}\sum_{j\neq i}^{N}\frac{q_i q_j}{r_{ij}}\left(\frac{1}{\epsilon_{\text{out}}}-1\right), \qquad 4.$$

where the factor $(1/\epsilon_{\text{out}} - 1)$ appears in the pairwise terms because the Coulombic interactions are rescaled by the change of dielectric constant upon going from vacuum to solvent. Of course, in real molecules, the atomic spheres are not necessarily far from one another, so one needs to go a step further.

The project of GB theory can be thought of as an effort to find a relatively simple analytical formula, resembling Equation 4, which for realistic molecular geometries will capture as much as possible the physics of the Poisson equation. The linearity of the Poisson equation (or the linearized Poisson–Boltzmann equation) assures that ΔG_{solv} will indeed be quadratic in the source charges, and so it is natural to generalize Equation 4 to

$$\Delta G_{\text{solv}} \simeq \left(1-\frac{1}{\epsilon_{\text{out}}}\right)\frac{1}{2}\sum_{ij}\frac{q_i q_j}{f_{ij}^{\text{GB}}}, \qquad 5.$$

where f^{GB} is some reasonably simple function. Here, the self ($i = j$) f^{GB} terms can be thought of as effective Born radii, whereas for the off-diagonal terms, it becomes an effective interaction distance. Befittingly, the collective name for the resulting models is the generalized Born. A particularly successful version of the GB kernel, f_{ij}^{GB} in Equation 5, was proposed in 1990 (105):

$$f_{ij}^{\text{GB}}(r_{ij}) = \left[r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_iR_j)\right]^{1/2}. \qquad 6.$$

Here, the R_i terms are the effective Born radii of the atoms, which generally depend not only on a_i, the intrinsic radius of atom (i), but on the radii and relative positions of all other atoms. Qualitatively, the effective Born radius of an atom corresponds to its degree of shielding from solvent by the surrounding atoms. In what follows, we call the GB model based on Equations 5 and 6 the canonical GB model.

While historically, the GB model was not conceived as a direct approximation to the Poisson model, and alternative interpretations exist (17), a number of arguments (87) can be made to support this viewpoint, including the very similar structure of the canonical GB and that of Equation 2—the exact solution of the Poisson problem for a sphere. One can further argue (88) that the $\exp(-r_{ij}^2/4R_iR_j)$ factor in Equation 6 attempts to account, in some average sense, for the nonspherical shape of realistic molecules by attenuating charge–charge interactions along the long dimension of the molecule. The view of the GB as an approximation to the Poisson model has been instrumental in the development of the GB for macromolecular applications and continues to help reasoning in developing new GB-like theories.

2.3. The Coulomb Field Approximation

In the classical electrostatics of a linearly polarizable media (49), the work required to assemble a charge distribution can be formulated either in terms of a product of the charge distribution with the electric potential or in terms of the scalar product of the electric field **E** and the electric displacement **D**:

$$G = \frac{1}{2}\int_{\Omega} \rho(\mathbf{x})\psi(\mathbf{x})d\mathbf{x} = \frac{1}{8\pi}\int_{\Omega} \mathbf{E}\cdot\mathbf{D}d\mathbf{x}. \quad 7.$$

We now introduce the essential approximation used in most early forms of GB theory: that the electric displacement is Coulombic in form and remains so even as the exterior dielectric is altered from 1 to ϵ_{out} in the solvation process. In other words, the displacement due to the charge of atom i (which is here presumed to lie on the origin) is

$$\mathbf{D}_i \approx \frac{q_i\mathbf{r}}{r^3}. \quad 8.$$

This is called the Coulomb field approximation (CFA). It is exact for a charge at the center of a sphere, but substantial deviations arise in more complex geometries. The work of placing a charge at the origin within a molecule whose interior dielectric constant is ϵ_{in}, surrounded by a medium of dielectric constant ϵ_{out} and in which no other charges have yet been placed, is then

$$G_i = \frac{1}{8\pi}\int (\mathbf{D}/\epsilon)\cdot\mathbf{D}d\mathbf{x} \approx \frac{1}{8\pi}\int_{\text{in}} \frac{q_i}{r^4\epsilon_{\text{in}}}d\mathbf{x} + \frac{1}{8\pi}\int_{\text{out}} \frac{q_i}{r^4\epsilon_{\text{out}}}d\mathbf{x}. \quad 9.$$

The electrostatic component of the solvation free energy is found by taking the difference as the exterior dielectric is changed from 1.0 to ϵ_{out}:

$$\Delta G_{\text{solv},i} = \frac{1}{8\pi}\left(\frac{1}{\epsilon_{\text{out}}} - 1\right)\int_{\text{out}} \frac{q_i}{r^4}d\mathbf{x}, \quad 10.$$

where the contribution due to the interior region has canceled in the subtraction. Comparing Equation 10 to Equations 4 or 5, we conclude that the effective Born radius should be

$$R_i^{-1} = \frac{1}{4\pi}\int_{\text{out}} \frac{q_i}{r^4}d\mathbf{x}. \quad 11.$$

It is convenient to rewrite this in terms of integration over the interior region, excluding a radius a_i around the origin, as

$$R_i^{-1} = a_i^{-1} - \frac{1}{4\pi}\int_{\text{in},\ r>a_i} \frac{1}{r^4}d\mathbf{x}. \quad 12.$$

Note that in the case of a monatomic ion, where the molecular boundary is simply the sphere of radius a_i, this equation becomes $R_i = a_i$ and the Born formula is recovered exactly. However, in general, the CFA overestimates the effective Born radii; this problem is partially alleviated by some of the approximations used to compute them, as discussed in Section 2.4.

With these approximations, the original two-dielectric Poisson equation, which implies the need for a (numerical) solution of a three-dimensional partial differential equation, has been

replaced by the much simpler task of estimating the three-dimensional integral over a molecular volume in Equation 12. In the next section, we survey some of the approaches used to do this, which often involve further approximations.

2.4. Estimating Effective Born Radii

Equations 11 and 12 provide a formal approach to computing effective Born radii, but in practice a variety of approximations and heuristic models are used, both for the sake of efficiency and to inject some empirical understanding into the resulting values.

2.4.1. The notion of "perfect radii." In principle, R_i could be chosen so that if one were to solve the Poisson equation for a charge q_i at the position of atom i and no other charges and a dielectric boundary determined by the molecular shape, then the self-energy of charge i in its reaction field would be equal to $-(q^2/2R_i)(1 - 1/\epsilon_{\text{out}})$. This self-energy could be computed by solving the Poisson equation numerically, and the resulting values of R_i have been called perfect radii; these are known to give a reasonably good approximation to Poisson theory when used in conjunction with Equations 5 and 6 (86). Obviously, this procedure would have no practical advantage over a direct calculation of ΔG_{solv} using a numerical solution of the Poisson equation, but comparisons to perfect radii are useful in testing more approximate (and more rapid) models. In some applications [e.g., coarse-grained DNA models (93, 94)], the perfect radii are computed once and then kept constant throughout the simulation.

2.4.2. Volume integration via quadrature. Perhaps the most straightforward approach to Equation 12 is to make use of quadrature schemes, often loosely adapted from quantum chemistry calculations. For example, the GBSW (GB simple switching) model (7, 44) uses a spherical quadrature model to sample the atomic density surrounding each atom to determine its contribution to the effective Born radius. A switching function is used to blur the sharp boundary between high- and low-dielectric regions, which fills in many of the small voids shown as red regions in **Figure 1**. (A correction to the CFA, discussed in Section 2.5, is also applied.) The use of quadratures results in a method that is not fully rotationally invariant; however, the errors can be small, and a version optimized for GPUs provides performance on par with the nonquadrature models, depending on system size (7).

The GBMV (GB molecular volume) models work harder to evaluate Equation 12 over the true molecular volume, generally trading speed for accuracy (62). The use of an analytical molecular volume model (35), with some heuristics to speed up table lookups, allows Born radii to be obtained that are quite close to the exact radii.

The Gaussian generalized Born model (36, 37) mimics GBSW in abandoning the model of a sharp division between high- and low-dielectric regions and estimates the integrals needed to obtain Born radii by adopting a Gaussian shape model. This approach has the advantage of providing analytical derivatives without the need for numerical grids.

2.4.3. Volume integrations over collections of overlapping spheres. A separate path to Equation 12 treats molecules as collections of overlapping spheres with sharp boundaries, as illustrated in **Figure 1**. If the molecule consisted of a set of nonoverlapping spheres of radius a_j at positions $\mathbf{r}_{ij}$ relative to atom i, then Equation 12 could be written as a sum of integrals over

spherical volumes, which can be computed analytically (95):

$$R_i^{-1} = a_i^{-1} - \sum_j \frac{a_j}{2(r_{ij}^2 - a_j^2)} - \frac{1}{4r_{ij}} \log \frac{r_{ij} - a_j}{r_{ij} + a_j}. \quad 13.$$

An analytical expression is also available for the case of atom j overlapping with the central atom, i, provided j does not overlap any other atom j' (95). Although in practice, the atoms j do overlap with one another to some extent, these overlaps can be neglected to a first approximation, and empirical corrections can be introduced to compensate for the lack of overlap. This is referred to as the pairwise descreening approximation (40). Hawkins et al. (40, 41) have introduced such a scheme based on rescaling the van der Waals radii by factors S_j. The expression for the GB radii takes the form

$$R_i^{-1} = a_i^{-1} - \sum_j H(r_{ij}, S_j a_j), \quad 14.$$

where H is a rather complex expression that, apart from rescaling, is essentially Equation 13 if i and j do not overlap and has different functional forms in overlapping cases (41). In the analytical generalized Born plus nonpolar (AGBNP) model of Gallicchio & Levy (28), the scaling factors S_{ij} are not independent constants but are computed on the fly from a Gaussian-based decomposition of the molecular volume into atomic contributions.

A serious problem with representing molecular volume as a set of overlapping atomic spheres is the neglect of interstitial spaces between the atomic spheres in the interior of the molecule (**Figure 1**). In the approximation, these crevices are treated as if they belonged to the solvent space—that is, filled with high dielectric, which is unphysical for many types of calculations (83); the corresponding effective radii are underestimated. (The use of the CFA leads to a certain cancellation of errors in this case since the CFA tends to overestimate the effective radii.) For biopolymers, this neglect of interstitial space leads to appreciable underestimation of the effective radii, compared to the perfect radii introduced above (86).

Efforts to correct this deficiency while preserving computational efficiency of the pairwise approximation have led to a series of GB flavors. In one of them, GB^{OBC} (now available in many modeling packages), an empirical correction is introduced (84, 85) that modifies the pairwise integration method to reduce the effect of interstitial high dielectrics. However, by design, the GB^{OBC} approach compensates for missing interstitial volume only on average, in a geometry-independent manner. To further improve the GB accuracy, an additional correction to the pairwise procedure was introduced (74) that brings in elements of molecular volume, in a pairwise sense: An additional term is added to Equation 14 that reintroduces the molecular volume between each pair of atoms missed by the original approximation. The integral over this neck-shaped region can be approximated by a simple analytical function with a negligible additional computational expense relative to GB^{OBC}. It took a significant reparameterization effort to make the original GB-neck model a success for both proteins (81) and the DNA (80)—all of these models are now available in Amber.

2.5. Beyond the Coulomb Field Approximation

As mentioned above, the CFA can lead to significant errors in the effective Born radii. In fact, even for a perfectly spherical solute, the CFA is exact only for a charge located in the center of the sphere, while for a charge near the boundary, it overestimates the effective radius by a factor of

2 (9). Consequently, the CFA overestimates effective radii for realistic molecular geometries as well (62). This problem has been well known for quite some time, and various empirical corrections to the CFA have been proposed (44, 60, 62). These typically take the form of a simple linear or a rational combination of correction terms such as

$$\alpha_N = \left(\frac{1}{4\pi}(N-3) \int_{ex} \frac{dV}{|\mathbf{r}-\mathbf{r}_i|^N} \right)^{\frac{1}{N-3}}, \qquad 15.$$

where $N \geq 4$. Specifically, Lee et al. (62) were the first to propose an alternative to the CFA along these lines—namely, an expression involving α_4 and α_5—and later, an even more accurate expression based on α_4 and α_7 (60). These corrections are utilized in the GBSW and GBMV models discussed above.

2.6. Expressions Based on Exact Spherical Limits

The canonical GB model becomes exact for a perfect sphere in the conductor limit $\epsilon_{out} \to \infty$, assuming exact effective Born radii. However, the effective radii computed via the CFA, and via other, more complex integral forms mentioned above, are not exact even for a sphere, leading to hard-to-control inaccuracies for realistic shapes.

An alternative expression to compute the effective R6 radii was proposed by A. Svrcek-Seiler (personal communication) and independently by Grycuk (38) as

$$R_i^{-1} = \left(\frac{3}{4\pi} \int_{out} \frac{d\mathbf{V}}{|\mathbf{r}-\mathbf{r}_i|^6} \right)^{1/3} = \left(a_i^{-3} - \frac{3}{4\pi} \int_{r>a_i}^{solute} |\mathbf{r}|^{-6} d\mathbf{V} \right)^{1/3}, \qquad 16.$$

where in the first expression, the integral (out) is taken over the region outside the molecule and in the second integral, the origin is moved to the center of atom i. The above expression is an integral equivalent of Equation 3—that is, it gives the effective Born radii that are exact for any charge location within a perfect spherical solute in the conductor limit.

The potential advantage of Equation 16 over the CFA and its extensions for practical computation became clear when it was shown (75) that the corresponding R6 radii can be very close to the perfect radii for realistic biomolecular shapes, resulting in solvation energies in close agreement with perfect radii. It was also demonstrated (100) that the use of $\epsilon_{out} \to \infty$ limit in computing the effective Born radii results in more accurate estimates of ΔG_{solv} via canonical GB, at least for single globular molecules. Implementations of the R6 GB model in which Equation 16 is integrated analytically over the van der Waals volume of the solute are available (58, 121). Performing the integral analytically over the molecular volume, which has traditionally been the target for GB models intended for use in MD simulations, proved difficult, and a number of approximations had to be made to produce a fully analytical expression (2); pilot implementations in MD produced mixed results. Compared to the CFA, the R6 model is apparently less forgiving to approximations to the molecular volume, and there is no fortuitous cancellation of error that helps the CFA.

Surface-based R6 formulations have also been developed, in which the effective Born radii are calculated via

$$R_i^{-3} = \left(-\frac{1}{4\pi} \oint_{\partial V} \frac{\mathbf{r}-\mathbf{r_i}}{|\mathbf{r}-\mathbf{r_i}|^6} \cdot d\mathbf{S} \right), \qquad 17.$$

where ∂V represents the molecular surface of the molecule, d**S** is the infinitesimal surface element vector, r_i is the position of atom i, and r represents the position of the infinitesimal surface element. [This follows the lead of earlier work to turn volume integrals into surface ones (32).] A version of the model, GBNSR6 (1, 27) (numerical surface R6 GB), is available in Amber. It has been tested in calculation of small molecule hydration energies (1) and in protein ligand binding (45), where its accuracy is noteworthy (47).

2.7. Alternative Forms for Charge–Charge Interactions

The canonical form of the GB kernel, f_{ij}^{GB} in Equation 6, is not the only one proposed and tested. For example, values of the empirical factor other than 4 in $\exp(-r_{ij}^2/4R_iR_j)$ have been considered (37, 38, 51, 60, 69), in the range from 1 to 8. That the canonical value of 4 remains the most widely used indicates that it is close to a so-called general purpose optimum and that further accuracy improvements may require a different functional from of the GB equation.

A substantially different form of the GB kernel f_{ij}^{GB} was recently proposed (59) on the basis of a carefully examined connection between the GB and conductor-like polarizable continuum models (114):

$$f_{ij} = r_{ij} + \left(1 + \frac{1.028 r_{ij}}{16\sqrt{R_iR_j}}\right)^{-16}. \qquad 18.$$

Noticeable improvements over the canonical GB in both accuracy and speed were reported (59); however, the testing has so far been limited.

GB-like models also exist that go beyond the canonical GB Green function itself (i.e., Equation 5). One such model is the analytical linearized Poisson–Boltzmann (ALPB) (99, 100):

$$\Delta G_{el} = -\frac{1}{2}\left(\frac{1}{\epsilon_{\mathrm{in}}} - \frac{1}{\epsilon_{\mathrm{out}}}\right)\frac{1}{1+\beta\alpha}\sum_{ij} q_i q_j, \qquad 19.$$

where $\beta = \epsilon_{\mathrm{in}}/\epsilon_{\mathrm{out}}$, $\alpha = 0.571412$, and A is the electrostatic size of the molecule, which is essentially the overall size of the structure. In practice, the value of A can be estimated analytically (99). Moreover, it was shown that keeping the value of A constant in MD simulation is acceptable (99). In the limit $\frac{\epsilon_{\mathrm{in}}}{\epsilon_{\mathrm{out}}} \to 0$, the canonical GB and ALPB coincide, but outside of this limit, the latter model shows a closer agreement with the Poisson theory. For the case of aqueous solvation, the model provides a small but consistent improvement over the canonical GB.

Going beyond the spherical shape as the basis for derivation of beyond-GB models led to a model that accounted for the existence of two modes in the solution of the Poisson equation for nonspherical shapes: longitudinal and transverse (88). The former more or less corresponds to the canonical GB, while the latter is very different; the overall Green function interpolates between the two modes on the basis of the values of the local gradients of the effective Born radii. Compared to the canonical GB, the new model resulted in significantly fewer gross errors in pairwise charge–charge interactions, with the numerical Poisson solution taken as reference. However, the testing was so far limited to R6 effective Born radii estimated numerically.

GB-like theories can also explicitly incorporate effects that go even beyond the linear-response Poisson theory. For example, an extension of the Born model was proposed (77) that explicitly accounts for charge hydration asymmetry (CHA)—strong dependence of the hydration free energy

on the sign of the solute charge:

$$\Delta G \simeq -\left(1-\frac{1}{\epsilon}\right)\frac{q^2}{2(R+R_s)}\left(1-\mathrm{sgn}[q]\frac{\delta}{R+R_w}\right). \quad 20.$$

Here, ϵ is the dielectric constant of water, R_w is the radius of water molecule, q and R are the ion charge and ionic radius, respectively, $R_s = 0.52$ Å is a constant shift to the dielectric boundary (77), and δ is the symmetry breaking parameter (78). The charge-asymmetric Born equation (Equation 20) contains no fitting parameters yet describes experimental ion hydration energies to within 5% of experiment (77). The CHA effects can be introduced into the effective Born radii by analogy with Equation 20, leading to a charge-asymmetric GB-like model CHA-GB (76), implemented in AmberTools. The introduction of CHA into the GB has improved its ability to predict hydration free energies of small molecules and amino acids simultaneously, including the charged ones.

2.8. Effects of Ionic Screening

Mobile ions in the solvent can be very effective at screening charge–charge interactions, augmenting the dielectric effect itself. The linearized (or Debye–Hückel) model can be solved analytically for a sphere (54, 109), and an extension can be made to general shapes in the same way as the salt-free case discussed above. This leads to the simple ansatz

$$\left(1-\frac{1}{\epsilon}\right) \rightarrow \left(1-\frac{\exp(-\kappa f_{\mathrm{GB}})}{\epsilon}\right), \quad 21.$$

where κ is the Debye–Hückel inverse screening length.

Both the canonical GB (and the Poisson–Boltzmann equation for that matter) utilize a mean-field description of ions, which does not account for ion–ion correlations or discreteness of ions near the charged solute surface. While in the case of monovalent ions, the correlation effects are small and can often be neglected, the correlations between multivalent ions can introduce significant corrections to ion distributions and electrostatic potentials around solutes. Recently, a GB-like model was constructed (113) to handle ions explicitly. Modifications to the canonical model were required, including modifications to account for multiple interacting solutes, which are characterized by a disconnected dielectric boundary around the solute–ion or ion–ion pairs. For a duplex DNA example, the monovalent (Na^+) and trivalent ($CoHex^{3+}$) counterion distributions produced by the model are in close agreement with all-atom explicit water MD simulations used as reference.

2.9. Adopting Generalized Born for Membrane Environments

A specific challenge arises if one wants to use the GB to describe the effects of the essentially heterogeneous dielectric environment of biological membranes and water/membrane interface. Several empirical modifications (24, 102, 112, 116) to the canonical GB have been proposed that so far utilize the same general idea: keep the main GB formalism intact and use the effective Born radii to account for the presence of additional dielectric boundaries. For example, the heterogeneous dielectric generalized Born (HDGB) flavor is an extension of the GBMV approach, where the original expression for the effective Born radii now includes an explicit dependence on ϵ_{in} and ϵ_{out} via an analytical formula for R_i: $R_i = R_i(\epsilon_{\mathrm{in}}, \epsilon_{\mathrm{out}})$. The model (112) partitions the membrane

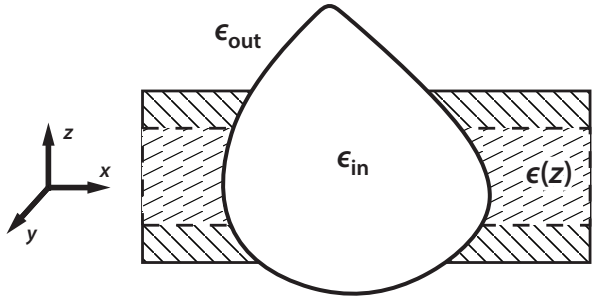

Figure 2

A schematic of an idealized, multi-dielectric membrane environment mimicked by some of the generalized Born models.

slab into several regions of constant dielectric (**Figure 2**), approximating a realistic scenario in which the dielectric properties of the membrane $\epsilon(z)$ vary continuously across the bilayer.

A recent extension of the HDGB model, HDGBvdW (HDGB with a van der Waals term) (21), provides a more accurate description of the nonpolar components of the free energy of solvation. Compared to the original HDGB, the extension improves free energy estimates in the hydrophobic interior of the membrane, where nonpolar interactions are significant.

2.10. Speeding It Up for Large Systems

Although the GB equations are more complex than simple molecular-mechanics force fields, they often perform well in parallel CPU architectures, and adaptations to GPUs are available in several popular packages (7, 22, 33). However, until recently, available GB implementations were poorly suited to handle very large structures, because traditional implementations scale as $O(N^2)$ with the number of solute atoms N. A certain amount of speedup, up to a factor of 3 or so for a 25,000-atom system, can be achieved by using a soft cutoff in the calculation of the effective Born radii (82), but the Ewald-based procedures that address the $O(N^2)$ problem for explicit solvent simulations cannot be readily adapted to the nonperiodic geometries used in implicit solvent models. However, cutoffs applied to charge–charge interaction can lead to artifacts, especially for highly charged systems (46).

To overcome the problem, a method of hierarchical charge partitioning (HCP) was proposed (3, 4, 6), which bears some similarity with the fast multipole approach but is better suited for biomolecular simulations. Specifically, HCP is a multi-scale, yet fully atomistic, approach to perform MD simulations based on the GB model, mainly intended for very large structures. HCP employs a charge coarse-graining scheme that takes advantage of the natural hierarchical partitioning of large biomolecules into smaller structural components (e.g., amino acids, chains). The charge coarse graining is performed via a rigorous mathematical procedure, and no reparameterization of the original atomistic force field is needed. MD simulations based on HCP [GB-HCPO (46), available in Amber] scale as $N \log N$ with the number of atoms in the solute, which means that conformations of very large (millions of atoms) fully atomistic systems can be sampled efficiently, without cut-offs.

3. APPROACHES TO NONPOLAR SOLVATION

The electrostatic component of solvation, including stabilization of polar and charged groups, and the screening of charge–charge interactions, are generally the most influential terms for

biomolecules in (salt-) water. The electrostatic component corresponds to the free energy of discharging the solute (i.e., of setting all its charges to zero). What remains is a very nonpolar object (something like an alkane) that has the same shape as the original solute. This remaining (hypothetical) object has a nonzero solvation free energy, which is generally smaller than the electrostatic component in absolute terms, but which can have a significant influence on conformational equilibria.

These interactions have proved to be hard to model in a simple fashion, primarily because one must consider two competing effects: There is a (unfavorable) free energy change required to create an empty cavity in the solute to accommodate the solute, and this is partially offset by a (favorable) dispersion interaction between the solute (once it is inserted into the cavity) and the surrounding solvent molecules. The simplest, and still most widely used, approach is to assume that the nonpolar contribution is proportional to the solvent-exposed surface area, even though nonpolar contributions are known to depend upon the solute size and shape in a more complex fashion (14, 30, 31, 48, 61, 65). To the extent that they "work," simple surface area terms may capture changes in solvation over narrow ranges of conformations, and experience suggests that the relatively smooth changes associated with solvent-accessible surface areas can avoid pitfalls that arise with more complex models (108).

Even within this model, one has the additional challenge of estimating surface areas in an efficient fashion. The Amber programs use the linear combination of pairwise overlaps model (119), which uses parameters characteristic of common combinations of neighboring atoms to a molecule surface area. Gallicchio & Levy (28) have adapted an earlier model based on a superposition of Gaussian functions representing atomic volumes (34) to compute a surface area estimate (and its derivatives) in an efficient manner. In this AGBNP model, they also base the solute–solvent dispersion estimates on already-computed Born radii.

4. APPLICATION EXAMPLES

Implicit solvent models have been in use in biomolecular simulations for a long time, and it is not feasible to attempt any fair overview of applications. Instead, we list here a few examples, chosen from those we are familiar with, to illustrate a range of applications where particular aspects of implicit solvation are important.

4.1. The Molecular Mechanics–Generalized Born Surface Area Approach to Free Energies

The free energy of solvation estimates provided by implicit solvent models include both enthalpy and entropy contributions, and this opens up a novel end-point approach to free energy calculations (103) where the free energy of a given state (A) is written as

$$G(A) = \langle E_{\mathrm{MM}} \rangle (A) - TS_{\mathrm{config}}(A) + \langle \Delta G_{\mathrm{solv}} \rangle. \qquad 22.$$

Here, E_{MM} is the molecular-mechanics estimate of the average energy of the solute in the absence of the solvent, S_{config} is a configurational entropy arising from the solute degrees of freedom, and ΔG_{solv} gives the free energy contributions from the solvent. The average energy and ΔG_{solv} values are generally computed over the configurations sampled by an MD trajectory, since no single configuration can represent thermal motion in a given basin. The configurational entropy of the solute can be estimated by normal mode or quasiharmonic analysis for fairly rigid molecules (52, 70) and by a variety of approaches for floppier systems (26). Free energy differences for

conformational transitions (say, between bound and free states in ligand–receptor interactions) are made simply by subtracting the free energy estimates of the two end states.

This approach (in its modern incarnation for biomolecules) was first applied to study the A to B helix transition in DNA and RNA (103) and was rather quickly extended to study ligand binding events (57). It has since been widely applied, with varying levels of agreement with observed data (29, 118). One obvious limitation lies in the inability of available implicit solvent models to capture subtle changes in ΔG_{solv} as a function of conformation. But obtaining good estimates of configurational entropies can be equally troubling, especially as the extent of disorder increases; in fact, the reason that this end-point approach is nearly useless for explicit solvent models arises from the difficulty of extracting entropy estimates for the solvent in explicit MD simulations.

4.2. Second Derivatives and Normal Modes

Many of the GB flavors discussed above are fully analytical, and higher-order derivatives can be computed by straightforward (if not tedious) applications of the chain rule. This has been implemented for the original Hawkins–Cramer–Truhlar model (40) in the Amber suite of programs (13). One can then compute normal modes (as with isolated molecules) but where solvent electrostatic effects are incorporated. This can be a very useful model, for example, in studying mechanical properties of biopolymers, such as the stretching and bending rigidity of DNA, with varying degrees of salt (11). Normal mode analysis also provides estimates of configurational entropies, which are useful for end-point analyses like the molecular mechanics–generalized Born surface area approach.

4.3. Large Scale Motions, Very Large Structures

One of the key advantages of the implicit solvation approach is that the effective viscosity of water can be set to a much lower value than that of real water; this can speed up conformational transitions significantly (5). The speedup is particularly impressive for transitions that involve large parts of the biomolecule moving essentially unimpeded in the solvent, in which case 100-fold speedup of conformational sampling compared to the explicit solvent is easily within reach (5), in addition to any algorithmic speedups that the GB can offer for the system in question. The first atomistic simulation of a long (147–base pair) DNA fragment free in solution is an example that illustrates the point: Only several nanoseconds of the simulation (91) had revealed significant, unexpected flexibility of the double helix. Note that even setting up a traditional explicit solvent simulation box of the appropriate size would be quite cumbersome, as the contour length of the 147–base pair DNA is ~500 Å.

Atomistic simulation of large biomolecular systems, especially those with flexible parts, is another area where GB-based simulations can be useful. The nucleosome—a complex of 8 histone proteins and 147 base pairs of DNA wrapped around it—is a relevant example here. GB-based MD simulations (23, 67) were utilized to study highly flexible N-terminal tail regions of the histone proteins, implicated in chromatin remodeling. Recently, partially assembled intermediate states of the nucleosome, unavailable from experiment, were constructed (92) through GB-based simulations followed by refinement in the explicit solvent.

Multimillion atom systems require special multi-scale treatment. One such GB-based model [GB-HCPO (46), described above] was used to refine the atomistic structure of a chromatin fiber fragment consisting of 40 nucleosomes, starting from a low-resolution cryo–electron microscopy input (89) (**Figure 3**).

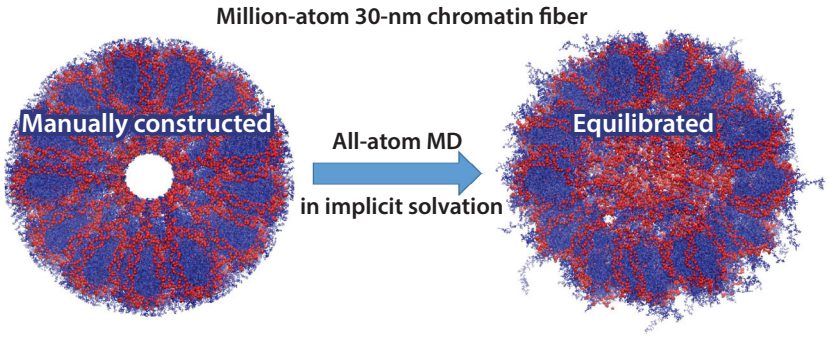

Figure 3

Atomistic simulation (46) of a cryo–electron microscopy–consistent model of 30-nm chromatin fiber reveals important details consistent with experiment: The linker DNA fills the core region, and the H3 histone tails interact with the linker DNA. Abbreviation: MD, molecular dynamics.

4.4. Protein Folding

Protein folding at the atomic resolution—once called "the grand challenge of computational science"—is arguably one of the most illustrious success stories of the GB. Whereas intermediate states resembling the native state were observed in a pioneering simulation of the folding process in explicit solvent (18), it was not until later that a complete folding of a small protein from a fully unfolded state to the native was achieved in a GB-based MD simulation (101). Many other GB-based folding studies followed (15, 50, 64), including those aimed at protein design (25, 68). Recent comparisons of the Amber GB model to Rosetta scoring functions have shown GB results that match or exceed those from Rosetta in terms of protein loop modeling or folding landscape characterization (90).

In a recent landmark study (79), folding simulations of 17 proteins (**Figure 4**) were performed on a commodity personal computer within days (79). In implicit solvent, correct native states of small proteins were identified as minimum energy snapshots (5, 101) in straightforward simulations starting from completely extended conformations—the task that is not nearly as straightforward in explicit solvent. For example, a recent study (66) of folding–unfolding transitions of 12 of the fastest folding proteins required extremely long simulations on a one-of-a-kind specialized supercomputer and the use of significantly elevated temperature to overcome kinetic traps. It is unclear for how many of the same proteins a truly de novo (e.g., from a linear peptide) prediction of the native structure could have been made at 300 K.

4.5. Biomolecules in the Presence of Biological Membranes

Translocation of molecular structures through membranes may involve significant molecular movements and conformational changes; membranes are generally large structures. These qualities make membrane systems good candidates for implicit solvent simulations, such as simulation of whole membrane proteins (103, 111, 112, 116). Recent applications include predicting the hydrophobic length of membrane proteins (19) and protein structure refinement (20).

4.6. Acid–Base and Redox Transitions

One application area that illustrates the potential power of implicit solvent models is constant pH simulations. In an early (but still popular) approach (72, 73, 106, 107), conventional MD

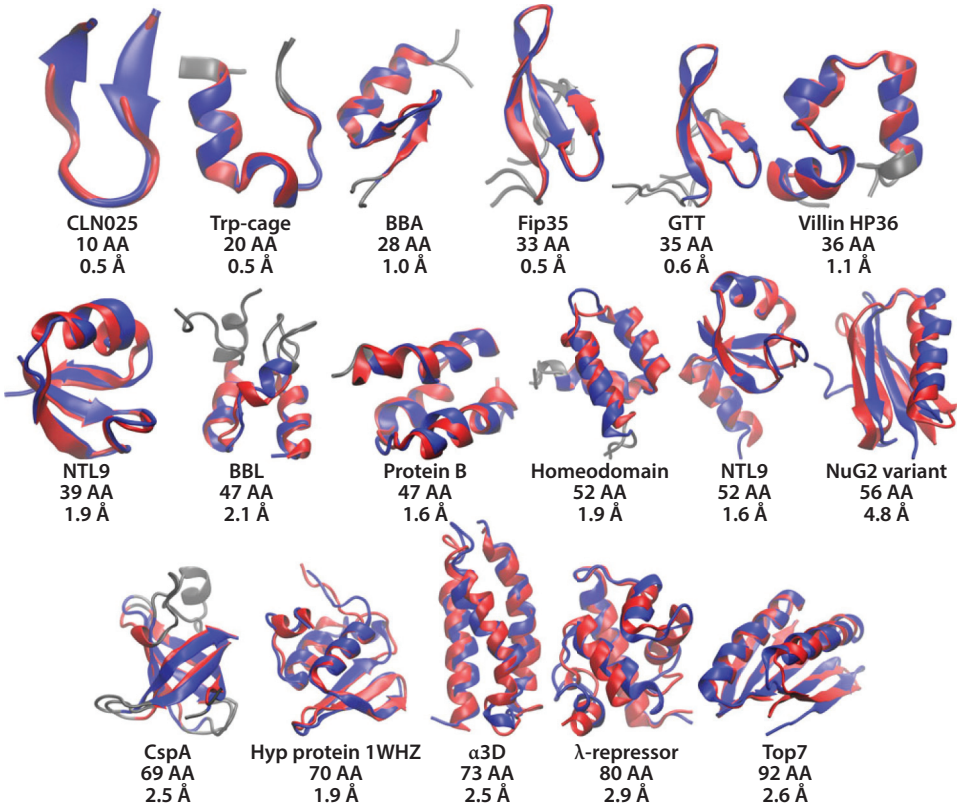

Figure 4

Generalized Born (GB)-based molecular dynamics simulations [GB-neck2 (81), Amber] of a number of proteins starting from completely unfolded states sample conformations (*blue*) that are close to the experimental native structures (*red*). Lowest root-mean-square deviation distance to the experimental native structure, in angstroms, is indicated under each protein. Figure adapted from Reference 79, courtesy of Carlos Simmerling. Abbreviation: AA, amino acids.

simulations are interrupted at intervals by Monte Carlo attempts to transfer protons between the solute and a hypothetical reservoir maintained at a given pH. An attempted move like this would almost always fail in an explicit solvent simulation, since waters arranged around a charged titration site (e.g., a protein side chain) would be in a high-energy configuration for a site that suddenly neutralized, and vice versa. In an implicit solvent model, by contrast, the solvent model can instantaneously respond to a change in charge, and Monte Carlo move attempts can have reasonable acceptance rates. A similar use of this instantaneous response is active in models that treat the transition between charged and neutral sites in terms of a continuous auxiliary variable (53, 63, 117).

A key problem with the implicit solvent models is that results are often in poorer agreement with experiment than one would like. Such errors might arise from limitations of solute force fields or from incomplete sampling, as well as from deficiencies in the solvent model itself. There have been quite a few efforts to extend the constant pH idea to explicit solvent simulations (including approaches that use a hybrid explicit/implicit model), and this continues to be an active area of current research (43).

5. CONCLUSIONS

We have reviewed the overall physical foundations and specific implementations of one of the most widely used fast implicit solvation models—the GB approximation. Implementations are available in a variety of molecular modeling packages, and many thousands of applications have been reported.

The specific choice of GB flavor depends on one's needs: A protein folding simulation may call for one flavor of the GB model (79, 85), while a study of a protein in a membrane environment will need quite another (20, 71). The model one may prefer for estimates of ligand binding energies (47) can be different from GB flavors one may recommend for simulations of the DNA (16, 80, 115). Yet a different approach is needed if the molecular charge distribution is described by (polarizable) multipoles, rather than by fixed point charges (96). Some recent updates in parameters have led to significant improvements in the quality of GB models applied to proteins and their complexes with nucleic acids (80, 81). This diversity of approaches reflects a lack of generality: No single implicit model works well everywhere. This is hardly surprising, given the complexity of solvent effects that one is trying to fold into a simple and fast model. Furthermore, some of the success of GB models certainly arises from fortuitous cancellation of errors and from parameterization schemes that hide defects in the solvation model or in the underlying solute force field. In spite of these limitations, we have described a number of advantages of such schemes and have documented some of the recent and novel ideas that continue to drive research in this area. We expect that there will continue to be an important place for implicit solvent models in biomolecular simulation for some time to come.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by NIH grant GM122086.

LITERATURE CITED

1. Aguilar B, Onufriev AV. 2012. Efficient computation of the total solvation energy of small molecules via the R6 generalized Born model. *J. Chem. Theory Comput.* 8:2404–11
2. Aguilar B, Shadrach R, Onufriev AV. 2010. Reducing the secondary structure bias in the generalized Born model via R6 effective radii. *J. Chem. Theory Comput.* 6:3613–30
3. Anandakrishnan R, Baker C, Izadi S, Onufriev AV. 2013. Point charges optimally placed to represent the multipole expansion of charge distributions. *PLOS ONE* 8:e67715
4. Anandakrishnan R, Daga M, Onufriev AV. 2011. An *n* log *n* generalized Born approximation. *J. Chem. Theory Comput.* 7:544–59
5. Anandakrishnan R, Drozdetski A, Walker RC, Onufriev AV. 2015. Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations. *Biophys. J.* 108:1153–64
6. Anandakrishnan R, Onufriev AV. 2010. An *N* log *N* approximation based on the natural organization of biomolecules for speeding up the computation of long range interactions. *J. Comput. Chem.* 31:691–706
7. Arthur EJ, Brooks CL III. 2016. Parallelization and improvements of the generalized Born model with a simple sWitching function for modern graphics processors. *J. Comput. Chem.* 37:927–39
8. Baker NA, Bashford D, Case DA. 2006. Implicit solvent electrostatics in biomolecular simulation. In *New Algorithms for Macromolecular Simulation*, ed. B Leimkuhler, C Chipot, R Elber, A Laaksonen, A Mark, et al., pp. 263–95. Berlin: Springer-Verlag

9. Bashford D, Case DA. 2000. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* 51:129–52
10. Bashford D, Karplus M. 1990. pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* 29:10219–25
11. Bomble YJ, Case DA. 2008. Multiscale modeling of nucleic acids: insights into DNA flexibility. *Biopolymers* 89:722–31
12. Born M. 1920. Volumes and heats of hydration of ions. *Z. Phys.* 1:45–48
13. Brown RA, Case DA. 2006. Second derivatives in generalized Born theory. *J. Comput. Chem.* 27:1662–75
14. Chen J, Brooks C. 2008. Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions. *Phys. Chem. Chem. Phys.* 10:471–81
15. Chen J, Im W, Brooks C. 2006. Balancing solvation and intramolecular interactions: toward a consistent generalized Born force field. *J. Am. Chem. Soc.* 128:3728–36
16. Chocholoušová J, Feig M. 2006. Implicit solvent simulations of DNA and DNA–protein complexes: agreement with explicit solvent versus experiment. *J. Phys. Chem. B* 110:17240–51
17. Cramer CJ, Truhlar DG. 2008. A universal approach to solvation modeling. *Acc. Chem. Res.* 41:760–68
18. Duan Y, Kollman PA. 1998. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 282:740–44
19. Dutagaci B, Feig M. 2017. Determination of hydrophobic lengths of membrane proteins with the HDGB implicit membrane model. *J. Chem. Inform. Model.* 57:3032–42
20. Dutagaci B, Heo L, Feig M. 2018. Structure refinement of membrane proteins via molecular dynamics simulations. *Proteins* 86:738–50
21. Dutagaci B, Sayadi M, Feig M. 2017. Heterogeneous dielectric generalized Born model with a van der Waals term provides improved association energetics of membrane-embedded transmembrane helices. *J. Comput. Chem.* 38:1308–20
22. Eastman P, Pande V. 2010. Efficient nonbonded interactions for molecular dynamics on a graphics processing unit. *J. Comput. Chem.* 31:1268–72
23. Erler J, Zhang R, Petridis L, Cheng X, Smith JC, Langowski J. 2014. The role of histone tails in the nucleosome: a computational study. *Biophys. J.* 107:2902–13
24. Feig M, Im W, Brooks CL. 2004. Implicit solvation based on generalized Born theory in different dielectric environments. *J. Chem. Phys.* 120:903–11
25. Felts AK, Gallicchio E, Chekmarev D, Paris KA, Friesner RA, Levy RM. 2008. Prediction of protein loop conformations using the AGBNP implicit solvent model and torsion angle sampling. *J. Chem. Theory Comput.* 4:855–68
26. Fenley A, Killian B, Hnizdo V, Fedorowicz A, Sharp D, Gilson M. 2014. Correlation as a determinant of configurational entropy in supramolecular and protein systems. *J. Phys. Chem. B* 118:6447–55
27. Forouzesh N, Izadi S, Onufriev AV. 2017. Grid-based surface generalized Born model for calculation of electrostatic binding free energies. *J. Chem. Inform. Model.* 57:2505–13
28. Gallicchio E, Levy R. 2004. AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* 25:479–99
29. Genheden S, Essex J. 2015. A simple and transferable all-atom/coarse-grained hybrid model to study membrane processes. *J. Chem. Theory Comput.* 11:4749–59
30. Genheden S, Kongsted J, Soderhjelm P, Ryde U. 2010. Nonpolar solvation free energies of protein–ligand complexes. *J. Chem. Theory Comput.* 6:3558–68
31. Genheden S, Mikulskis P, Hu L, Kongsted J, Soderhjelm P, Ryde U. 2011. Accurate predictions of nonpolar solvation free energies require explicit consideration of binding-site hydration. *J. Am. Chem. Soc.* 133:13081–92
32. Ghosh A, Rapp C, Friesner R. 1998. Generalized Born model based on a surface integral formulation. *J. Phys. Chem. B* 102:10983–90
33. Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. 2012. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* 8:1542–55
34. Grant J, Pickup B. 1995. A Gaussian description of molecular shape. *J. Phys. Chem.* 99:3503–10

35. Grant J, Pickup B, Nicholls A. 2001. A smooth permittivity function for Poisson–Boltzmann solvation methods. *J. Comput. Chem.* 22:608–41
36. Grant J, Pickup B, Sykes M, Kitchen C, Nicholls A. 2007. A simple formula for dielectric polarisation energies: the Sheffield Solvation Model. *Chem. Phys. Lett.* 441:163–66
37. Grant J, Pickup B, Sykes M, Kitchen C, Nicholls A. 2007. The Gaussian Generalized Born model: application to small molecules. *Phys. Chem. Chem. Phys.* 9:4913–22
38. Grycuk T. 2003. Deficiency of the Coulomb-field approximation in the generalized Born model: an improved formula for Born radii evaluation. *J. Chem. Phys.* 119:4817–26
39. Havranek J, Harbury P. 1999. Tanford–Kirkwood electrostatics for protein modeling. *PNAS* 96:11145
40. Hawkins G, Cramer C, Truhlar D. 1995. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* 246:122–29
41. Hawkins G, Cramer C, Truhlar D. 1996. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* 100:19824–39
42. Hoijtink GJ, de Boer E, van der Meij PH, Weijland WP. 1956. Reduction potentials of various aromatic hydrocarbons and their univalent anions. *Recl. Trav. Chim. Pays-Bas* 75:487–503
43. Huang Y, Chen W, Wallace JA, Shen J. 2016. All-Atom continuous constant pH molecular dynamics with particle mesh ewald and titratable water. *J. Chem. Theory Comput.* 12:5411–21
44. Im W, Lee M, Brooks CL III. 2003. Generalized Born model with a simple smoothing function. *J. Comput. Chem.* 24:1691–702
45. Izadi S, Aguilar B, Onufriev AV. 2015. Protein–ligand electrostatic binding free energies from explicit and implicit solvation. *J. Chem. Theory Comput.* 11:4450–59
46. Izadi S, Anandakrishnan R, Onufriev AV. 2016. Implicit solvent model for million-atom atomistic simulations: Insights into the organization of 30-nm chromatin fiber. *J. Chem. Theory Comput.* 12:5946–59
47. Izadi S, Harris RC, Fenley MO, Onufriev AV. 2018. Accuracy comparison of generalized Born models in the calculation of electrostatic binding free energies. *J. Chem. Theory Comput.* 14:1656–70
48. Izairi R, Kamberaj H. 2017. Comparison study of polar and nonpolar contributions to solvation free energy. *J. Chem. Inf. Model.* 57:2539–53
49. Jackson J. 1975. *Classical Electrodynamics*. New York: Wiley and Sons
50. Jang S, Kim E, Shin S, Pak Y. 2003. Ab initio folding of helix bundle proteins using molecular dynamics simulations. *J. Am. Chem. Soc.* 125:14841–46
51. Jayaram B, Liu Y, Beveridge D. 1998. A modification of the generalized Born theory for improved estimates of solvation energies and *p*K shifts. *J. Chem. Phys.* 109:1465–71
52. Karplus M, Kushick J. 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules* 14:325–32
53. Khandogin J, Raleigh D, Brooks C. 2007. Folding intermediate in the villin headpiece domain arises from disruption of a N-terminal hydrogen-bonded network. *J. Am. Chem. Soc.* 129:3056–57
54. Kirkwood J. 1934. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J. Chem. Phys.* 2:351–61
55. Kirkwood J. 1939. The dielectric polarization of polar liquids. *J. Chem. Phys.* 7:911–19
56. Klamt A, Schüürmann G. 1993. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans.* 2:799–805
57. Kollman P, Massova I, Reyes C, Kuhn B, Huo S, et al. 2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accts. Chem. Res.* 33:889–97
58. Labute P. 2008. The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. *J. Comput. Chem.* 29:1693–98
59. Lange AW, Herbert JM. 2012. Improving generalized Born models by exploiting connections to polarizable continuum models. I. An improved effective Coulomb operator. *J. Chem. Theory Comput.* 8:1999–2011
60. Lee MS, Feig M, Salsbury FR, Brooks CL. 2003. New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* 24:1348–56

61. Lee MS, Olson M. 2013. Comparison of volume and surface area nonpolar solvation free energy terms for implicit solvent simulations. *J. Chem. Phys.* 139:044119
62. Lee MS, Salsbury FR, Brooks CL. 2002. Novel generalized Born methods. *J. Chem. Phys.* 116:10606–14
63. Lee MS, Salsbury FR, Brooks CL. 2004. Constant-pH molecular dynamics using continuous titration coordinates. *Proteins* 56:738–52
64. Lei H, Duan Y. 2007. Two-stage folding of HP-35 from ab initio simulations. *J. Mol. Biol.* 370:196–206
65. Levy R, Zhang L, Gallicchio E, Felts A. 2003. On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute–solvent interaction energy. *J. Am. Chem. Soc.* 125:9523–30
66. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. 2011. How fast-folding proteins fold. *Science* 334:517–20
67. Liu H, Duan Y. 2008. Effects of posttranslational modifications on the structure and dynamics of histone H3 N-terminal peptide. *Biophys. J.* 94:4579–85
68. Lopes A, Alexandrov A, Bathelt C, Archontis G, Simonson T. 2007. Computational sidechain placement and protein mutagenesis with implicit solvent models. *Proteins* 67:853–67
69. Marenich AV, Cramer CJ, Truhlar DG. 2009. Universal solvation model based on the generalized Born approximation with asymmetric descreening. *J. Chem. Theory Comput.* 5:2447–64
70. McQuarrie D. 1976. *Statistical Mechanics*. New York: Harper and Row
71. Mirjalili V, Feig M. 2015. Interactions of amino acid side-chain analogs within membrane environments. *J. Phys. Chem. B* 119:2877–85
72. Mongan J, Case DA. 2005. Biomolecular simulations at constant pH. *Curr. Opin. Struct. Biol.* 15:157–63
73. Mongan J, Case DA, McCammon JA. 2004. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* 25:2038–48
74. Mongan J, Simmerling C, McCammon JA, Case DA, Onufriev A. 2007. Generalized Born model with a simple, robust molecular volume correction. *J. Chem. Theory Comput.* 3:156–69
75. Mongan J, Svrcek-Seiler WA, Onufriev A. 2007. Analysis of integral expressions for effective Born radii. *J. Chem. Phys.* 127:185101
76. Mukhopadhyay A, Aguilar BH, Tolokh IS, Onufriev AV. 2014. Introducing charge hydration asymmetry into the generalized Born model. *J. Chem. Theory Comput.* 10:1788–94
77. Mukhopadhyay A, Fenley AT, Tolokh IS, Onufriev AV. 2012. Charge hydration asymmetry: the basic principle and how to use it to test and improve water models. *J. Phys. Chem. B* 116:9776–83
78. Mukhopadhyay A, Tolokh IS, Onufriev AV. 2015. Accurate evaluation of charge asymmetry in aqueous solvation. *J. Phys. Chem. B* 119:6092–100
79. Nguyen H, Maier J, Huang H, Perrone V, Simmerling C. 2014. Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *J. Am. Chem. Soc.* 136:13959–62
80. Nguyen H, Pérez A, Bermeo S, Simmerling C. 2015. Refinement of generalized Born implicit solvation parameters for nucleic acids and their complexes with proteins. *J. Chem. Theory Comput.* 11:3714–28
81. Nguyen H, Roe DR, Simmerling C. 2013. Improved generalized Born solvent model parameters for protein simulations. *J. Chem. Theory Comput.* 9:2020
82. Onufriev A. 2010. Continuum electrostatics solvent modeling with the generalized Born model. In *Modeling Solvent Environments*, ed. M Feig, pp. 127–65. Weinheim, Ger: Wiley. 1st ed.
83. Onufriev AV, Aguilar B. 2014. Accuracy of continuum electrostatic calculations based on three common dielectric boundary definitions. *J. Theor. Comput. Chem.* 13:1440006
84. Onufriev AV, Bashford D, Case DA. 2000. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* 104:3712–20
85. Onufriev AV, Bashford D, Case DA. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins* 55:383–94
86. Onufriev AV, Case DA, Bashford D. 2002. Effective Born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.* 23:1297–304
87. Onufriev AV, Izadi S. 2018. Water models for biomolecular simulations. *WIRES: Comput. Mol. Sci.* 8:e1347

88. Onufriev AV, Sigalov G. 2011. A strategy for reducing gross errors in the generalized Born models of implicit solvation. *J. Chem. Phys.* 134:164104–15
89. Robinson PJJ, Fairall L, Huynh VAT, Rhodes D. 2006. Em measurements define the dimensions of the 30-nm chromatin fiber: evidence for a compact, interdigitated structure. *PNAS* 103:6506–11
90. Rubenstein AB, Blacklock K, Nguyen H, Case DA, Khare SD. 2018. Systematic comparison of Amber and Rosetta energy functions for protein structure evaluation. *J. Chem. Theory Comput.* 14:6015–25
91. Ruscio JZ, Onufriev A. 2006. A computational study of nucleosomal DNA flexibility. *Biophys. J.* 91:4121–32
92. Rychkov GN, Ilatovskiy AV, Nazarov IB, Shvetsov AV, Lebedev DV, et al. 2017. Partially assembled nucleosome structures at atomic detail. *Biopys. J.* 112:460–72
93. Savin AV, Kikot IP, Mazo MA, Onufriev AV. 2013. Two-phase stretching of molecular chains. *PNAS* 110:2816–21
94. Savin AV, Mazo MA, Kikot IP, Manevitch LI, Onufriev AV. 2011. Heat conductivity of the DNA double helix. *Phys. Rev. B* 83:245406
95. Schaefer M, Froemmel C. 1990. A precise analytical method for calculating the electrostatic energy of macromolecules in aqueous solution. *J. Mol. Biol.* 216:1045–66
96. Schnieders M, Ponder J. 2007. Polarizable atomic multipole solutes in a generalized Kirkwood continuum. *J. Chem. Theory Comput.* 3:2083–97
97. Schutz C, Warshel A. 2001. What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* 44:400–17
98. Sharp KA, Honig B. 1990. Electrostatic interactions in macromolecules: theory and experiment. *Annu. Rev. Biophys. Biophys. Chem.* 19:301–32
99. Sigalov G, Fenley A, Onufriev A. 2006. Analytical electrostatics for biomolecules: beyond the generalized Born approximation. *J. Chem. Phys.* 124:124902
100. Sigalov G, Scheffel P, Onufriev A. 2005. Incorporating variable dielectric environments into the generalized Born model. *J. Chem. Phys.* 122:094511
101. Simmerling C, Strockbine B, Roitberg AE. 2002. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* 124:11258–59
102. Spassov VZ, Yan L, Szalma S. 2002. Introducing an implicit membrane in generalized Born/solvent accessibility continuum solvent models. *J. Phys. Chem. B* 106:8726–38
103. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA. 1998. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices. *J. Am. Chem. Soc.* 120:9401–9
104. Srinivasan J, Miller J, Kollman P, Case DA. 1998. Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* 16:671–82
105. Still W, Tempczyk A, Hawley R, Hendrickson T. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–29
106. Swails J, Roitberg A. 2012. Enhancing conformation and protonation state sampling of hen egg white lysozyme using pH replica exchange molecular dynamics. *J. Chem. Theory Comput.* 8:4393–404
107. Swails J, York D, Roitberg A. 2014. Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: implementation, testing, and validation. *J. Chem. Theory Comput.* 10:1341–52
108. Tan C, Tan Y, Luo R. 2007. Implicit nonpolar solvent models. *J. Phys. Chem. B* 111:12263–74
109. Tanford C, Kirkwood J. 1957. Theory of titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* 79:5333–39
110. Tanford C, Roxby R. 1972. Interpretation of protein titration curves. *Biochemistry* 11:2192–98
111. Tanizaki S. 2006. Molecular dynamics simulations of large integral membrane proteins with an implicit membrane model. *J. Phys. Chem. B* 110:548
112. Tanizaki S, Feig M. 2005. A generalized Born formalism for heterogeneous dielectric environments: application to the implicit modeling of biological membranes. *J. Chem. Phys.* 122:124706
113. Tolokh IS, Thomas DG, Onufriev AV. 2018. Explicit ions/implicit water generalized Born model for nucleic acids. *J. Chem. Phys.* 148:195101

114. Tomasi J, Mennucci B, Cammi R. 2005. Quantum mechanical continuum solvation models. *Chem. Rev.* 105:2999–3094
115. Tsui V, Case DA. 2000. Molecular dynamics simulations of nucleic acids using a generalized Born solvation model. *J. Am. Chem. Soc.* 122:2489–98
116. Ulmschneider M, Ulmschneider J, Sansom M, DiNola A. 2007. A generalized Born implicit-membrane representation compared to experimental insertion free energies. *Biophys. J.* 92:2338–49
117. Wallace JA, Wang Y, Shi C, Pastoor KJ, Nguyen BL, et al. 2011. Toward accurate prediction of pK_a values for internal protein residues: the importance of conformational relaxation and desolvation energy. *Proteins* 79:3364–73
118. Wang C, Greene D, Xiao L, Qi R, Luo R. 2018. Recent developments and applications of the MMPBSA method. *Front. Mol. Biosci.* 4:87
119. Weiser J, Shenkin P, Still W. 1999. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* 20:217–30
120. Xu Z, Cai W. 2011. Fast analytical methods for macroscopic electrostatic models in biomolecular simulations. *SIAM Rev.* 53:683–723
121. Zhou B, Trinajstic N. 2007. On the largest eigenvalue of the distance matrix of a connected graph. *Chem. Phys. Lett.* 447:384–87
122. Zhou H-X, Pang X. 2018. Electrostatic interactions in protein structure, folding, binding, and condensation. *Chem. Rev.* 118:1691–741

Annual Review of Biophysics
Volume 48, 2019

Contents