

Improving Generalized Born Models by Exploiting Connections to Polarizable Continuum Models. I. An Improved Effective Coulomb Operator

Adrian W. Lange[†] and John M. Herbert*

Department of Chemistry, The Ohio State University, Columbus, Ohio 43210, United States

ABSTRACT: We investigate the generalized Born (GB) implicit solvation model in comparison with polarizable continuum models (PCMs). We show that the GB model is intimately connected to the conductor-like PCM (C-PCM), a method that is accurate for high-dielectric solvents but less so for weakly polar and nonpolar solvents. The formal connection between C-PCM and the GB model suggests that C-PCM calculations place a limit on the accuracy that one should expect from GB models but also demonstrates that comparison of GB and C-PCM calculations directly interrogates the accuracy of the effective Coulomb operator that is used in the pairwise GB energy expression. We introduce a simple alternative to the “canonical” pairwise interaction operator first proposed by Still et al. and show that this alternative reduces the cost of the pairwise GB energy summation by as much as a factor of 3. At the same time, the new operator reduces statistical errors in solvation energies (as compared to C-PCM benchmarks) by 0.3% with respect to the canonical operator that exhibits an error of roughly 1.0%.

I. INTRODUCTION

Solvent effects are crucial in biochemical simulations,^{1–4} but explicitly accounting for the multitude of solvent molecules surrounding a solute can often make a simulation too expensive to carry out. To avoid the unfeasible, various “implicit” solvent methods have been developed in an effort to capture the most important solvent effects without introducing explicit solvent molecules. The mean-field electrostatic solvation effect, which is the focus of the present work, is commonly modeled using a linear, isotropic dielectric continuum solvent governed by the Poisson equation (PE). For a solute whose charge distribution is $\rho(\vec{r})$, this equation is

$$\hat{\nabla}^2[\epsilon(\vec{r})U(\vec{r})] = -4\pi\rho(\vec{r}) \quad (1.1)$$

where $U(\vec{r})$ is the total electrostatic potential and $\epsilon(\vec{r})$ is the position-dependent dielectric function. The merits of (and problems associated with) various numerical PE solvers have been discussed at length,^{3–14} but this is not the focus of the present work. Instead, we focus on a class of boundary-element methods known as “apparent surface charge” polarizable continuum models (PCMs).¹⁵ As compared to direct solution of eq 1.1, these models offer a more efficient means to compute the electrostatic solvation energy, approximately but often with high accuracy.

For macromolecular simulations, implicit solvation using an accurate, grid-based PE solver might yet be too computationally expensive, and an even less expensive—but possibly less accurate—method becomes necessary. Generalized Born (GB) models offer precisely this.^{7,12,16–18} Such models approximate the electrostatic solvation energy arising from a set of point charges (the solute) in a pairwise-additive way using a modified Coulomb interaction. This approach is far less expensive as compared to sophisticated PE solvers yet in many cases exhibits reasonable accuracy.¹⁹ Consequently, GB models are quite popular in molecular mechanics (MM) applications to biomolecules.

The continuing development of GB theory has been primarily aimed at closing the gap in accuracy between GB models and PE solvers. Disagreement in energy between the two is most often attributed to the effective Born radii used in the GB calculations. To isolate errors attributable to the pairwise-additive electrostatic interaction term, one may compute “perfect” effective Born radii,²⁰ as discussed in section II of this work. The perfect radius for a given atom is defined to be the radius such that the Born ion formula,²¹ using the perfect radius, affords the exact electrostatic solvation energy for a system in which only that particular atom’s charge is present but where the solute cavity is maintained as if all solute atoms were present. When perfect radii are used, any error in the GB solvation energy is attributable solely to the choice of the effective Coulomb operator that is used in the pairwise-additive energy formula. Several methods exist by means of which nearly perfect radii can be computed at fairly low cost.^{22–28}

Further improvements to GB theory may be possible by considering the effective Coulomb operator. The operator introduced by Still et al.¹⁶ more than 20 years ago remains the most widely used choice; other effective Coulomb operators have been proposed since that time,^{20,24,29–32} but none has become widely adopted, and these alternative operators are typically unavailable in common MM software packages. This may be due to the fact that these alternatives tend to be more complicated than the “canonical” (Still et al.) operator and therefore increase the computational expense of GB calculations. For example, Onufriev and Sigalov³² recently presented a very thorough analysis of effective Coulomb operators and also proposed a new one that improves upon the canonical choice, but this improvement comes at the price of a significant increase in computational complexity (e.g., depend-

Received: February 7, 2012

Published: April 13, 2012

ence on the gradients of the effective Born radii). The Still operator will likely continue to be the preferred choice unless an alternative can be found that is more accurate but not substantially more expensive.

In this work, we examine GB models in the context of a comparison to PCMs. We show that for the conventional GB model both the perfect effective Born radii and the effective Coulomb operator must depend on the dielectric constant in order to reproduce the exact electrostatic solvation energy. In the conductor limit, however, the dielectric-dependent terms vanish and an equivalence between GB theory and the conductor-like PCM (C-PCM,³³ also known as GCOS-MO^{34,35}) can be established rigorously. If one assumes, as in conventional GB theory, that the radii and effective Coulomb operator are independent of the solvent dielectric constant then this equivalence implies that the accuracy of GB models is limited by the accuracy of C-PCM calculations. This equivalence provides the means for an incisive investigation of the accuracy of various effective Coulomb operators in GB theory. We exploit this to test the accuracy of some novel pairwise interaction operators that are less expensive to evaluate as compared to Still's prescription; remarkably, we are able to find an operator that is both *more* accurate and *less* expensive than the canonical choice.

Our presentation is organized as follows. We briefly review the theory of GB models (section II) and then draw comparisons to PCM theory (section III). In particular, the equivalence between GB theory and C-PCM is derived in section III.E. We next consider effective Coulomb operators in the GB model (section IV) and propose a new one that is less expensive to evaluate. Finally, we present a series of calculations on 16 small proteins and 54 larger biomolecules in order to compare the accuracy of various effective Coulomb operators against C-PCM results (section V). Atomic units are used in the equations (so there is no $4\pi\epsilon_0$ in the Coulomb potential), but numerical results are presented in kcal/mol.

II. GENERALIZED BORN MODELS

II.A. Conventional GB Model. In the GB model and other implicit solvent models, the mean-field electrostatics of a bulk solvent are approximated as a homogeneous dielectric medium with a characteristic dielectric constant, ϵ . The solute resides in a cavity carved out of the dielectric medium. For simplicity, we assume throughout this work that $\epsilon = 1$ inside of the cavity, which is appropriate since electrostatic interactions are treated *explicitly* within the cavity via the MM force field. The electric field due to the solute charge distribution polarizes the medium, and in response, the dielectric produces an electrostatic reaction field that interacts with the solute.³⁶ The total electrostatic energy of this system can be written as

$$G_{\text{tot}} = G_0 + G_{\text{pol}} \quad (2.1)$$

where G_0 is the gas-phase electrostatic energy of the solute charge density, $\rho(\vec{r})$, and G_{pol} is the polarization energy arising from interaction with the solvent reaction field. In the conventional GB model, the solute charge distribution is modeled with a set of atom-centered point charges, q_i , and the polarization energy is expressed as

$$G_{\text{pol}}^{\text{GB}} = \frac{1}{2} \left(\frac{1 - \epsilon}{\epsilon} \right) \sum_{i,j} \frac{q_i q_j}{f_{ij}} \quad (2.2)$$

The quantity f_{ij}^{-1} is the effective Coulomb operator. This operator typically depends upon atom-specific information such as effective Born radii and the distance r_{ij} between the atoms. (A more detailed discussion of effective Coulomb operators is given in section IV.)

For future reference, we define pairwise contributions to the electrostatic solvation energy (cf. eq 2.2)

$$G_{\text{pol},ij} = \frac{1}{2} \left(\frac{1 - \epsilon}{\epsilon} \right) \frac{q_i q_j}{f_{ij}} \quad (2.3)$$

If $i = j$ or in the limit that $r_{ij} \rightarrow 0$, one demands that f_{ij} reduce to the effective Born radius of the i th atom, R_i . This implies that the self-interaction terms ($i = j$ in eq 2.2) reduce to the Born ion expression²¹

$$G_{\text{pol},ii} = \frac{1}{2} \left(\frac{1 - \epsilon}{\epsilon} \right) \frac{q_i^2}{R_i} \quad (2.4)$$

The effective radius R_i can be loosely interpreted as an average measure of the distance from the i th atomic center to the cavity surface where a strong reaction field potential is produced. The "perfect" value of R_i is obtained by solving the PE for a system in which q_i is the only charge present, yet the solute cavity is the same as if all atoms were present.²⁰ The polarization energy thus obtained, $G_{\text{pol},ii}^{\text{PE}}$ can then be substituted for $G_{\text{pol},ii}$ in eq 2.4 to obtain an expression for the "perfect" value of R_i ,

$$R_i = \frac{1}{2} \left(\frac{1 - \epsilon}{\epsilon} \right) \frac{q_i^2}{G_{\text{pol},ii}^{\text{PE}}} \quad (2.5)$$

Computing perfect radii in this manner is absurd in practical applications, since it involves numerous expensive PE calculations that GB methods seek to avoid in the first place. Instead, various procedures designed to approximate the perfect R_i have been developed, including surface integral techniques,^{25,37,38} volumetric integration methods,^{22,23,26,27} and highly parametrized empirical formulas.³⁹

Given these definitions, eq 2.2 can be rewritten as

$$G_{\text{pol}}^{\text{GB}} = \left(\frac{1 - \epsilon}{\epsilon} \right) \sum_i \left(\frac{q_i^2}{2R_i} + \sum_{j>i} \frac{q_i q_j}{f_{ij}} \right) \quad (2.6)$$

The first term in this equation is a sum of atomic self-energies, while the second term is a pairwise-additive approximation to the interatomic Coulomb energy.

II.B. GB ϵ Model. The "GB ϵ model",³⁰ also known as the *analytical linearized Poisson–Boltzmann* (ALPB) model,³¹ is a relatively elaborate GB model that employs a pairwise interaction energy of the form

$$G_{\text{pol}}^{\text{GB}\epsilon} = \frac{1}{2} \left(\frac{1 - \epsilon}{\epsilon} \right) \sum_{i,j} \frac{q_i q_j}{1 + \alpha/\epsilon} \left(\frac{1}{f_{ij}^\epsilon} + \frac{\alpha}{\epsilon A} \right) \quad (2.7)$$

Here, α is a special nonadjustable parameter that depends upon both ϵ and the solute cavity and supposedly can be derived from theory, although doing so is not straightforward without a priori knowledge of the exact solution to the PE for the given solute cavity. However, this parameter is found to be rather insensitive to variations in both cavity shape and ϵ , and the value $\alpha = 0.579$ has been suggested for general use.^{30,40} The parameter A in eq 2.7 is the *electrostatic size* of the solute, which provides a rough measure of cavity size. For a spherical cavity, A

is equal to the cavity radius but more elaborate formulas have been proposed for obtaining A in general.^{30,31} Equation 2.7 reduces to the familiar Born ion model for a single charge centered in a spherical cavity. Note also that this equation contains an effective Coulomb operator, $1/f_{ij}^e$, and this is where the effective Born radii enter into the GB ϵ model. We will show in section III.C how the functions f_{ij} and f_{ij}^e are related.

The difference between the conventional GB model (eq 2.2) and the GB ϵ model (eq 2.7) is the presence of the additional factors and terms in the latter that involve α , A , and ϵ . In the limit $\alpha/\epsilon \rightarrow 0$, the GB ϵ model becomes equivalent to the conventional GB model. Using the recommended value $\alpha = 0.579$ for room-temperature water ($\epsilon \approx 80$), one has $\alpha/\epsilon \approx 0.0072$, meaning that the ϵ -dependent terms are quite small. Our numerical calculations (section V) focus exclusively on water; hence, we expect little difference between the GB and the GB ϵ models. For low-dielectric solvents, discrepancies between GB and GB ϵ calculations may be more pronounced, just as discrepancies between C-PCM and exact PCM calculations are large when ϵ is small but disappear as $\epsilon \rightarrow \infty$.⁴¹

III. COMPARISON TO POLARIZABLE CONTINUUM MODELS

III.A. Background. Polarizable continuum models (PCMs) are a family of implicit solvent methods that utilize an *apparent surface charge* formalism to determine the electrostatic solvation energy (either exactly or approximately, depending upon which PCM is used) via a boundary-element approach.^{15,42} This formalism only requires the evaluation of two-dimensional integrals over the solute cavity surface, rather than three-dimensional volumetric integration, and is therefore significantly less expensive as compared to grid-based PE solvers. PCMs assume a sharp boundary at the cavity surface separating the cavity interior from the continuum exterior, as in the Born ion model that forms the basis of GB models. PCMs are applicable to any solute charge density, whereas the GB and GB ϵ models are defined only for point charges, meaning that more complicated charge densities must first be collapsed onto point charges in order to use GB models.⁴³ Because $\rho(\vec{r})$ can be arbitrary, PCMs are very popular for quantum-mechanical electronic structure calculations in which the solvent reaction field is iterated to self-consistency with the electron density. Oddly, PCMs have not enjoyed much attention in the MM simulation community, which has largely favored three-dimensional PE solvers for assessing the accuracy of GB models.

The most sophisticated and accurate PCMs are known as the *integral equation formalism* (IEF-PCM)⁴⁷ and also the *surface and simulation of volume polarization for electrostatics* [SS(V)PE] model.⁴⁸ At the level of analytical integral equations, IEF-PCM and SS(V)PE are equivalent, but the two methods differ in discretization and implementation details, which can have a profound effect on the computed solvation energy.⁴¹ If $\rho(\vec{r})$ is contained entirely within the solute cavity, as is the case for classical solutes comprised of atomic partial charges, then the solvation energy obtained using either IEF-PCM or SS(V)PE is exactly the same as what would be obtained by solving the PE, for any cavity shape. If some portion of the solute charge density exists outside the cavity (e.g., the tail of an electron density) then these methods are exact for the interior charge density and provide an approximation for the *volume polarization* that results from the exterior charge density.⁴⁸ This approximation has been shown to be quite accurate.⁴²

IEF-PCM and SS(V)PE calculations, while less expensive than three-dimensional PE solvers, remain fairly expensive for macromolecular solutes. A computationally simpler alternative is the *conductor-like screening model* (COSMO).⁴⁹ In this work, we focus on a slightly modified version of the original COSMO that is known variously as either “generalized COSMO” (GCOSMO) or the conductor-like PCM (C-PCM),^{33–35} which is better justified on formal theoretical grounds,⁴¹ although the numerical differences are small in high-dielectric solvents. (Explicit equations for these models are given in Appendix A.)

Unlike the IEF-PCM and SS(V)PE approaches, C-PCM calculations provide only an approximate electrostatic solvation energy, even for solutes composed of classical point charges. However, the accuracy of this approximation improves as ϵ increases, such that for $\epsilon \approx 80$ (and in some cases for dielectric constants as low as $\epsilon \approx 10$) the error is negligible as compared to IEF-PCM/SS(V)PE.⁴¹ In the conductor limit ($\epsilon \rightarrow \infty$), all of these PCMs are equivalent and provide the exact electrostatic solvation energy for a given cavity shape, provided that $\rho(\vec{r})$ is contained entirely within the cavity.

Previous comparisons of PCM and GB results have been mostly superficial, either comparing the computed total solvation energies^{5,25,50–53} or in some cases simply stating that a solvent reaction field is obtainable with either a PCM or a GB model.^{5,15,54} Such comparisons overlook how these methods are related at a more fundamental level. We seek to highlight this relationship. In what follows we derive analytical equations that place PCMs and GB models on a common theoretical footing that facilitates comparison of these models.

III.B. Polarization Energy. In PCM theory, the polarization of the dielectric medium is represented by an “apparent” (i.e., effective) surface charge density that resides on the surface of the solute cavity.¹⁵ A point on the cavity surface will be denoted by \vec{s} , whereas an arbitrary point in space will be denoted \vec{r} . The surface charge, $\sigma(\vec{s})$, produces a reaction-field potential, $\chi(\vec{r})$, that interacts with the solute charge density, $\rho(\vec{r})$. The total electrostatic solvation energy for any PCM can be expressed as (cf. eq 2.1)

$$G_{\text{tot}}^{\text{PCM}} = G_0 + G_{\text{pol}}^{\text{PCM}} \quad (3.1)$$

The electrostatic solvation energy, $G_{\text{pol}}^{\text{PCM}}$, is

$$G_{\text{pol}}^{\text{PCM}} = \frac{1}{2} \int d\vec{r} \rho(\vec{r}) \chi(\vec{r}) \quad (3.2)$$

or equivalently

$$G_{\text{pol}}^{\text{PCM}} = \frac{1}{2} \int d\vec{s} \sigma(\vec{s}) \phi(\vec{s}) \quad (3.3)$$

where $\phi(\vec{s})$ is the electrostatic potential due to $\rho(\vec{r})$, evaluated at the cavity surface. To determine $\sigma(\vec{s})$, we introduce an integral operator \hat{Q} (called the *response operator*) that maps $\phi(\vec{s})$ onto $\sigma(\vec{s})$:

$$\sigma(\vec{s}) = \hat{Q}\phi(\vec{s}) \quad (3.4)$$

Each PCM is defined by its particular form for \hat{Q} . Definitions for \hat{Q} and other PCM integral operators can be found in Appendix A.

To make a useful comparison with GB models, we assume henceforth that the solute is composed of point charges,

$$\rho(\vec{r}) = \sum_i q_i \delta(\vec{r} - \vec{r}_i) \quad (3.5)$$

We then partition $\phi(\vec{s})$ into the sum of contributions from each individual point charge,

$$\phi(\vec{s}) = \sum_i \phi_i(\vec{s}) \quad (3.6)$$

The mapping in eq 3.4 then partitions $\sigma(\vec{s})$ according to

$$\sigma(\vec{s}) = \sum_i \sigma_i(\vec{s}) \quad (3.7)$$

where $\sigma_i = \hat{Q} \phi_i$. Inserting these partitions into eq 3.3 we have

$$G_{\text{pol}}^{\text{PCM}} = \frac{1}{2} \sum_{i,j} \int d\vec{s} \sigma_i(\vec{s}) \phi_j(\vec{s}) \quad (3.8)$$

We conclude that the PCM electrostatic solvation energy is decomposable into a form analogous to eq 2.6 for GB models. The $i = j$ terms in eq 3.8 constitute the self-energy energy, and the $i \neq j$ terms comprise the pairwise-additive energy.

III.C. Effective Coulomb Operator. Comparing eqs 2.2 and 3.8 we obtain the following analytical relationship involving the GB effective Coulomb operator:

$$\left(\frac{1 - \epsilon}{\epsilon} \right) \frac{q_i q_j}{f_{ij}} = \int d\vec{s} \sigma_i(\vec{s}) \hat{Q}^{-1} \sigma_j(\vec{s}) \quad (3.9)$$

The right side of this equation is a complicated surface integral and does not provide an obvious route to defining a better f_{ij} . Nevertheless, it is informative to substitute the IEF-PCM/SS(V)PE version of the response operator (see Appendix A) into eq 3.9. Upon rearranging, the result is

$$\frac{q_i q_j}{f_{ij}} = k_\epsilon^2 \int d\vec{s} \sigma_i(\vec{s}) \hat{C}_\epsilon \sigma_j(\vec{s}) \quad (3.10)$$

where

$$k_\epsilon = \frac{\epsilon}{\epsilon - 1} \quad (3.11)$$

and

$$\hat{C}_\epsilon = \left[\hat{I} + \frac{1}{\epsilon} (2\hat{Y}^{-1} - \hat{I}) \right] \hat{S} \quad (3.12)$$

The operators in eq 3.12 are defined in Appendix A. Note the conductor limit, $\hat{C}_\infty = \hat{S}$ and $k_\infty = -1$.

Defining

$$\tilde{\phi}_i(\vec{s}) = \frac{\phi_i(\vec{s})}{q_i} \quad (3.13)$$

we can rearrange eq 3.10 using eq 3.4 to obtain

$$\frac{1}{f_{ij}} = \int d\vec{s} \tilde{\phi}_i(\vec{s}) \hat{C}_\epsilon^{-1} \tilde{\phi}_j(\vec{s}) \quad (3.14)$$

This equation implies that in order to reproduce the exact electrostatic solvation energy, the operator f_{ij} in the conventional GB model must depend on ϵ , since \hat{C}_ϵ depends explicitly on ϵ . In the conductor limit, however, the ϵ -dependent terms vanish:

$$\frac{1}{f_{ij}^\infty} = \int d\vec{s} \tilde{\phi}_i(\vec{s}) \hat{S}^{-1} \tilde{\phi}_j(\vec{s}) \quad (3.15)$$

This result is interesting because the canonical choice for f_{ij} , originally proposed by Still et al.,¹⁶ does *not* depend on ϵ (see section IV) nor do most alternatives. Therefore, even if perfect effective Born radii are obtained, the pairwise interactions in the conventional GB model will be incorrect to some extent without an ϵ -dependent Coulomb operator. This discrepancy vanishes as $\epsilon \rightarrow \infty$, which explains how these models can still perform reasonably well for water.

We can make an analogous comparison to the pairwise interactions in the GB ϵ model with the following result:

$$\frac{1}{1 + \alpha/\epsilon} \left(\frac{1}{f_{ij}^\infty} + \frac{\alpha}{\epsilon A} \right) = \int d\vec{s} \tilde{\phi}_i(\vec{s}) \hat{C}_\epsilon^{-1} \tilde{\phi}_j(\vec{s}) \quad (3.16)$$

Here, we have explicitly written that the effective Coulomb operator used in the GB ϵ model, which we previously called f_{ij}^ϵ , must be evaluated in the conductor limit (f_{ij}^∞), which is easily seen by taking the limit $\epsilon \rightarrow \infty$ to recover eq 3.15. This important feature has been known since the original derivation of the GB ϵ model,³⁰ and the GB ϵ model does incorporate a dependence on ϵ into the pairwise interactions. It seems reasonable to expect that one could use eq 3.16 to arrive at explicit expressions for α and/or A for nonspherical cavities, but our attempts to do so have not yielded anything more fruitful than what has already been suggested by other authors.

III.D. Effective Born Radii. The function f_{ij} reduces to an effective Born radius in the limit that $r_{ij} \rightarrow 0$. On the basis of the analysis in section III.C we deduce the following expression for the perfect effective Born radius in the conventional GB model:

$$\frac{1}{R_i} = \int d\vec{s} \tilde{\phi}_i(\vec{s}) \hat{C}_\epsilon^{-1} \tilde{\phi}_i(\vec{s}) \quad (3.17)$$

The dependence on ϵ vanishes only in the conductor limit, in which case the perfect radius is given by

$$\frac{1}{R_i^\infty} = \int d\vec{s} \tilde{\phi}_i(\vec{s}) \hat{S}^{-1} \tilde{\phi}_i(\vec{s}) \quad (3.18)$$

Note that $f_{ij}^\infty = R_i^\infty$ for $i = j$.

III.E. Equivalence of the GB Model to C-PCM. Consider the GB polarization energy in the conductor limit. In that limit, both the GB and the GB ϵ models afford

$$G_{\text{pol}}^{\text{GB},\infty} = -\frac{1}{2} \sum_{i,j} \frac{q_i q_j}{f_{ij}^\infty} \quad (3.19)$$

To recover an expression akin to eq 2.2 one only needs to scale by $(1 - \epsilon)/\epsilon$:

$$G_{\text{pol}}^{\text{GB}} = \left(\frac{1 - \epsilon}{\epsilon} \right) G_{\text{pol}}^{\text{GB},\infty} \quad (3.20)$$

Using eq 3.15 for f_{ij}^∞ we can rewrite eq 3.20 as

$$G_{\text{pol}}^{\text{GB}} = -\frac{1}{2} k_\epsilon^{-1} \sum_{i,j} \int d\vec{s} \phi_i(\vec{s}) \hat{S}^{-1} \phi_j(\vec{s}) \quad (3.21)$$

Recognizing that (see Appendix A)

$$\hat{Q}^{\text{C-PCM}} = -k_\epsilon^{-1} \hat{S}^{-1} \quad (3.22)$$

one finds that the right side of eq 3.21 is precisely the C-PCM polarization energy. We thus obtained the remarkable equivalence that

$$G_{\text{pol}}^{\text{GB}} = G_{\text{pol}}^{\text{C-PCM}} \quad (3.23)$$

This equivalence between GB and C-PCM polarization energies has been derived under the assumption of employing R_i^∞ and f_{ij}^∞ , which are arguably the best possible ϵ -independent choices for use in conventional GB theory, since any other ϵ -independent choices would necessarily lack the correct asymptotic limit as $\epsilon \rightarrow \infty$. In other words, conventional GB theory will, at best, reproduce C-PCM results when using perfect effective Born radii, obtained in the conductor limit, along with an effective Coulomb operator f_{ij}^∞ that is appropriate for the conductor limit. Unless ϵ -dependent radii and an ϵ -dependent effective Coulomb operator are employed, which is atypical, the accuracy of conventional GB models—in terms of reproducing the exact electrostatic solvation energy obtained from Poisson's equation—is limited by the accuracy of C-PCM. Note that unlike the IEF-PCM/SS(V)PE method, the C-PCM approach is *not* exact for finite ϵ , even for point-charge solutes, except in certain special cases such as the Born ion model. Therefore, if conventional GB results happen to be more accurate than C-PCM, this can only result from some fortuitous cancellation of errors or system-specific tailoring of parameters.

We stress that eq 3.23 holds only when $\rho(\vec{r})$ consists of point charges and not for the more complicated charge distributions that arise in electronic structure calculations or the higher order multipoles that are sometimes employed in polarizable force fields. C-PCM can handle these more complicated cases directly, whereas to apply a GB model, the electrostatic potential would first have to be expressed in terms of appropriate point charges. In Appendix B, we show that the pairwise-additive approximation in GB theory can be extended to higher order multipoles, as is done in the generalized Kirkwood model.^{44,45}

The equivalence expressed in eq 3.23 has a few additional implications worth mentioning. Whereas GB models are often regarded as empirical constructions born from dubious approximations, eq 3.23 provides a direct connection to a model (C-PCM) that can be derived from Poisson's equation using well-defined approximations.⁵⁵ These approximations can furthermore be shown to introduce errors of order ϵ^{-1} ;⁵⁵ hence, the accuracy is often quite good in high-dielectric solvents. Thus, the empiricism in GB calculations is restricted to how one computes approximate values for R_i^∞ and what functional form is chosen for f_{ij}^∞ in practice.

Note that the equivalence in eq 3.23 is valid for arbitrary ϵ , although it does require that the GB model use Born radii and an effective Coulomb operator that are appropriate for the conductor limit. Equation 3.23 further implies that if one uses C-PCM calculations to obtain perfect effective Born radii then these radii will be R_i^∞ (see eq 3.18) no matter what value of ϵ is used in the C-PCM calculation. This is a useful fact because it means that when we compare C-PCM calculations to GB results that employ perfect radii obtained from the C-PCM calculation any discrepancy between the two solvation energies is directly attributable to the effective Coulomb operator, $(f_{ij}^\infty)^{-1}$. This makes comparisons to C-PCM especially useful for developing improvements to the function f_{ij}^∞ . As compared to three-dimensional PE solvers, we find that it is generally easier and less expensive to converge C-PCM calculations with

respect to the quadrature grid that is used to discretize the cavity surface,⁴¹ so the equivalence expressed in eq 3.23 is very convenient from the standpoint of benchmarking.

We confirmed that the equivalence in eq 3.23 holds, numerically, to better than 10^{-6} kcal/mol for the case of a spherical cavity containing any number of point charges located at arbitrary interior positions. For this special case, analytic equations for R_i^∞ and f_{ij}^∞ are available.²⁴ These tests furthermore confirm that the GB ϵ model is extremely accurate for all values of ϵ , including smaller values for which C-PCM is significantly less accurate than IEF-PCM/SS(V)PE.

IV. EFFECTIVE COULOMB OPERATORS

IV.A. Traditional Models. Defining the effective Coulomb operator, f_{ij}^{-1} , poses a major challenge for GB theory because its analytic expression is known only in two limiting cases, yet its definition is crucial to the success of GB models. One of these cases is that of two point charges located at arbitrary positions inside of a spherical cavity surrounded by a conductor. In this case, the effective Coulomb operator

$$f_{ij}^{\text{sphere}} = \sqrt{r_{ij}^2 + R_i^\infty R_j^\infty} \quad (4.1)$$

affords the exact G_{pol} if R_i^∞ and R_j^∞ are both perfect effective Born radii.²⁴ The perfect radii can be computed from an analytic expression in this case; see section V.B. Equation 4.1 naturally reduces to the self-energy expression when $r_{ij} = 0$.

Equation 4.1 was derived by Grycuk,²⁴ but well before that time it was recognized by Still et al.¹⁶ that f_{ij}^{sphere} tends to underestimate the longer-range interaction of two proximal charges centered in spheres that are fully exposed to the dielectric. In this case,

$$f_{ij}^{\text{long}} = r_{ij} \quad (4.2)$$

is the exact Coulomb operator, provided that the two spheres do not overlap.^{55,56}

Equations 4.1 and 4.2 are the only exact limits for which the effective Coulomb operator is known. For all other cases, which constitute the overwhelming majority of interactions in practice, simple pairwise-additive analytic expressions for f_{ij} are not known. Knowledge of these exact limits seems to have been what prompted Still et al.¹⁶ to introduce an empirical function, f_{ij}^{still} , that interpolates between them:

$$f_{ij}^{\text{still}} = \sqrt{r_{ij}^2 + R_i R_j e^{-B_{ij}}} \quad (4.3)$$

Here,

$$B_{ij} = \frac{r_{ij}^2}{c R_i R_j} \quad (4.4)$$

and c is an empirical parameter. The value $c = 4$ was proposed by Still et al., but subsequent authors have suggested other values including $c = 1.64$,²⁹ 3.7 ,⁵⁷ 5.249 ,²⁷ and 8 .²³

The concept of the effective Coulomb operator as an interpolation function can be expressed more generally as³²

$$f_{ij} = \sqrt{r_{ij}^2 + R_i R_j \Psi_{ij}} \quad (4.5)$$

where Ψ_{ij} is some function chosen to obtain the proper limits. The choice $\Psi_{ij}^{\text{sphere}} \equiv 1$ recovers eq 4.1, whereas $\Psi_{ij}^{\text{long}} \equiv 0$ recovers eq 4.2 and $\Psi_{ij}^{\text{still}} = \exp(-B_{ij})$ yields the canonical effective Coulomb operator.

IV.B. Simplified Models. One school of thought for improving the effective Coulomb operator in the GB model is to introduce further complexity to account for the pieces that the canonical (Still et al.) operator is missing. This is almost certainly the correct approach if one desires to develop a GB model whose accuracy is competitive with that of PE solvers. However, more complexity is almost invariably tied to greater computational expense, and for biomolecular simulations the increased cost may not be justified if the gain in accuracy is modest.

We take a somewhat different approach here and attempt to enhance the computational efficiency of GB calculations while maintaining, if not improving upon, the accuracy of the canonical operator. Our strategy is to identify the CPU-intensive parts of f_{ij}^{still} and replace these with less expensive operations. After all, the effective Coulomb operator is known only in two limits, so there is really no theoretical justification for the CPU-intensive function calls (exponential and square-root functions) in eqs 4.3 and 4.5. In what follows we assume—as is typically the case—that the atomic coordinates and effective Born radii are stored and available in core memory. Note that r_{ij}^2 can be computed from atomic coordinates without calling the square-root function.

The square-root function in eq 4.5 can be obviated by a clever choice of the interpolating function:

$$\Psi_{ij} = \frac{2\Omega_{ij}r_{ij}}{\sqrt{R_iR_j}} + \Omega_{ij}^2 \quad (4.6)$$

With this choice, eq 4.5 for f_{ij} becomes

$$f_{ij} = r_{ij} + \Omega_{ij}\sqrt{R_iR_j} \quad (4.7)$$

The function Ω_{ij} will be chosen below in order to satisfy two limiting conditions:

$$\Omega_{ij} \rightarrow 1 \text{ as } r_{ij}^2/R_iR_j \rightarrow 0 \quad (4.8a)$$

$$\Omega_{ij} \rightarrow 0 \text{ as } r_{ij}^2/R_iR_j \rightarrow \infty \quad (4.8b)$$

First, however, let us comment on computational considerations. Although eq 4.7 may initially appear to complicate matters, we can choose to store $R_i^{1/2}$ in core memory rather than R_i , thereby eliminating calls to the square-root function. This represents a very minor (and linear-scaling) addition to the computational overhead that can be performed prior to the quadratic-scaling pairwise loop. In addition, r_{ij} should be available from other pairwise interactions, namely, the vacuum Coulomb interaction for G_0 (eq 2.1). Thus, if the pairwise GB energy is computed inside of the vacuum Coulomb pairwise loop, which is common practice, then evaluating eq 4.7 only requires computing the yet-to-be-defined function Ω_{ij} .

A simple choice for Ω_{ij} that satisfies the conditions in eqs 4.8a and 4.8b and also provides good accuracy, according to calculations discussed in section V, is

$$\Omega_{ij}^{\text{exp}} = \exp\left(\frac{-\zeta r_{ij}}{\sqrt{R_iR_j}}\right) \quad (4.9)$$

with $\zeta = 1.021$. However, we prefer to avoid the costly exponential function and instead make use of the following approximation

$$e^{-x} \approx \left(\frac{1}{1+x/p}\right)^p \quad (4.10)$$

which is valid for $p \geq 1$ and becomes exact in the limit $p \rightarrow \infty$. A low-order approximation to the exponential function can thus be obtained using only a few floating-point operations. The quantity $(1+x/p)^{-1}$ can be computed for about the same cost as computing B_{ij} in eq 4.4 and can then be successively self-multiplied $\log_2 p$ times to obtain the approximation in eq 4.10. One may view the value of p as just another adjustable parameter or merely as a part of selecting an interpolation function, analogous to choosing $\exp(-B_{ij})$ in the case of the canonical operator. We find that $p = 16$ is satisfactory, in terms of both accuracy and efficiency, when accompanied with a slight adjustment to $\zeta = 1.028$.

In summary, we propose to use the following effective Coulomb operator

$$f_{ij}^{p16} = r_{ij} + \Omega_{ij}^{p16} \sqrt{R_iR_j} \quad (4.11)$$

with

$$\Omega_{ij}^{p16} = \left(1 + \frac{\zeta r_{ij}}{16\sqrt{R_iR_j}}\right)^{-16} \quad (4.12)$$

and $\zeta = 1.028$. In the numerical calculations reported in section V, we will also test another effective Coulomb operator, f_{ij}^{exp} , which is analogous to f_{ij}^{p16} except that it uses Ω_{ij}^{exp} (eq 4.9), with $\zeta = 1.021$, rather than Ω_{ij}^{p16} .

We note that f_{ij}^{p16} is essentially a simpler, more efficient version of the Still operator, and as such it does not necessarily account for “transverse” interactions that make $\Psi_{ij} > 1$ in some cases.³² (A value $\zeta < 1$ can make $\Psi_{ij} > 1$ for f_{ij}^{p16} , but we have not found this useful in practice so far. One could imagine choosing ζ to be some simple function that becomes less than unity when it detects “transverse” interactions, but we have not attempted to do this.) We do not expect f_{ij}^{p16} to rival the accuracy achieved by a more sophisticated alternative operator like that introduced recently by Onufriev and Sigalov.³² Nevertheless, our numerical results indicate that f_{ij}^{p16} is a noticeable improvement upon the canonical Still operator.

V. NUMERICAL TESTS

V.A. Computational Details. C-PCM calculations were carried out for all of the systems investigated here using a solvent dielectric constant, $\epsilon = 78.4$, that is characteristic of room-temperature water and a dielectric constant of unity inside of the solute cavity. The atomic spheres used to construct the solute cavity surface are chosen in the customary manner for PCM calculations,¹⁵ namely, by scaling atomic van der Waals radii⁵⁸ by a factor of 1.2. We then use these atomic spheres to construct Connolly’s “solvent-excluded” cavity surface⁵⁹ using a new algorithm that fits within the framework of the switching Gaussian (SWIG) surface discretization method that we developed.^{60,61} (The technical details involved in extending the SWIG algorithm to construction of a Connolly surface will be reported in a future publication, but the end result is a discretized version of the surface proposed by Connolly.⁵⁹) A solvent probe radius of 1.4 Å, corresponding to water, is used to construct the Connolly surface. The average grid resolution of the surface in all calculations is 0.07 Å² per surface element (14 grid points/Å²), and we expect discretization errors of <1 kcal/mol in G_{pol} for the C-PCM

Table 1. Error Statistics^a in the Total Solvation Energy, G_{pol} , Using Perfect Radii and Various Forms of f_{ij} for a Training Set of Small Proteins

	sphere operator ^b	Still operator ^c		new operators	
		$c = 4.0$	$c = 5.5$	exp ^d	p16 ^e
MSE ^f	8.97 (1.44)	2.76 (0.10)	5.45 (0.52)	2.79 (0.22)	1.10 (0.07)
SD ^g	8.61 (1.72)	9.06 (1.10)	7.14 (0.65)	6.32 (0.61)	5.59 (0.52)
RMSE ^h	12.43 (2.24)	9.47 (1.10)	8.98 (0.83)	6.91 (0.65)	5.70 (0.53)
MAE ⁱ	10.64 (1.52)	7.75 (0.91)	6.85 (0.66)	5.09 (0.52)	4.04 (0.41)
Max. AE ^j	21.04 (6.69)	20.33 (2.44)	21.63 (1.70)	18.02 (1.34)	15.62 (1.16)

^aErrors in kcal/mol, with % errors in parentheses. ^bEquation 4.1. ^cEquation 4.3. ^dEquations 4.7 and 4.9. ^eEquation 4.11. ^fMean signed error. ^gStandard deviation about the MSE. ^hRoot-mean-square error. ⁱMean absolute error. ^jMaximum absolute error.

Table 2. Error Statistics^a in the Pairwise Interaction Energies, $G_{\text{pol},ij}$ (eq 2.3), Using Perfect Radii and Various forms of f_{ij} for a Training Set of Small Proteins

	sphere operator	Still operator		new operators	
		$c = 4.0$	$c = 5.5$	exp	p16
MSE $\times 10^{-6}$	30.7 (-0.077)	9.4 (0.085)	18.7 (0.064)	9.5 (0.051)	3.8 (0.037)
SD	0.14 (2.73)	0.11 (2.71)	0.085 (2.16)	0.080 (1.95)	0.077 (1.77)
RMSE	0.14 (2.73)	0.11 (2.71)	0.085 (2.16)	0.080 (1.95)	0.077 (1.77)
MAE	0.045 (2.17)	0.032 (1.63)	0.025 (1.28)	0.022 (1.13)	0.021 (1.02)
max AE	6.04 (14.46)	4.66 (21.66)	4.03 (19.64)	3.85 (19.02)	3.71 (18.12)
no. of gross errors ^b	9620	4991	2479	2128	2056

^aErrors in kcal/mol with % errors in parentheses for 4 677 283 interaction pairs. Other nomenclature is the same as in Table 1. ^bNumber of errors larger than 1.2 kcal/mol.

calculations.⁶¹ Atomic point charges for the solute are taken from the AMBER99 force field.⁶²

Perfect effective Born radii are obtained from the C-PCM calculations following eq 2.5, wherein a separate C-PCM calculation is performed for each individual atom of the solute. Because we use C-PCM, these perfect Born radii are in fact the R_i^∞ of eq 3.18. We also use C-PCM to compute the pairwise interactions, $G_{\text{pol},ij}^{\text{C-PCM}}$, for all atom pairs in every system. By virtue of the equivalence between C-PCM and the conventional GB model, this is the same as using eq 3.15. The C-PCM matrix equations are solved via direct matrix inversion for the 16 small protein training set (section V.B) to provide highly precise benchmark data. For the larger biomolecule set (section V.D), matrix inversion is no longer feasible given our computational resources and we instead employ an efficient, low-memory conjugate gradient solver with a tight convergence criterion. For the data set of 16 small proteins discussed in section V.B, the difference between the conjugate gradient and direct inversion results is -0.002 ± 0.006 kcal/mol. All calculations are performed using a locally modified version of the Q-CHEM software.⁶³

V.B. Small Protein Training Set. The GB model is most commonly used in simulations of biochemical macromolecules, so in order to provide a relevant gauge of performance, we investigate the accuracy and efficiency of various GB models for a set of 16 small proteins selected from the training set used by Feig et al.¹⁹ This data set, which is listed in Appendix C, was downloaded from the Protein Data Bank (PDB, www.pdb.org)⁶⁴ and set up for calculation using Tinker, v. 4.2.⁶⁵ The proteins range from 515–997 atoms (37–69 residues) and were selected mainly for being feasible in the context of the C-PCM calculations performed here.

We first consider errors

$$\Delta G_{\text{pol}} = G_{\text{pol}}^{\text{C-PCM}} - G_{\text{pol}}^{\text{GB}} \quad (5.1)$$

in the total electrostatic solvation energy, obtained using perfect effective Born radii. As discussed in section III.E, these errors are attributable solely to the effective Coulomb operator and we test five different operators, including f_{ij}^{sphere} (eq 4.1) and also f_{ij}^{still} (eq 4.3), the latter with both $c = 4.0$ (the most widely used value of this parameter) and $c = 5.5$, which was optimized to minimize the percentage root-mean-square error (RMSE) for this particular training set. (The optimal value that we obtain is close to the value $c = 5.249$ recommended by Grant et al.²⁷) Finally, we test the two new operators proposed in section IV.B, namely, f_{ij}^{exp} as defined by eqs 4.7 and 4.9, and f_{ij}^{p16} as defined in eq 4.11. These two operators differ only in the numerical implementation of the exponential function that defines Ω_{ij}^{exp} , and the ζ parameters that were reported in section IV.B have been optimized to minimize the percentage RMSE for this training set.

A statistical summary of the errors in total solvation energies (ΔG_{pol}) is presented in Table 1, and in Table 2 we present a statistical analysis of errors in $G_{\text{pol},ij}$ for all atom pairs (i,j) in the protein training set. Perhaps not surprisingly, the operator f_{ij}^{sphere} is the least accurate among the interaction operators that we consider. More interesting is the fact that f_{ij}^{p16} improves upon f_{ij}^{still} for both the total polarization energy and the pairwise polarization energies. Optimizing the value of c in f_{ij}^{still} to minimize the RMSE in G_{pol} also has the effect of reducing the mean absolute error (MAE) relative to the canonical choice $c = 4.0$, but this change does bias the mean signed error (MSE) toward underestimation of the polarization energy, that is, optimization makes $G_{\text{pol}}^{\text{GB}}$ more negative on average. Looking at the pairwise interactions, however, the choice $c = 5.5$ is clearly superior to $c = 4.0$.

Most significantly, f_{ij}^{p16} reduces the RMSE and MAE of the total polarization energy by 3.28 and 2.81 kcal/mol, respectively, relative to f_{ij}^{still} with $c = 5.5$. In terms of percentages, this is a reduction in the statistical error by 0.30% and 0.25%, respectively, an appreciable improvement

given that f_{ij}^{still} with $c = 5.5$ has an RMSE of only 0.83% and a MAE of only 0.66%. On an individual basis, f_{ij}^{p16} is not always the best operator, but on a statistical basis it is definitely the best choice for this particular data set.

We also track the number of “gross” errors in the pairwise energies. Following Onufriev and Sigalov,³² we define a gross error to be one for which $|\Delta G_{\text{pol},ij}| > 2k_{\text{B}}T$ and set $2k_{\text{B}}T = 1.2$ kcal/mol here, corresponding to $T = 300$ K. From Table 2 we see that f_{ij}^{p16} has the fewest gross errors, more than halving the number of gross errors obtained using the canonical Still operator with $c = 4.0$. It is nevertheless remarkable that of all the 4 677 283 pairwise interactions in the data set, f_{ij}^{still} with $c = 4.0$ manages to place 99.9% of them within the prescribed gross error threshold.

In addition to perfect radii, C-PCM calculations provide the means to compute “perfect” values of Ψ_{ij} . To do this we use eq 4.5 to obtain Ψ_{ij} taking f_{ij} to be whatever value is necessary to reproduce the value of $G_{\text{pol},ij}$ obtained from a C-PCM calculation. The resulting “perfect” value of Ψ_{ij} which we denote $\Psi_{ij}^{\text{C-PCM}}$, is therefore different for each atom pair and defined only for a single value of r_{ij} . We plot these values in Figure 1 for one particular protein in the data set; other proteins exhibit similar trends.

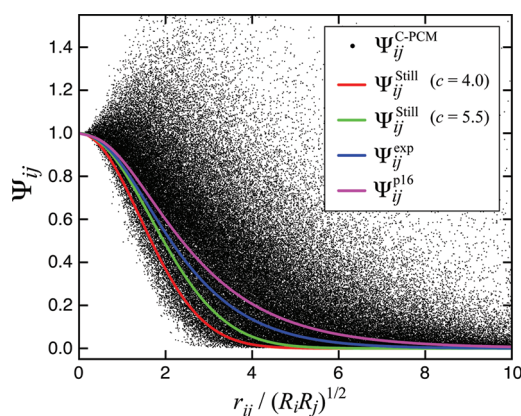


Figure 1. Values of $\Psi_{ij}^{\text{C-PCM}}$ for the protein 1AJJ, along with analytic interpolation functions corresponding to the effective Coulomb operators f_{ij}^{still} , f_{ij}^{exp} , and f_{ij}^{p16} . For f_{ij}^{sphere} , the corresponding interpolation function is constant ($\Psi_{ij}^{\text{sphere}} \equiv 1$) and is not plotted.

The C-PCM data confirm (as has been shown previously²⁰) that the interpolating function Ψ_{ij} is not a simple, single-valued function of pairwise distance and/or effective Born radii. As such, by choosing some analytic function $\Psi_{ij}(r_{ij})$ to define a GB model, one may hope, at best, to obtain a reasonable average over what is essentially a structured scatter plot of $\Psi_{ij}^{\text{C-PCM}}$ values. In particular, we note that there are numerous atomic pairs for which $\Psi_{ij}^{\text{C-PCM}} > 1$, behavior that Onufriev and Sigalov³² attributed to “transverse” pairwise interactions. These data points are outside the range of the various empirical choices for Ψ_{ij} that are considered here.

Note also from Figure 1 the absence of data points for which $\Psi_{ij}^{\text{C-PCM}} < 0$. Negative values of Ψ_{ij} imply an unphysical Coulomb interaction wherein $|G_{0,ij}| < |G_{\text{pol},ij}|$ for $i \neq j$; hence, “exact” calculations should never exhibit $\Psi_{ij}^{\text{C-PCM}} < 0$. In practice, we find that C-PCM calculations can exhibit unphysical values $\Psi_{ij}^{\text{C-PCM}} < 0$ unless the C-PCM linear equations are solved to very high numerical accuracy. These unphysical values typically correspond to larger values of r_{ij} and

therefore make little difference to the total solvation energy (total energy differences of ~ 0.002 kcal/mol, depending on numerical thresholds), but they do paint an unrealistic picture of the Ψ_{ij} distribution. For this reason, all of the C-PCM calculations for the protein training set (and thus all of the C-PCM data in Figure 1) solve the C-PCM equations by matrix inversion rather than via a conjugate gradient algorithm.

Also plotted in Figure 1 are several different choices for the analytic interpolation function $\Psi_{ij}(r_{ij})$ corresponding to various analytic functions f_{ij} . These plots reveal why GB models work at all: the interpolating functions Ψ_{ij} clearly pick up the structure of the scatter plot of exact $\Psi_{ij}^{\text{C-PCM}}$ values, in an average way, while achieving the correct limits as $r_{ij} \rightarrow 0$ and $r_{ij} \rightarrow \infty$. Moreover, the differences between various choices for Ψ_{ij} help to explain the relative accuracy of the corresponding GB models.

When the quantity $r_{ij}/(R_i R_j)^{1/2}$ is small, all of the analytic interpolation functions are in rough agreement with $\Psi_{ij}^{\text{C-PCM}}$ values and each achieves the correct limit of $\Psi_{ij} = 1$ for $r_{ij} = 0$. (Note that $\Psi_{ij}^{\text{sphere}} \equiv 1$.) We see that Ψ_{ij}^{still} with $c = 4.0$ is too short ranged as compared to the $\Psi_{ij}^{\text{C-PCM}}$ data, but the choice $c = 5.5$ increases the extent of Ψ_{ij}^{still} and improves its agreement with $\Psi_{ij}^{\text{C-PCM}}$. However, the functions Ψ_{ij}^{exp} and Ψ_{ij}^{p16} are even longer ranged, and on average, these functions provide a better fit to the $\Psi_{ij}^{\text{C-PCM}}$ data for the middle-range values of $r_{ij}/(R_i R_j)^{1/2}$. This explains why these new interpolation functions improve the energy statistics for the protein data set. The corollary of this observation is that for systems in which the “perfect” Ψ_{ij} might be predominantly short-ranged, Ψ_{ij}^{still} could be more accurate than Ψ_{ij}^{exp} or Ψ_{ij}^{p16} . However, our tests here on small proteins and on the larger biomolecules in section V.D seem to indicate that these macromolecular systems tend to have more long-ranged pairwise interactions, again leading to the observed improvement in accuracy when Ψ_{ij}^{exp} or Ψ_{ij}^{p16} is used in place of Ψ_{ij}^{still} .

V.C. R6 Effective Born Radii. In order to examine larger macromolecules, we must abandon perfect effective Born radii because they become too cumbersome to compute. Instead, we turn to the integral-based methodology sometimes referred to as “R6 radii” (for reasons that will become apparent).

Grycuk²⁴ has shown that for the case of a spherical cavity embedded in a conductor, the following integral yields the exact polarization energy for a point charge located at an arbitrary point \vec{r}_i inside of the sphere:

$$G_{\text{pol},ii}^{\text{Grycuk}} = -\frac{q_i^2}{2} \left(\frac{3}{4\pi} \int_{\text{ext}} \frac{d\vec{r}}{|\vec{r} - \vec{r}_i|^6} \right)^{1/3} \quad (5.2)$$

The integration region is the volume exterior to the cavity, but this can be transformed into a surface integral using Gauss’ theorem, resulting in the following equivalent expression:

$$G_{\text{pol},ii}^{\text{Grycuk}} = -\frac{q_i^2}{2} \left(\frac{1}{4\pi} \int d\vec{s} \frac{(\vec{s} - \vec{r}_i) \cdot \vec{n}_{\vec{s}}}{|\vec{s} - \vec{r}_i|^6} \right)^{1/3} \quad (5.3)$$

Here, the integral runs over the spherical cavity surface and $\vec{n}_{\vec{s}}$ is the outward-pointing vector normal to the cavity surface at the point \vec{s} . These integrals immediately afford the perfect effective Born radii:

$$R_{\text{R6},i} = \left(\frac{1}{4\pi} \int d\vec{s} \frac{(\vec{s} - \vec{r}_i) \cdot \vec{n}_{\vec{s}}}{|\vec{s} - \vec{r}_i|^6} \right)^{-1/3} \quad (5.4)$$

Numerically, one can compute R6 radii via a discretized version of eq 5.4,

$$R_{R6,i} = \left(\frac{1}{4\pi} \sum_k \frac{(\vec{s}_k - \vec{r}_i) \cdot \vec{n}_k a_k}{|\vec{s}_k - \vec{r}_i|^6} \right)^{-1/3} \quad (5.5)$$

Here, \vec{r}_i is the location of the i th atom and the summation runs over all surface grid points \vec{s}_k with outward-pointing surface normals \vec{n}_k and surface areas a_k . By using eq 5.5 in our GB calculations hereafter, we are essentially comparing C-PCM to a version of the surface GB (S-GB) method.³⁷

Equation 5.4 is exact for a spherical cavity in a conductor; hence, $R_{R6,i} = R_i^\infty$ in this case. For nonspherical cavities, the R6 integral does not yield the exact R_i^∞ , although the inverse R6 radii are found to be quite close to $1/R_i^\infty$ apart from a roughly constant offset.²⁸ As such, Mongan et al.²⁸ propose augmenting the R6 radii to shift them closer to perfect radii on average. Thus, they propose to use so-called “R6* radii” defined by

$$\frac{1}{R_{R6^*,i}} = \frac{1}{R_{R6,i}} + a \quad (5.6)$$

where a is an empirical constant. While acknowledging that the optimal value of a likely depends on the choice of cavity definition, Mongan et al.²⁸ suggest using the value $a = 0.028 \text{ \AA}^{-1}$.

Returning now to the testing of effective Coulomb operators, we investigate the effect of using R6 radii computed via eq 5.5, where the cavity surface is the same Connolly surface that is used in the C-PCM calculations. Because the same surface integration grid is used to compute both the R6 radii and the C-PCM energies, we eliminate any possible error arising from incongruities in the discretization, which might be present, for example, if one were to compute R6 radii with a volumetric integration grid and then compare to perfect radii derived from a boundary-element method such as C-PCM. For the protein training set and cavity surface definition considered here, we find that a shift of $a = 0.011 \text{ \AA}^{-1}$ best reproduces the perfect radii (see Figure 2).

Energy statistics are reported in Table 3 for the small protein training set using the R6* radii. Essentially the same trends are observed as in the case of the perfect effective Born radii (Table 1), confirming that the augmentation in eq 5.6 is a rather robust correction. The R6* radii seem to incur only a small but constant error for $f_{ij}^{\text{still}} (c = 5.5)$, f_{ij}^{exp} and f_{ij}^{p16} , raising the RMSE

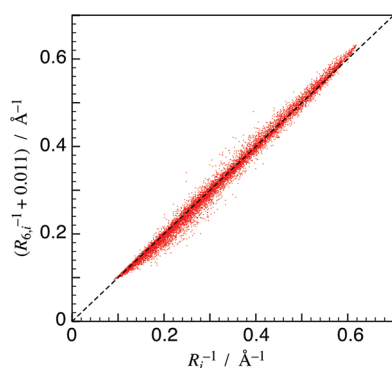


Figure 2. Inverse augmented R6* radii (eq 5.6 with $a = 0.011 \text{ \AA}^{-1}$) versus perfect effective Born radii obtained from C-PCM calculations for all 12 407 atoms of the small protein training set. The RMS deviation is 0.012 \AA^{-1} .

by about 0.1% as compared to results obtained with perfect radii.

V.D. Larger Biomolecules. Three of the five effective Coulomb operators considered here have parameters that are optimized to minimize errors in the protein training set introduced above. Next, we consider a data set for which we have not performed any parameter optimization. Specifically, we select a second data set consisting of 54 larger biomolecules (ranging from 758–6186 atoms) and compute ΔG_{pol} using the R6* radii optimized for the training set. This new data set, which is listed in Appendix C, is composed of 41 randomly selected proteins from a study by Feig et al.¹⁹ along with the 12 protein structures and the one B-DNA structure considered by Onufriev and Sigalov.³² The data set includes a variety of structures and charge distributions and serves as a blind test of the accuracy of f_{ij}^{exp} and f_{ij}^{p16} .

Energy statistics are reported in Table 4, and overall the trends are fairly similar to what was observed for the small protein training set. The operator f_{ij}^{sphere} is still the worst for accuracy, and f_{ij}^{still} with $c = 5.5$ offers a small improvement over the corresponding operator with $c = 4.0$. The new operators f_{ij}^{exp} and f_{ij}^{p16} are statistically better than f_{ij}^{still} with $c = 5.5$, offering up to $\sim 0.2\%$ reduction in the RMSE and MAE. Further tests will be necessary to evaluate the extent to which these new operators improve the accuracy of GB calculations in different molecular systems as well as in applications aside from computing $G_{\text{pol}}^{\text{GB}}$, but so far these new operators appear to be promising alternatives to the canonical operator of Still et al.¹⁶

V.E. Timings. The most time-consuming step in a GB calculation is typically evaluation of effective Born radii. Nonetheless, evaluating the pairwise GB energies scales quadratically with the number of atoms and can eventually become a significant computational expense. A more efficient choice of f_{ij} should certainly be a welcome reduction in CPU time for GB simulations, provided that the accuracy does not suffer.

Here, we consider the CPU time required by the various effective Coulomb operators. We measure the CPU time to compute $G_{\text{pol}}^{\text{GB}}$ for the largest molecule in either of our test sets (PDB code 1DBF), provided the augmented R6* radii defined in eq 5.6 are given as input, that is, timings reported here do not include the time required to perform the numerical surface integration to obtain R6* radii. Calculations on the other proteins display a similar trend. The calculations are run in serial on a single 2.5 GHz Opteron core. Timings reported here are those computed within our development version of Q-CHEM,⁶³ but we also confirmed that the same trends can be observed using a stand-alone program that eliminates any possible overhead introduced by the Q-CHEM program. The pairwise loops for each operator are coded as similarly as possible to facilitate a fair comparison. Charges, effective Born radii, and atomic positions are stored in core memory, but we also compare the CPU time required to compute r_{ij} (or r_{ij}^2 for f_{ij}^{still} and f_{ij}^{sphere}) versus having r_{ij} stored in memory. The small overhead associated with computing $R_i^{1/2}$ is included in the timings for f_{ij}^{exp} and f_{ij}^{p16} . For efficiency in evaluating Ω_{ij}^{p16} (eq 4.12), we first compute the quantity

$$x = \frac{(R_i R_j)^{1/2}}{(R_i R_j)^{1/2} + \zeta r_{ij}/16} \quad (5.7)$$

and then successively self-multiply it four times. The product $(R_i R_j)^{1/2}$ need only be computed once per atom pair, and the

Table 3. Error Statistics^a in the Total Solvation Energy, G_{pol} , Using R6* Radii^b and Various Forms of f_{ij} for a Training Set of Small Proteins

	sphere operator	Still operator		new operators	
		$c = 4.0$	$c = 5.5$	exp	p16
MSE	7.63 (1.31)	2.43 (0.10)	4.82 (0.48)	2.03 (0.17)	0.26 (0.01)
SD	9.01 (1.77)	10.72 (1.15)	8.80 (0.78)	8.00 (0.71)	7.21 (0.62)
RMSE	11.81 (2.20)	11.00 (1.16)	10.04 (0.91)	8.25 (0.73)	7.22 (0.62)
MAE	10.14 (1.43)	8.73 (0.96)	7.67 (0.75)	6.05 (0.57)	5.06 (0.46)
Max. AE	19.48 (7.10)	22.14 (2.03)	21.91 (1.63)	18.45 (1.65)	18.65 (1.67)

^aErrors in kcal/mol with % errors in parentheses. Other nomenclature is the same as in Table 1. ^bRadii defined by eq 5.6 with $a = 0.011 \text{ \AA}^{-1}$.

Table 4. Error Statistics^a in the Total Solvation Energy, G_{pol} , for a Set of Larger Biomolecules Using R6* Radii Optimized for the Protein Training Set^b

	sphere operator	Still operator		new operators	
		$c = 4.0$	$c = 5.5$	exp	p16
MSE	13.74 (1.43)	4.69 (0.19)	9.29 (0.60)	2.86 (0.27)	-1.40 (0.09)
SD	24.13 (1.58)	12.16 (1.04)	10.08 (0.74)	9.91 (0.76)	10.85 (0.76)
RMSE	27.76 (2.13)	13.03 (1.05)	13.71 (0.95)	10.32 (0.81)	10.94 (0.76)
MAE	21.64 (1.60)	10.48 (0.76)	10.86 (0.74)	8.62 (0.61)	8.25 (0.57)
Max. AE	78.50 (8.49)	30.93 (4.48)	33.23 (2.54)	20.47 (2.97)	31.84 (2.85)

^aErrors in kcal/mol with % errors in parentheses. Other nomenclature is the same as in Table 1. ^bRadii defined by eq 5.6 with $a = 0.011 \text{ \AA}^{-1}$.

constant $\zeta/16$ can be held in memory. As such, we can compute Ω_{ij}^{p16} using just eight floating point operations, of which only one is a division operation.

Timings are presented in Figure 3. Of the operators considered here, the canonical Still operator, f_{ij}^{Still} , is clearly

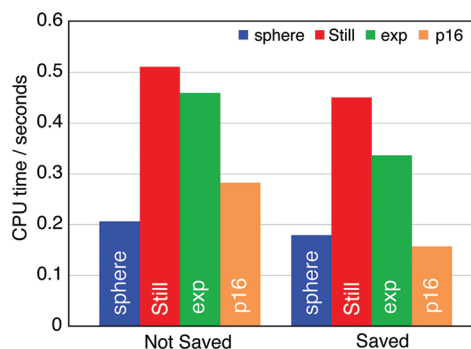


Figure 3. CPU timings for computing $G_{\text{pol}}^{\text{GB}}$ for the protein 1DBF for various choices of f_{ij} . Data are shown for saved r_{ij} (right) as well as unsaved r_{ij} (left).

the most expensive to compute with or without the availability of pairwise distances. Without saved distances, f_{ij}^{sphere} is the least expensive but is also the least accurate in our tests. The operator f_{ij}^{p16} is more expensive than f_{ij}^{sphere} when the r_{ij} are not saved, because computing r_{ij} from $r_{ij}^2 = \vec{r}_i \cdot \vec{r}_j$ requires a call to the square root function. However, since f_{ij}^{p16} is more accurate than f_{ij}^{sphere} (for general nonspherical cavity shapes), this is a small price to pay for the significantly improved accuracy of f_{ij}^{p16} . On the other hand, when the r_{ij} are stored, the function f_{ij}^{p16} is actually slightly faster to evaluate as compared to f_{ij}^{sphere} . More importantly, f_{ij}^{p16} achieves a speedup of a factor of 3 relative to f_{ij}^{Still} , owing to the fact that f_{ij}^{p16} does not call the square root function inside of the pairwise loop. Since f_{ij}^{p16} is also the most accurate interaction operator considered here, it is a very attractive choice for GB calculations.

VI. CONCLUSIONS

We made a formal comparison of the GB model (in both its conventional and “GB ϵ ” forms^{30,31}) to PCM theory¹⁵ and, specifically, to the conductor-like PCM, known in the literature as either C-PCM or GCOSMO.^{33,35} In addition, we investigated some alternatives to the “canonical” effective Coulomb operator¹⁶ that is widely used in GB calculations.

The following list summarizes our principal results.

- (1) The conventional GB model is equivalent to C-PCM insofar as one employs an effective Coulomb operator and “perfect” effective Born radii that are both exact in the conductor limit. (This is the only choice that is asymptotically correct for high-dielectric solvents.) The accuracy of conventional GB calculations is therefore limited by the accuracy of C-PCM. The latter provides an approximation (albeit a fairly accurate one in high-dielectric solvents) to the electrostatic solvation energy that would be computed from Poisson’s equation for the same solute cavity and dielectric constant.
- (2) As found in previous studies,²⁸ effective radii obtained from “R6” integrals are quite close to perfect effective Born radii obtained in the conductor limit. As such, the accuracy of conventional GB calculations with R6 radii approaches the accuracy of C-PCM calculations.
- (3) We propose a simplified effective Coulomb operator for conventional GB calculations. For a set of small proteins, where perfect effective Born radii can be computed using C-PCM, this alternative operator reduces the error in the electrostatic solvation energy by roughly 0.3% relative to the canonical operator introduced by Still et al.,¹⁶ which exhibits a statistical error of about 1.0%. For a set of larger biomolecules, the alternative operator continues to perform better than the canonical operator when augmented R6 radii are employed. Furthermore, the alternative operator is significantly less expensive to compute as compared to the Still et al. operator, by nearly a factor of 3.

It is our hope that these results will help to guide the future development of the GB model as well as enhance its efficiency and accuracy.

APPENDIX A

PCM Integral Operators

Here, we briefly define the PCM integral operators encountered in this work. More thorough discussions can be found elsewhere.^{15,42,66}

The electrostatic potential produced by a solute charge density, ρ , at an arbitrary point \vec{r} , is

$$\phi(\vec{r}) = \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3\vec{r}' \quad (\text{A1})$$

If ρ consists of a single point charge, q_i , located at position \vec{r}_i , then the potential is

$$\phi_i(\vec{r}) = \frac{q_i}{|\vec{r}_i - \vec{r}|} \quad (\text{A2})$$

The apparent surface charge density in PCM theory is denoted $\sigma(\vec{s})$, where \vec{s} represents a point on the solute cavity surface. The reaction-field potential generated by $\sigma(\vec{s})$ at any point in space is

$$\chi(\vec{r}) = \int \frac{\sigma(\vec{s})}{|\vec{r} - \vec{s}|} d^2\vec{s} \quad (\text{A3})$$

For a point on the cavity surface, the self-adjoint integral operator \hat{S} is defined by the equation

$$\hat{S}\sigma(\vec{s}) = \chi(\vec{s}) \quad (\text{A4})$$

The outward-pointing normal component of the electric field, arising from the apparent surface charge σ and evaluated at the point \vec{s} on the cavity surface, is equal to the directional derivative $\partial_{\vec{s}}\chi(\vec{s})$. This is used to define the adjoint of another integral operator, \hat{D} :

$$\hat{D}^\dagger\sigma(\vec{s}) = \partial_{\vec{s}}\chi(\vec{s}) \quad (\text{A5})$$

The integral operators \hat{D} and \hat{S} obey the identity $\hat{D}\hat{S} = \hat{S}\hat{D}^\dagger$, but this identity typically does not hold once the integral operators are discretized.⁴¹ For convenience, we also define

$$\hat{Y} = \hat{I} - \frac{1}{2\pi}\hat{D} \quad (\text{A6})$$

where \hat{I} is the identity operator. The operator \hat{Y} appears in the definition of \hat{C}_e in eq 3.12.

The form of the response operator \hat{Q} in eq 3.4 depends upon the particular flavor of PCM that is used. Response operators for each of the PCMs discussed in this work are listed in Table 5. Note that the C-PCM response operator lacks the term

Table 5. Response Operators for Various PCMs

model	response operator, \hat{Q}
conductor	$-\hat{S}^{-1}$
C-PCM/GCOSMO ^a	$-k_e^{-1}\hat{S}^{-1}$
COSMO	$-[(\epsilon - 1)/(\epsilon + 0.5)]\hat{S}^{-1}$
IEF-PCM/SS(V)PE ^{a,b}	$-k_e^{-1}\hat{C}_e^{-1}$

^aThe factor $k_e = \epsilon/(\epsilon - 1)$. ^bThe operator \hat{C}_e is defined in eq 3.12.

proportional to ϵ^{-1} that is present in the IEF-PCM/SS(V)PE operator; this term is required for an exact result but is quite small for $\epsilon \approx 80$.⁴¹ We also note that each of these PCMs is exact in the conductor limit, $\epsilon \rightarrow \infty$. If the surface charge computed in the conductor limit is scaled by a factor of $(\epsilon - 1)/\epsilon$, the result is precisely the working equation for the C-PCM method. This is the sense in which C-PCM is “conductor-like”. (However, we recently presented a more rigorous and satisfying derivation of this model, which does not require ad hoc scaling of the surface charge.⁵⁵)

APPENDIX B

Comparison to the Generalized Kirkwood Model

In section III we derived several formal connections between PCMs and GB models for solute charge densities composed of point charges only. Here, we illustrate that it is possible to extend the pairwise-additive approximation for the electrostatic solvation energy to higher-order multipoles, as in the generalized Kirkwood model.^{44,45}

Suppose that the MM solute consists of a set of atom-centered multipoles. We can partition the electrostatic potential at the cavity surface, $\phi(\vec{s})$, into contributions from each multipole order (cf. eq 3.6)

$$\phi(\vec{s}) = \sum_i \sum_{l=0}^{\infty} \phi_i^{(l)}(\vec{s}) \quad (\text{B1})$$

Here, $\phi_i^{(l)}(\vec{s})$ is the electrostatic potential arising from the l th-order multipole centered on atom i . The surface charge density can be partitioned in a similar way,

$$\sigma(\vec{s}) = \sum_i \sum_{l=0}^{\infty} \sigma_i^{(l)}(\vec{s}) \quad (\text{B2})$$

where $\sigma_i^{(l)} = \hat{Q}\phi_i^{(l)}$ and \hat{Q} is one of the response operators defined in Table 5.

Substitution of the multipole expansions of $\phi(\vec{s})$ and $\sigma(\vec{s})$ into eq 3.3 affords an analytic expression for the electrostatic solvation energy in PCM theory,

$$G_{\text{pol}}^{\text{PCM}} = \frac{1}{2} \sum_{i,j} \sum_{l,m} \int d\vec{s} \sigma_i^{(l)}(\vec{s}) \phi_j^{(m)}(\vec{s}) \quad (\text{B3})$$

Next, define

$$G_{\text{pol},ij}^{\text{PCM}} = \frac{1}{2} \sum_{l,m} \int d\vec{s} \sigma_i^{(l)}(\vec{s}) \phi_j^{(m)}(\vec{s}) \quad (\text{B4})$$

To obtain a GB-like approximation, one must find appropriate analytical expressions for the self-energies, $G_{\text{pol},ij}^{\text{PCM}}$, and pairwise-additive energies, $G_{\text{pol},ij}^{\text{PCM}}$. The Kirkwood model⁴⁶ of an arbitrary multipole in a spherical cavity accomplishes this for the self-energies, just as the Born ion model (which is the $l = 0$ case of the Kirkwood model) does in traditional GB theory. Then, appropriate choices for the multipole–multipole effective Coulomb operator must be made for the pairwise-additive terms, which is the topic of a recent study.⁴⁵

APPENDIX C

Biomolecule Data Sets

The Protein Data Bank (PDB) abbreviations of the biomolecules used in this work are listed below. Molecules were prepared with the Tinker software,⁶⁵ v. 4.2. Any water

molecules present in the PDB structures were removed. Where multiple conformers are available in the PDB, the structure "A" or "1" was selected.

1. *Small Protein Training Set*. 1AJJ, 1BBL, 1BOR, 1BPI, 1CBN, 1FCA, 1FXD, 1HPT, 1MBG, 1PTQ, 1R69, 1SH1, 1UXC, 1VII, 1VJW, 2ERL.

2. *Large Biomolecule Data Set*. 1AB3, 1AB7, 1APC, 1AZ6, 1BDD, 1BH4, 1BJ8, 1BKU, 1BMX, 1BNO, 1BRV, 1BY1, 1BYY, 1BZG, 1CG7, 1CMR, 1CN2, 1D7Q, 1DBF, 1DMC, 1EPO, 1ERD, 1EXG, 1FVL, 1FYC, 1G2S, 1GIO, 1HP8, 1IHV, 1LXL, 1MFN, 1MUN, 1NLS, 1PCP, 1PMS, 1QDP, 1QQF, 1RCH, 1ROT, 1SGG, 1TLE, 1VII, 2BNA, 2EZK, 2HMX, 2JV3, 2LZT, 2MB5, 2SOB, 2TRX, 2VIK, 3VUB, 4ULL, 7A3H.

AUTHOR INFORMATION

Corresponding Author

*E-mail: herbert@chemistry.ohio-state.edu.

Present Address

†Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, IL 60439, USA.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by an NSF CAREER award (CHE-0748448). Calculations were performed at the Ohio Supercomputer Center (project no. PAS-0291). J.M.H. is an Alfred P. Sloan Foundation fellow and a Camille Dreyfus Teacher-Scholar. A.W.L. acknowledges a Presidential Fellowship from The Ohio State University.

REFERENCES

- Warshel, A.; Åqvist, J. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 267–298.
- Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144–1149.
- Feig, M.; Brooks, C. L., III *Curr. Opin. Struct. Biol.* **2004**, *14*, 217–224.
- Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *Biochim. Biophys. Acta* **2006**, *1764*, 1647–1676.
- Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161–2200.
- Baker, N. A. *Method. Enzymol.* **2004**, *383*, 94–118.
- Baker, N. A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 137–143.
- Grochowski, P.; Trylska, J. *Biopolymers* **2008**, *89*, 93–113.
- Lu, B. Z.; Zhou, Y. C.; Holst, M. J.; McCammon, J. A. *Commun. Comput. Phys.* **2008**, *3*, 973–1009.
- Wang, J.; Tan, C.; Tan, Y.-H.; Lu, Q.; Luo, R. *Commun. Comput. Phys.* **2008**, *3*, 1010–1031.
- Onufriev, A. *Annu. Rep. Comput. Chem.* **2008**, *4*, 125–137.
- Onufriev, A. In *Modeling Solvent Environments: Applications to Simulations of Biomolecules*; Feig, M., Ed.; Wiley-VCH: Hoboken, NJ, 2010; Chapter 6, pp 127–165.
- Marchand, F.; Caflich, A. In *Modeling Solvent Environments: Applications to Simulations of Biomolecules*; Feig, M., Ed.; Wiley-VCH: Hoboken, NJ, 2010; Chapter 9, pp 209–232.
- Cramer, C. J.; Truhlar, D. G. In *Trends and Perspectives in Modern Computational Science*; Maroulis, G., Simos, T. E., Eds., Vol. 6 of *Lecture Series on Computer and Computational Sciences*; Brill/VSP: Leiden, 2006; pp 112–140.
- Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.
- Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J. *Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824–19839.
- Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–52.
- Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., III *J. Comput. Chem.* **2004**, *25*, 265–284.
- Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297–1304.
- Born, M. *Z. Phys.* **1920**, *1*, 45–48.
- Lee, M. S.; F. R. Salsbury, J.; Brooks, C. L., III *J. Chem. Phys.* **2002**, *116*, 10606–10614.
- Lee, M. S.; Feig, M.; F. R. Salsbury, J.; Brooks, C. L., III *J. Comput. Chem.* **2003**, *24*, 1348–1356.
- Grycuk, T. *J. Chem. Phys.* **2003**, *119*, 4817–4826.
- Romanov, A. N.; Jabin, S. N.; Martynov, Y. B.; Sulimov, A. V.; Grigoriev, F. V.; Sulimov, V. B. *J. Phys. Chem. A* **2004**, *108*, 9323–9327.
- Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. *J. Chem. Theory Comput.* **2007**, *3*, 156–169.
- Grant, J. A.; Pickup, B. T.; Sykes, M. J.; Kitchen, C. A.; Nicholls, A. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4913–4922.
- Mongan, J.; Svrcek-Seiler, W. A.; Onufriev, A. *J. Chem. Phys.* **2007**, *127*, 185101:1–10.
- Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465–1471.
- Sigalov, G.; Scheffel, P.; Onufriev, A. *J. Chem. Phys.* **2005**, *122*, 094511:1–15.
- Sigalov, G.; Fenley, A.; Onufriev, A. *J. Chem. Phys.* **2006**, *124*, 124902:1–14.
- Onufriev, A.; Sigalov, G. *J. Chem. Phys.* **2011**, *134*, 164104:1–14.
- Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
- Truong, T. N.; Stefanovich, E. V. *Chem. Phys. Lett.* **1995**, *240*, 253–260.
- Truong, T. N.; Nguyen, U. N.; Stefanovich, E. V. *Int. J. Quantum Chem. Symp.* **1996**, *30*, 1615–1622.
- Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486–1493.
- Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983–10990.
- Yu, Z.; Jacobson, M. P.; Friesner, R. A. *J. Comput. Chem.* **2005**, *27*, 72–89.
- Qiu, D.; Shenkin, P.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.
- Fenley, A. T.; Gordon, J. C.; Onufriev, A. *J. Chem. Phys.* **2008**, *129*, 075101:1–10.
- Lange, A. W.; Herbert, J. M. *Chem. Phys. Lett.* **2011**, *509*, 77–87.
- Chipman, D. M. *Theor. Chem. Acc.* **2002**, *107*, 80–89.
- So-called *generalized Kirkwood models* bear mention in this context.^{44,45} These models extend the pairwise-additive approximation in GB theory to arbitrary multipoles, in analogy to the Kirkwood model of a multipole in a spherical cavity.⁴⁶ Unlike PCMs, however, the cavity must be spherical in these generalized Kirkwood models.
- Kong, Y.; Ponder, J. W. *J. Chem. Phys.* **1997**, *107*, 481–492.
- Schnieders, M. J.; Ponder, J. W. *J. Chem. Theory Comput.* **2007**, *3*, 2083–2097.
- Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351–361.
- Mennucci, B.; Cancès, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506–10517.
- Chipman, D. M. *J. Chem. Phys.* **2000**, *112*, 5558–5565.
- Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.
- Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6532–6542.
- Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 877–887.
- Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760–768.
- Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- Marenich, A. V.; Cramer, C. J.; Truhlar, D. G.; Guido, C. A.; Mennucci, B.; Scalmani, G.; Frisch, M. J. *Chem. Sci.* **2011**, *2*, 2143–2161.

- (55) Lange, A. W.; Herbert, J. M. *J. Chem. Phys.* **2011**, *134*, 204110:1–15.
- (56) Lotan, I.; Head-Gordon, T. *J. Chem. Theory Comput.* **2006**, *2*, 541–555.
- (57) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 2447–2464.
- (58) Rowland, R. S.; Taylor, R. *J. Phys. Chem.* **1996**, *100*, 7384–7391.
- (59) Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.
- (60) Lange, A. W.; Herbert, J. M. *J. Phys. Chem. Lett.* **2010**, *1*, 556–561.
- (61) Lange, A. W.; Herbert, J. M. *J. Chem. Phys.* **2010**, *133*, 244111:1–18.
- (62) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (63) Shao, Y.; Fusti-Molnar, L.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P., Jr.; R. A., D.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L., III; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F., III; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.
- (64) 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*, 235–242.
- (65) TINKER, version 4.2; <http://dasher.wustl.edu/tinker>.
- (66) Chipman, D. M. *J. Chem. Phys.* **2004**, *120*, 5566–5575.