Published in final edited form as:

J Chem Theory Comput. 2007 November ; 3(6): 2083–2097. doi:10.1021/ct7001336.

Polarizable Atomic Multipole Solutes in a Generalized Kirkwood Continuum

Michael J. Schnieders[1] and **Jay W. Ponder**[2,*]

[1]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO 63130

[2]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110

Abstract

The generalized Born (GB) model of continuum electrostatics is an analytic approximation to the Poisson equation useful for predicting the electrostatic component of the solvation free energy for solutes ranging in size from small organic molecules to large macromolecular complexes. This work presents a new continuum electrostatics model based on Kirkwood's analytic result for the electrostatic component of the solvation free energy for a solute with arbitrary charge distribution.[1] Unlike GB, which is limited to monopoles, our generalized Kirkwood (GK) model can treat solute electrostatics represented by any combination of permanent and induced atomic multipole moments of arbitrary degree. Here we apply the GK model to the newly developed Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) force field, which includes permanent atomic multipoles through the quadrupole and treats polarization via induced dipoles. A derivation of the GK gradient is presented, which enables energy minimization or molecular dynamics of an AMOEBA solute within a GK continuum. For a series of 55 proteins, GK electrostatic solvation free energies are compared to the Polarizable Multipole Poisson-Boltzmann (PMPB) model[2] and yield a mean unsigned relative difference of 0.9%. Additionally, the reaction field of GK compares well to that of the PMPB model, as shown by a mean unsigned relative difference of 2.7% in predicting the total solvated dipole moment for each protein in this test set. The CPU time needed for GK relative to vacuum AMOEBA calculations is approximately a factor of three, making it suitable for applications that require significant sampling of configuration space.

1. INTRODUCTION

The solvent environment influences the structure and behavior of solutes within it. For example, the scaling of the radius of gyration of a polymer with chain length in dilute aqueous solution can be predicted by considering whether solvent molecules prefer interactions amongst themselves to those with the polymer.[3] This scaling law serves to

*Corresponding Author: Tel: (314) 362-4195, Fax: (314) 362-7183, ponder@dasher.wustl.edu.

SUPPORTING INFORMATION AVAILABLE

Intermediate terms in the derivation of the solvent field approximation are provided in Appendix A and Tables A-1 through A-4. Appendix B and Tables B-1 through B-10 provide a procedure for factoring the auxiliary tensors needed to compute the gradient of the GK potential. This material is available free of charge via the Internet at http://pubs.acs.org.

emphasize that rigorous results can be obtained without treating the solvent in explicit atomic detail. Here we present an analytic model of the electrostatic interactions between a solute represented by polarizable atomic multipoles and a continuum environment characterized by its permittivity, dispensing with the expense of representing explicit solvent molecules.

Our approach can be traced to work presented by Born in 1920 to describe the electrostatic solvation energy of a charged, spherical ion in terms of macroscopic continuum theory.[4] In 1934, Kirkwood extended this approach to a spherical particle with arbitrary electrostatic multipole moments with application to the study of zwitterions, which have a large dipole moment.[1] More recently, Kong and Ponder revisited Kirkwood's theory to allow analytic treatment of off-center point multipoles.[5] For a single spherical particle in isolation, therefore, the theoretical foundations to enable use of macroscopic continuum theory have already been established.

However, a general analytic solution to the Poisson equation for an arbitrarily spaced collection of spherical dielectric particles embedded in solvent is tenable only via approximations. For example, the generalization of Born's method to a collection of monopoles began to be considered in the 1990's by a number of groups including Schaefer *et al*.[6,7], Hawkins *et al*.[8,9], Still *et al*.[10,11], Feig *et al*.[12,13] and Onufriev *et al*.[14–18]. This generalized Born (GB) approach is intended to approximate the numerical solution of the Poisson equation for realistic molecular geometries and monopole charge distributions.[19–22] Given highly accurate self-energies, GB has been shown to be remarkably quantitative.[13,14,16,23,24] The goal of the present work is to extend the ideas underlying GB to more accurate charge distributions, specifically to the treatment of polarizable atomic multipoles, which might be termed generalized Kirkwood (GK) by analogy.

In order to further motivate the present work, we recall the electrostatic solvation energy is a key component of an implicit solvent model, which typically also includes apolar contributions due to cavitation and dispersion.[11,25,26] Given a solute potential and implicit solvent, a broad range of physical properties can be predicted, including conformational preferences such as radius of gyration, binding energies and pK_as.[24] Recent work by a number of groups to explicitly include higher order permanent moments and polarization within the functional form of empirical force field electrostatics may improve the quality of theoretical predictions based on implicit solvent approaches.[27–32] However, this step forward can only be realized if the improved detail of the molecular mechanics electrostatic model is propagated through to the reaction potential.

For an excellent introduction to the fundamentals of GB theory, including treatment of salt effects, we recommend the review by Bashford and Case.[33] Feig and Brooks present a review of recent improvements in GB methodology as well as novel applications.[12] Assuming this level of familiarity, we immediately outline the key components of GB that need to be further generalized in order to incorporate polarizable atomic multipoles.

1.1 Effective radii and the self-energy

Definition of the "perfect" effective radius a_i for site i under the GB approximation[16] guarantees an exact self-energy. It is based on the following equality

$$a_i = \frac{1}{2}\left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon_h}\right)\frac{{q_i}^2}{\Delta \mathrm{W}_{self,i}^{Poisson}} \quad (1)$$

where the factor of ½ accounts for the cost of polarizing the continuum, q_i is a partial charge, ε_h is the permittivity of a homogeneous reference state and ε_s is the permittivity of the solvent. The self-energy $\Delta \mathrm{W}_{self,i}^{Poisson}$ can be determined to high precision numerically. In this manner, the self-energy for each fixed partial charge of a solute is mapped onto the Born equation.[4] Alternatively, an analytic solution for the self-energy in terms of an energy density is possible after making the Coulomb field approximation

$$\Delta \mathrm{W}_{self,i}^{GB} = \frac{1}{2}\left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon_h}\right)\frac{q_i^2}{4\pi}\int_{solvent}\frac{1}{r^4}dV \quad (2)$$

although other methods will be elaborated below. Substituting for $\Delta \mathrm{W}_{self,i}^{Poisson}$ in Eq. (1) with $\Delta \mathrm{W}_{self,i}^{GB}$ from Eq. (2) and changing the limits of integration for convenience it can be shown that each effective Born radius is[33]

$$a_i = \left(\frac{1}{r_i} - \frac{1}{4\pi}\int_{solute, r>r_i}\frac{1}{r^4}dV\right)^{-1} \quad (3)$$

where the integration over the solute does not include the region within the atomic radius r_i. A number of analytic methods have been developed for determining this integral, notably the pairwise descreening method of Hawkins, Cramer and Truhlar that we will refer to as HCT[8,9], a method by Qiu *et al.* that assumes constant energy density within each descreening atom[11], and more recently a parameter free approach by Gallicchio *et al.*[34] Although effective radii determine the reaction potential at atomic centers, we note that the electrostatic solvation energy of a polarizable atomic multipole also depends on its higher order gradients.

After computing effective radii, the total self-energy of a solute within GB is

$$\Delta \mathrm{W}_{self}^{GB} = \frac{1}{2}\left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon_h}\right)\sum_i \frac{q_i^2}{a_i}. \quad (4)$$

For permanent multipoles, the self-energy of higher order components must be considered. Furthermore, if the solute is polarizable, self-consistent induced moments elicit a reaction potential that leads to an additional contribution to the electrostatic solvation free energy. We will avoid decomposing the polarization energy into self-energy and cross-term contributions, since it is inherently many-body and therefore any partitioning is somewhat artificial.

1.2 Cross-term energy

An analytic continuum electrostatics model designed to match results from the Poisson equation must also include an estimate of the pairwise cross-term energy between all multipole pairs. Given effective radii, the GB cross-term energy for fixed partial charges is given by

$$\Delta \mathrm{W}_{cross}^{GB} = \frac{1}{2}\left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon_h}\right)\sum_i \sum_{j \neq i} \frac{q_i q_j}{f} \quad (5)$$

where the empirical generalizing function f usually takes the form[10]

$$f = \sqrt{r_{ij} + a_i a_j e^{-r_{ij}/c_f a_i a_j}} \quad (6)$$

where r_{ij} is the distance between sites i and j and the tuning parameter c_f is chosen in the range 2–8. As r_{ij} goes to zero, the Born formula is recovered, such that the self-energy is simply a special case of the cross-term energy. Derivation of a general form for the pairwise cross-term energy between two multipole components will be presented, which is similar in spirit to GB in that the limiting cases of superimposition and wide separation for a pair of solvated multipoles are reproduced. The accuracy of the proposed interpolation at intermediate separations will be investigated via a series of tests ranging from simple systems consisting of only two sites up to the electrostatic solvation energy and dipole moment of proteins.

Our tests of GK rely on the Polarizable Multipole Poisson-Boltzmann (PMPB) model[2] as a standard of accuracy, which has been implemented for solutes described by the Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) force field.[35–37] In our previous work, excellent agreement was seen in the electrostatic response of proteins solvated by the PMPB continuum when compared to ensemble average explicit water simulations, indicating that at the length scale of proteins treatment of solvent as a continuum is valid. As an alternative to numerical PMPB electrostatics, the analytic GK formulation for the AMOEBA force field is orders of magnitude more efficient.

The description of GK will be subdivided into four sections. First, determination of the self-energy for a permanent multipole will be considered. Second, we will propose a functional form for the cross-term energy between arbitrary degree multipole moments. Third, given the underlying GK theory, we continue on to the derivation of the electrostatic solvation energy and gradient in the specific case of solutes described by the AMOEBA force field. Finally, we apply the GK continuum model to a series of proteins and compare their electrostatic solvation free energy and total dipole moment to analogous calculations with the PMPB continuum.

2 MULTIPOLE SELF-ENERGY

We begin by reiterating that the self-energy of a multipole depends on not only the reaction potential at atomic centers, but on the reaction field, the reaction field gradient, and so on. Unlike GB, the perfect effective radius is not enough information to guarantee the higher

order features of the reaction potential are correct, unless the multipole site happens to be at the center of a spherical cavity. Two methods have been investigated to describe the self-energy of a permanent atomic multipole. The first method reduces to the Coulomb-field approximation (CFA) for a monopole and requires knowledge of the analytic solution for the field in solvent based on a multipole at the center of a spherical dielectric cavity. We term this the solvent field approximation (SFA), as it is consistent with the CFA, but requires more information.[1] A second approach makes use of Grycuk's method for determining effective radii based on the reaction potential of an off-center charge within a spherical solute.[38] We refer to this approach as the reaction potential approximation (RPA).

Before detailing the SFA and RPA methods, a brief introduction to the electrostatic energy of a dielectric media will be given. The work required to assemble a fixed charge distribution in a linearly polarizable medium[33,39,40] can be formulated by a volume integral of the product of the charge density $\rho(r)$ with the potential $\varphi(r)$ or by the scalar product of the electric field **E** with the electric displacement **D**

$$W=\frac{1}{2}\int_V \rho(r)\phi(r)dV=\frac{1}{8\pi}\int_V \mathbf{E}\cdot\mathbf{D}dV \quad (7)$$

where the displacement is proportional to the electric field in regions of constant permittivity ε

$$\mathbf{D}=\varepsilon\mathbf{E}. \quad (8)$$

For our purposes, the system of interest is composed of a solute with a different permittivity than the solvent. The electrostatic free energy of this system relative to a homogeneous reference state is[33,39]

$$\Delta G=\frac{1}{8\pi}\int_V (\mathbf{E}\cdot\mathbf{D}-\mathbf{E}_h\cdot\mathbf{D}_h)dV \quad (9)$$

where in the homogeneous case the field is Coulombic and can be defined relative to the vacuum field as $\mathbf{E}_h = \mathbf{E}_{vac}/\varepsilon_h$ using the homogeneous permittivity ε_h. The homogeneous displacement is simply $\mathbf{D}_h = \mathbf{E}_{vac}$. A less intuitive, but equivalent definition of the electrostatic free energy given in Eq. (9) is[33,39]

$$\Delta G=\frac{1}{8\pi}\int_V (\mathbf{E}\cdot\mathbf{D}_h-\mathbf{D}\cdot\mathbf{E}_h)dV. \quad (10)$$

This expression can be subdivided into integrals over the solute and solvent volumes as

$$\Delta G=\frac{1}{8\pi}\int_{solute} (\mathbf{E}\cdot\mathbf{D}_h-\mathbf{D}\cdot\mathbf{E}_h)dV+\frac{1}{8\pi}\int_{solvent} (\mathbf{E}\cdot\mathbf{D}_h-\mathbf{D}\cdot\mathbf{E}_h)dV. \quad (11)$$

In both the homogeneous and mixed permittivity states the solute retains the homogeneous permittivity. By using the relationships for the homogeneous field and displacement described above it can be seen that the integral over the solute vanishes

$$\Delta G=\frac{1}{8\pi}\int_{solute}(\mathbf{E}\cdot\mathbf{E}_{vac}-\varepsilon_h\mathbf{E}\cdot\mathbf{E}_{vac}/\varepsilon_h)dV+\frac{1}{8\pi}\int_{solvent}(\mathbf{E}_s\cdot\mathbf{E}_{vac}-\varepsilon_s\mathbf{E}_s\cdot\mathbf{E}_{vac}/\varepsilon_h)dV \quad (12)$$

to leave only the integral over the solvent

$$\Delta G=\frac{1}{8\pi}\left(1-\frac{\varepsilon_s}{\varepsilon_h}\right)\int_{solvent}(\mathbf{E}_s\cdot\mathbf{E}_{vac})dV. \quad (13)$$

Having made no assumptions to this point, the remaining challenge can be simplified to defining the field within the solvent $\mathbf{E}_s$ for the mixed permittivity case. This is the starting point for the SFA. In general, the solvent field does not have an exact analytic form for a union of spheres. However, many molecular systems of interest are globular, and therefore an approximation based on the assumption of a spherical solute is not only qualitatively reasonable, but in many cases quantitative.

2.1 Solvent field approximation

The SFA is similar to the CFA, but is based on evaluating Eq. (13) using Kirkwood's solution for the field outside a spherical solute with a central multipole moment[1,41]

$$\mathbf{E}_s=\sum_{l=0}^{\infty}\frac{(2l+1)\varepsilon_h}{(l+1)\varepsilon_s+l\varepsilon_h}\mathbf{E}_{vac}^{(l)} \quad (14)$$

where $\mathbf{E}_{vac}^{(l)}$ is the vacuum field due to all multipole moments of degree *l*, defined using either irregular spherical harmonics or Cartesian tensors. Throughout the current work we neglect salt effects, although their addition to a future GK formulation is straightforward. This definition of the self-energy is equivalent to the CFA for a monopole and becomes approximate for off-center multipole sites or for non-spherical solute geometries.

Under the SFA, the self-energy of a permanent multipole site *i* is given by

$$\Delta G_i^{SFA}=\frac{1}{8\pi}\left(1-\frac{\varepsilon_s}{\varepsilon_h}\right)\int_{solvent}\mathbf{E}_{vac,i}\cdot\sum_{l=0}^{\infty}\left(\frac{(2l+1)\varepsilon_h}{(l+1)\varepsilon_s+l\varepsilon_h}\mathbf{E}_{vac,i}^{(l)}\right)dV. \quad (15)$$

It is possible to invert the integration domain by adding and subtracting an integral over the solute region outside the radius R_i of atom *i* to Eq. (15) giving

$$\begin{aligned}\Delta G_i^{SFA}=&\frac{1}{8\pi}\left(1-\frac{\varepsilon_s}{\varepsilon_h}\right)\int_{r>R_i}\mathbf{E}_{vac,i}\cdot\sum_{l=0}^{\infty}\left(\frac{(2l+1)\varepsilon_h}{(l+1)\varepsilon_s+l\varepsilon_h}\mathbf{E}_{vac,i}^{(l)}\right)dV\\&-\frac{1}{8\pi}\left(1-\frac{\varepsilon_s}{\varepsilon_h}\right)\int_{\substack{solute,\\ r>R_i}}\mathbf{E}_{vac,i}\cdot\sum_{l=0}^{\infty}\left(\frac{(2l+1)\varepsilon_h}{(l+1)\varepsilon_s+l\varepsilon_h}\mathbf{E}_{vac,i}^{(l)}\right)dV.\end{aligned} \quad (16)$$

The first integral is the solvation energy of a lone multipole ΔG_i^M and the second represents the effect of descreening sites. Substituting ΔG_i^M into Eq. (16) gives

$$\Delta G_i^{SFA} = \Delta G_i^M - \frac{1}{8\pi}\left(1-\frac{\varepsilon_s}{\varepsilon_h}\right) \int_{\substack{solute,\\ r>R_i}} \mathbf{E}_{vac,i} \cdot \sum_{l=0}^{\infty} \left(\frac{(2l+1)\varepsilon_h}{(l+1)\varepsilon_s + l\varepsilon_h} \mathbf{E}_{vac,i}^{(l)} \right) dV \quad (17)$$

where

$$\Delta G_i^M = \frac{1}{2}\left[c_0 \frac{q_i^2}{a_i} + c_1 \frac{\mu_{i,\alpha}^2}{a_i^3} + c_2 \frac{2}{3} \frac{\Theta_{i,\alpha\beta}^2}{a_i^5} \right] \quad (18)$$

and

$$c_l = \frac{1}{\varepsilon_h} \frac{(l+1)(\varepsilon_h - \varepsilon_s)}{(l+1)\varepsilon_s + l\varepsilon_h}. \quad (19)$$

In Eq. (18) we have assumed the Einstein convention for summation over Greek subscripts α and β which can take the value x, y, or z. The descreening integral in Eq. (17), which we will refer to as I_i, can be decomposed into a sum of pairwise integrals I_{ij}[8,9]

$$I_i(r_{ij}, R_i, R_j) = \sum_{j \neq i} \int_0^{\xi_{ij}} \int_0^{2\pi} \mathbf{E}_{vac,i} \cdot \sum_{l=0}^{\infty} \left(\frac{(2l+1)\varepsilon_h}{(l+1)\varepsilon_s + l\varepsilon_h} \mathbf{E}_{vac,i}^{(l)} \right) r^2 \sin\theta \, d\phi \, d\theta \, dr = \sum_{j \neq i} I_{ij}(r_{ij}, R_i, R_j) \quad (20)$$

where ξ_{ij} is the angle formed between the pairwise axis and any ray that begins at the center of atom *i* and passes through the circle of intersection between the integration shell and atom *j*

$$\xi_{ij} = \cos^{-1}\left(\frac{r_{ij}^2 - R_j^2 + r^2}{2 r_{ij} r} \right) \quad (21)$$

where r_{ij} is the distance between atoms *i* and *j*, R_j is the radius of atom *j* and *r* is the radial integration variable. The integration limits for the radial coordinate depend on what extent atoms *i* and *j* intersect, and therefore the solution to Eq. (20) is presented as an indefinite integral that is to be evaluated at limits described below. Typically the radius of the descreening atom is scaled down to prevent over counting due to atomic overlap, although parameter free approaches are being explored.[34] Specifically, R_j is replaced with sR_j where the HCT scale factor *s* is a parameter between 0 and 1 fit to reproduce PMPB results (see section 2.3 below).

Unlike the field due to a partial charge, the field due to a multipole of arbitrary order has an angular dependence. Our approach has been to represent the field using a spherical harmonic basis, rather than Cartesian tensors, to determine the analytic solution to Eq. (20) through quadrupole order. Additionally, it is assumed that the positive z-axis of the multipole frame is directed towards the center of the descreening atom. This imposes symmetry that greatly reduces the number of non-vanishing terms in the solution, but requires rotation of multipole moments for each pairwise descreening interaction.

A complex definition of spherical harmonics is commonly used in the formulation of quantum mechanics, however this work uses the following real form

$$Y_l^{(m)}(\theta,\phi)=\begin{cases}(-1)^m\sqrt{2}\sqrt{\frac{(l-m)!}{(l+m)!}}P_l^{(m)}(\cos\theta)\cos m\phi & m>0\\ \sqrt{\frac{(l-m)!}{(l+m)!}}P_l^{(m)}(\cos\theta) & m=0\\ (-1)^{|m|}\sqrt{2}\sqrt{\frac{(l-|m|)!}{(l+|m|)!}}P_l^{(|m|)}(\cos\theta)\sin|m|\phi & m<0\end{cases} \quad (22)$$

where $Y_l^{(m)}(\theta,\phi)$ is of degree $l \geq 0$ and order $|m| \leq l$, $P_l^{(m)}$ are the associated Legendre polynomials, the polar angle ranges from $0 \leq \theta \leq \pi$ and the azimuth ranges from $0 \leq \phi \leq 2\pi$.

We chose to use the Racah normalization, which has the property that $Y_l^{(0)}(0,0)=1$. In combination with our choice of phase factors, this ensures formulas for the conversion between Cartesian multipole moments and those consistent with this definition of real spherical harmonics are identical to the conversions commonly used for complex spherical harmonics. The conversion formulas through quadrupole degree are given in Table A-1 of the Supporting Information.[42]

The potential due to a unit magnitude multipole moment $\Phi_l^{(m)}(r,\theta,\phi)$ is obtained by multiplication of the real spherical harmonics by a radial factor of $1/r^{l+1}$ to give

$$\Phi_l^{(m)}(r,\theta,\phi)=\frac{Y_l^{(m)}(\theta,\phi)}{r^{l+1}} \quad (23)$$

and are listed in Table A-2 through quadrupole order. The unit field can then be calculated as the negative gradient of the unit potential

$$\mathbf{E}_l^{(m)}=-\nabla\Phi_l^{(m)}(r,\theta,\phi)=-\frac{\partial\Phi_l^{(m)}(r,\theta,\phi)}{\partial r}\hat{\mathbf{r}}-\frac{1}{r}\frac{\partial\Phi_l^{(m)}(r,\theta,\phi)}{\partial\theta}\hat{\boldsymbol{\theta}}-\frac{1}{r\sin\theta}\frac{\partial\Phi_l^{(m)}(r,\theta,\phi)}{\partial\phi}\hat{\boldsymbol{\phi}}. \quad (24)$$

The field for 9 multipole components through degree 2, which are listed in Table A-3 of the Supporting Information, lead to 36 scalar products that must be integrated via Eq. (20) to determine the descreening energy due to atom *j*. However, due to the symmetry of the integration domain only 14 scalar products lead to non-zero integrals, and these are listed in Table A-4. The integration results are given in Table 1, showing 11 unique terms and 3 duplicates. Schaeffer et al. originally presented the same result for a monopole[6,7], and the higher order formulas are presented here for the first time. If the descreening angle ξ_{ij} is π as a result of atom *j* completely engulfing atom *i*, then the indefinite integrals simplify to those given in Table 2. This situation can occur for hydrogen atoms bonded to a heavy atom, for example, or in more artificial structures where one still wishes to have a continuous potential.

We note that after performing the integral no angular dependence remains. Therefore, although the derivation is based on spherical harmonics, our solution is equally useful for Cartesian tensors by using the conversion formulas in Table A-1. We can now define the pairwise descreening integral for a permanent atomic multipole at site *i* being descreened by site *j* under the SFA as

$$I_{ij}\left(r_{ij},R_i,R_j\right)=\sum_{l_i=0}^{n}\frac{(2l_i+1)\varepsilon_h}{(l_i+1)\varepsilon_s+l_i\varepsilon_h}\sum_{m_i=-l_i}^{l_i}Q_{l_i}^{(m_i)}\sum_{l_j=0}^{n}\sum_{m_j=-l_j}^{l_i}Q_{l_j}^{(m_j)}D_{(l,m)_j,(l,m)_j}\left(r_{ij},R_i,R_j\right) \quad (25)$$

where $Q_{l_i}^{(m_i)}$ is the magnitude of a spherical harmonic of site *i*, $Q_{l_j}^{(m_j)}$ is the magnitude of a spherical harmonic of site *j* and $D_{(l,\,m)i,\,(l,\,m)j}\,(r_{ij},\,R_i,\,R_j)$ is given by

$$D_{(l,m)_i,(l,m)_j}\left(r_{ij},R_i,R_j\right)=\begin{cases}\delta_{(l_1,l_2)}\delta_{(m_1,m_2)}D_{l_i}\Big|_{r=R_i}^{r=R_j-r_{ij}}+D_{(l,m)_i,(l,m)_j}\left(r_{ij},R_j\right)\Big|_{r=R_j-r_{ij}}^{r=r_{ij}+R_j} & R_j-r_{ij}>R_i \\ \text{Case 1:Engulfment by the descreener} & \\ D_{(l,m)_i,(l,m)_j}\left(r_{ij},R_j\right)\Big|_{r=R_i}^{r=r_{ij}+R_j} & \begin{matrix}R_j-r_{ij}<=R_i\\ r_{ij}<R_i+R_j\end{matrix} \\ \text{Case 2:Partial overlap} & \\ D_{(l,m)_i,(l,m)_j}\left(r_{ij},R_j\right)\Big|_{r=r_{ij}-R_j}^{r=r_{ij}+R_j} & r_{ij}>R_i+R_j \\ \text{Case 3:No overlap} & \end{cases} \quad (26)$$

Radial limits are detailed for three cases including engulfment by the descreener, partial overlap and no overlap. These limits are applied in conjunction with the indefinite integrals $D_{(l,\,m)i,\,(l,\,m)j}\,(r_{ij},\,R_j)$ and D_{l_i} listed in Table 1 and Table 2, respectively. We note that the Kronecker delta functions δ specify that the engulfment integrals between orthogonal spherical harmonics vanish. In our implementation of Eq. (25), the magnitudes of the spherical harmonic moments are found via conversion from AMOEBA traceless Cartesian multipoles.

2.2 Reaction potential approximation

An alternative to the CFA for determining effective radii based on the analytic solution for the reaction potential of an off-center charge within a spherical dielectric cavity[1,43] has been proposed by Grycuk[38]. We briefly outline this RPA method and its application to the self-energy of a permanent multipole.

The reaction potential at ***r*** due to an off-center charge at $\boldsymbol{r_0}$ inside a spherical dielectric cavity of permittivity ε_h surrounded by solvent with permittivity ε_s is given by

$$\Phi(\mathbf{r})=\frac{q}{a\varepsilon_h}\sum_{l=0}^{\infty}\frac{(l+1)(\varepsilon_h-\varepsilon_s)}{(l+1)\varepsilon_s+l\varepsilon_h}\left(\frac{rr_0}{a^2}\right)^l P_l(\cos\,\theta) \quad (27)$$

where *a* is the cavity radius, *q* is the magnitude of the charge and P_l is the Legendre polynomial of degree *l* whose argument is the cosine of the angle θ between *r* and r_0.[1,43] The self-energy of a charge based on Eq. (27) is

$$W(d)=\frac{1}{2}\frac{q^2}{a\varepsilon_h}\sum_{l=0}^{\infty}\frac{(l+1)(\varepsilon_h-\varepsilon_s)}{(l+1)\varepsilon_s+l\varepsilon_h}\left(\frac{d^2}{a^2}\right)^l P_l(1) \quad (28)$$

where d is used to specify the distance between the multipole site and the center of the sphere. For d=0, all asymmetric self-interactions vanish, for example the charge with a dipole component, but for off-center multipole sites these interactions are generally nonzero.[5] Noting that P_l (1) = 1 for all l, the summation in Eq. (28) can be reduced to a closed form if the factor (l+1) can be canceled by setting $l\varepsilon_h$ in the denominator to (l+1)ε_h or to 0, giving quantities that are more positive W_+ (d) or more negative W_- (d) than the true self-energy, respectively

$$W_+(d)=\frac{1}{2}\frac{(\varepsilon_h-\varepsilon_s)}{\varepsilon_s\varepsilon_h+{\varepsilon_h}^2}\frac{q^2}{a}\sum_{l=0}^{\infty}\left(\frac{d^2}{a^2}\right)^l=\frac{1}{2}\frac{(\varepsilon_h-\varepsilon_s)}{(\varepsilon_s\varepsilon_h+{\varepsilon_h}^2)}q^2\frac{a}{(a^2-d^2)}. \quad (29)$$

$$W_-(d)=\frac{1}{2}\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)\frac{q^2}{a}\sum_{l=0}^{\infty}\left(\frac{d^2}{a^2}\right)^l=\frac{1}{2}\left(\frac{1}{\varepsilon_s}-\frac{1}{\varepsilon_h}\right)q^2\frac{a}{(a^2-d^2)}$$

Both the upper and lower bound approach the true self-energy if $\varepsilon_s \square \varepsilon_h$ allowing the simpler form to be used as an approximation

$$W(d)\approx W_-(d) \quad (30)$$

As shown by Grycuk, it is possible to calculate the factor $a_r = a/(a^2 - d^2)$, which is equivalent to the inverse of an effective radius, as

$$a_r=\left(\frac{3}{4\pi}\int_{ex}\frac{1}{r'^6}dV\right)^{1/3}. \quad (31)$$

This expression can be motivated by the analytic solution for a spherical geometry

$$\int_{ex}\frac{1}{r'^6}dV=2\pi\int_a^{\infty}\int_0^{\pi}\frac{r^2\sin\theta}{(r^2+d^2-2dr\cos\theta)^3}d\theta dr=\frac{4\pi}{3}\frac{a^3}{(a^2-d^2)^3} \quad (32)$$

As d approaches zero, the multipole site approaches the center of the dielectric sphere such that a_r equals the radius of the sphere a. In practice this integral is evaluated using the pairwise descreening approach described in the previous section for the SFA and elsewhere.[8,9,38] After determining effective radii, the self-energy for each permanent atomic multipole under the RPA is evaluated via Eq. (18).

2.3 Self-energy accuracy

We now demonstrate that for a series of proteins the RPA is superior to the SFA, which is consistent with findings for fixed partial charge models.[38,44] The perfect self-energy and perfect effective radii for all permanent atomic multipole sites in five proteins structures retrieved from the Protein Databank[45], including 1CRN[46], 1ENH[47], 1FSV[48], 1PGB[49] and 1VII[50], were determined using the PMPB model.[2] The grid size for all calculations was 257×257×257 using a grid spacing of 0.31 Å to give approximately 10 Å of continuum solvent between the low dielectric boundary and the grid boundary. The Bondi radii set (H

1.2, C 1.7, N 1.55, O 1.52, S 1.8) was used to define a step-function solute-solvent boundary with the solute dielectric set to unity and that of the solvent to 78.3.[51] Multiple Debye-Hückel boundary conditions were used to complete the definition of the Dirichlet problem. We also tried larger grids, up to 353×353×353, and therefore smaller grid spacing, which leads to the PMPB electrostatic solvation energy increasing by less than 2%. We opted for efficiency, since the important conclusion of this section, that the RPA is superior to the SFA, is not altered.

The SFA was fit using nonlinear optimization to determine one HCT scale factor per atomic number that minimized the RMS percent error in the permanent atomic multipole self-energies against numerical PMPB results for 3,032 data points. As discussed previously, these HCT parameters scale down the radius of the descreening atom to prevent over counting due to atomic overlap. This lead to a mean unsigned relative difference (MURD) between the perfect self-energy for each multipole site and the SFA self-energy of 5.5%. However, using only a single scale factor (0.568), rather than one per atomic number, increased the MURD by just 0.4 to 5.9%.

Similarly, the RPA was fit using nonlinear optimization to determine a second set of scale factors to minimize the RMS percent difference between analytic effective radii and perfect effective radii. The achieved MURD in the effective radii was 1.1%. Alternatively, using a single scale factor (0.690) increased the MUPD by only 0.2% to 1.3%. Therefore, given the negligible improvements of using one HCT parameter per atomic number, we prefer implementations of the SFA and RPA that are each based on a single parameter.

The total analytic self-energy for each protein is compared to the total computed by summing the numerical permanent multipole self-energies as shown in Table 3. Fitting of a single HCT parameter for each method as described above eliminated the systematic error for both the SFA and RPA. However, the mean unsigned percent difference of the RPA (0.5) is smaller than that of the SFA (0.8). Considering that the RPA is more efficient and more accurate than the SFA, it is our preferred method to compute effective radii and permanent multipole self-energies.

3 MULTIPOLE CROSS-TERM ENERGY

There are two concepts needed to extend the GB cross-term to the interaction between two arbitrary multipole components. First, we describe the simplest possible definition for the reaction potential of any multipole component in the presence of a second multipole site, where an effective radius characterizes each site. Second, using this auxiliary definition of the reaction potential for each site, we formulate the cross-term energy in a consistent fashion. The electrostatic solvation free energy for the interaction between multipole components will be reproduced in the limiting cases of superimposition and wide separation.

3.1 Generalized Kirkwood auxiliary reaction potential

The generalized Kirkwood auxiliary reaction potential is a building block for defining the interaction energy and its gradients for any pair of multipole components. It is motivated by noting that the only difference between the analytic solution for the reaction potential inside

and outside of a spherical solute with central multipole is exchange of the solute radius *a* in the former case with separation distance r_{ij} in the latter, where $\mathbf{r}_{ij} = (x_j - x_i, y_j - y_i, z_j - z_i)$.[1,41] For example, substitution for *f* in Eq. (33) below by *a* or r_{ij} gives the analytic formulas for the reaction potential inside and outside of the dielectric boundary, respectively.

Rather than using radial factors of $1/r_{ij}^{l+1}$ as was done earlier in defining the unit vacuum potential in terms of real spherical harmonics, the factor r_{ij}^l/f^{2l+1} is used to define the unit GK auxiliary reaction potential $A_l^{(m)}$ for a multipole component of degree *l* and order *m*,

$$A_l^{(m)}(\mathbf{r}_{ij}, a_i, a_j, \theta, \phi) = c_l \frac{r_{ij}^l}{f^{2l+1}} Y_l^{(m)}(\theta, \phi) \quad (33)$$

where *f* is the generalizing function defined in Eq. (6) and c_l is a function of the permittivity inside and outside the solute defined in Eq. (19). We note that for $r_{ij}^2 \gg a_i a_j, r_{ij}^l/f^{2l+1}$ approaches $1/r_{ij}^{l+1}$ to give the reaction potential in solvent. When $r_{ij} = 0$ and therefore $a_i = a_j = a$, then r_{ij}^l/f^{2l+1} simplifies to r_{ij}^l/a^{2l+1} to give the reaction potential at the center of the two concentric atoms. In this case the reaction potential is nonzero only for the monopole.

A definition in terms of Cartesian tensors is possible by first taking successive gradients of $1/r_{ij}$ and then substituting for factors of r_{ij} in the denominator with factors of *f*. For example, neglecting the *ij* subscript, the vacuum tensors are[42]

$$T = \frac{1}{r}$$

$$T_\alpha = \nabla_\alpha \frac{1}{r} = -\frac{r_\alpha}{r^3}$$

$$T_{\alpha\beta} = \nabla_\alpha \nabla_\beta \frac{1}{r} = \frac{3 r_\alpha r_\beta}{r^5} - \frac{\delta_{\alpha\beta}}{r^3}$$

$$T_{\alpha\beta\gamma} = \nabla_\alpha \nabla_\beta \nabla_\gamma \frac{1}{r} = -\frac{15 r_\alpha r_\beta r_\gamma}{r^7} + \frac{3\left(r_\alpha \delta_{\beta\gamma} + r_\beta \delta_{\alpha\gamma} + r_\gamma \delta_{\alpha\beta}\right)}{r^5} \quad (34)$$

$$\begin{aligned} T_{\alpha\beta\gamma\delta} = \nabla_\alpha \nabla_\beta \nabla_\gamma \nabla_\delta \frac{1}{r} &= \frac{105 r_\alpha r_\beta r_\gamma r_\delta}{r^9} \\ &- \frac{15\left(r_\alpha r_\beta \delta_{\gamma\delta} + r_\alpha r_\gamma \delta_{\beta\delta} + r_\alpha r_\delta \delta_{\beta\gamma} + r_\beta r_\gamma \delta_{\alpha\delta} + r_\beta r_\delta \delta_{\alpha\gamma} + r_\gamma r_\delta \delta_{\alpha\beta}\right)}{r^7} \\ &+ \frac{3\left(\delta_{\alpha\beta}\delta_{\gamma\delta} + \delta_{\alpha\gamma}\delta_{\beta\delta} + \delta_{\alpha\delta}\delta_{\beta\gamma}\right)}{r^5} \end{aligned}$$

where α, β, γ and δ can take the values x, y, or z and the Kronecker delta function is unity if its subscripts are equal, but zero otherwise. Applying the substitution gives

$$A = c_0 \frac{1}{f}$$

$$A_\alpha = -c_1 \frac{r_\alpha}{f^3}$$

$$A_{\alpha\beta} = c_2 \frac{3 r_\alpha r_\beta}{f^5} \quad (35)$$

$$A_{\alpha\beta\gamma} = -c_3 \frac{15 r_\alpha r_\beta r_\gamma}{f^7}$$

$$A_{\alpha\beta\gamma\delta} = c_4 \frac{105 r_\alpha r_\beta r_\gamma r_\delta}{f^9}$$

and represent the GK auxiliary reaction potential tensors. We have removed terms that require summing over a trace by requiring use of traceless multipoles. Unlike the vacuum case, the GK auxiliary reaction potential tensor of degree *l* is not simply a gradient of a degree *l*−1 tensor.

The total auxiliary reaction potential due to multipole *i*, up to quadrupole order, at site *j* is

$$\phi^{(i)}(\mathbf{r}_{ij}, a_i, a_j) = q_i A - \mu_{i,\alpha} A_\alpha + \frac{1}{3}\Theta_{i,\alpha\beta} A_{\alpha\beta} \quad (36)$$

where the Einstein convention for repeated summation over Greek subscripts is implied. The total auxiliary potential due to multipole *j*, up to quadrupole order, at site *i* is given by

$$\phi^{(j)}(\mathbf{r}_{ji}, a_i, a_j) = q_j A - \mu_{j,\alpha} A_\alpha + \frac{1}{3}\Theta_{j,\alpha\beta} A_{\alpha\beta} \quad (37)$$

where $\boldsymbol{r}_{ji}$ is defined from site *j* to site *i*.

3.2 Generalized Kirkwood cross-term

Given the auxiliary reaction potentials, we define the auxiliary cross-term energy using Eq. (36) to be

$$U^{(i)}(\mathbf{r}_{ij}, a_i, a_j) = \frac{1}{2}\left(q_j \phi^{(i)} + \mu_{j,\gamma} \nabla_\gamma \phi^{(i)} + \frac{1}{3}\Theta_{j,\gamma\delta} \nabla_\gamma \nabla_\delta \phi^{(i)}\right) \quad (38)$$

such that substituting for $\varphi^{(i)}$ gives

$$U^{(i)}(\mathbf{r}_{ij}, a_i, a_j) = \frac{1}{2}\left[q_j\left(q_i A - \mu_{i,\alpha}A_\alpha + \frac{1}{3}\Theta_{i,\alpha\beta}A_{\alpha\beta}\right) + \mu_{j,\gamma}\nabla_\gamma\left(q_i A - \mu_{i,\alpha}A_\alpha + \frac{1}{3}\Theta_{i,\alpha\beta}A_{\alpha\beta}\right) + \frac{1}{3}\Theta_{j,\gamma\delta}\nabla_\gamma\nabla_\delta\left(q_i A - \mu_{i,\alpha}A_\alpha + \frac{1}{3}\Theta_{i,\alpha\beta}A_{\alpha\beta}\right)\right] \quad (39)$$

while the auxiliary cross-term energy using Eq. (37) is

$$U^{(j)}(\mathbf{r}_{ji}, a_i, a_j) = \frac{1}{2}\left(q_i\phi^{(j)} + \mu_{i,\gamma}\nabla_\gamma\phi^{(j)} + \frac{1}{3}\Theta_{i,\gamma\delta}\nabla_\gamma\nabla_\delta\phi^{(j)}\right) \quad (40)$$

such that substituting for $\varphi^{(j)}$ gives

$$U^{(j)}(\mathbf{r}_{ji}, a_i, a_j) = \frac{1}{2}\left[q_i\left(q_j A - \mu_{j,\alpha}A_\alpha + \frac{1}{3}\Theta_{j,\alpha\beta}A_{\alpha\beta}\right) + \mu_{i,\gamma}\nabla_\gamma\left(q_j A - \mu_{j,\alpha}A_\alpha + \frac{1}{3}\Theta_{j,\alpha\beta}A_{\alpha\beta}\right) + \frac{1}{3}\Theta_{i,\gamma\delta}\nabla_\gamma\nabla_\delta\left(q_j A - \mu_{j,\alpha}A_\alpha + \frac{1}{3}\Theta_{j,\alpha\beta}A_{\alpha\beta}\right)\right] \quad (41)$$

In the case of superimposition, either $U^{(i)}$ or $U^{(j)}$ exactly reproduces the correct self-energies. In the case of wide separation, both $\varphi^{(i)}$ and $\varphi^{(j)}$ neglect the bending of field lines near the spherical dielectric cavity surrounding site *j* and site *i*, respectively. The density of field lines in the case of wide separation is not an issue for a fixed partial charge interaction, although neglect of this effect introduces an error of less than 1% for dipole interactions in the case of a solute with unit permittivity in water.

Gradients of the auxiliary reaction potential can easily be obtained, although it is important to note that

$$\nabla_\alpha A \neq A_\alpha. \quad (42)$$

Namely, $\nabla_\alpha A$ includes a factor of $\left(1 - e^{-r_{ij}^2/c_f a_i a_j}/c_f\right)$ relative to A_α such that equality is only achieved for r_{ij} equal to zero or infinity. This subtle point implies, not surprisingly, the auxiliary reaction potential is too simple for intermediate r_{ij}. An important consequence is that $U^{(i)} \neq U^{(j)}$. A consistent model requires that the α-component of the potential gradient at site *j* of a unit charge at site *i* should equal the potential at site *i* of the dipole's unit magnitude α-component at site *j*. This reciprocity condition is a well-known property of linear dielectric continuums.[40] We note that in practice $\nabla_\alpha A \approx A_\alpha$ and therefore we simply take the average of the energies to obtain a consistent interaction model.

$$\Delta G_{ij} = \frac{1}{2}\left(U^{(i)} + U^{(j)}\right) \quad (43)$$

The qualitative behavior of the GK cross-term formulation for multipole permutations through quadrupole degree is seen in Figure (1). The system is composed of two spheres,

each with a radius of 3.0 Å and unit permittivity, in a solvent with permittivity 78.3. The total electrostatic solvation energy was evaluated using the PMPB and GK models. In the case of superimposition, the GK value is exact. When the two spheres are widely separated, GK asymptotes to the PMPB results for all permutations. For intermediate separations, the behavior is promising, but not exact.

4 AMOEBA SOLUTES UNDER GENERALIZED KIRKWOOD

4.1 Electrostatic solvation free energy

Derivation of the electrostatic solvation free energy for an AMOEBA solute[35–37] within the GK continuum resembles the derivation of the PMPB electrostatic solvation free energy.[2] Each permanent atomic multipole site can be considered as a vector of coefficients including charge, dipole and quadrupole components

$$\mathbf{M}_i = [q_i, d_{i,x}, d_{i,y}, d_{i,z}, \Theta_{i,xx}, \Theta_{i,xy}, \Theta_{i,xz}, \ldots, \Theta_{i,zz}]^t \quad (44)$$

where the superscript *t* denotes the transpose. The interaction potential energy between two sites *i* and *j* separated by the distance r_{ij} in a homogeneous permittivity ε_h can then be represented in tensor notation as

$$\mathrm{U}(\mathbf{r}_{ij}) = \mathbf{M}_i^t \mathbf{T}_{ij} \mathbf{M}_j = \begin{bmatrix} q_i \\ d_{i,x} \\ d_{i,y} \\ d_{i,z} \\ \Theta_{i,xx} \\ \vdots \end{bmatrix}^t \begin{bmatrix} 1 & \frac{\partial}{\partial x_j} & \frac{\partial}{\partial y_j} & \frac{\partial}{\partial z_j} & \cdots \\ \frac{\partial}{\partial x_i} & \frac{\partial^2}{\partial x_i \partial x_j} & \frac{\partial^2}{\partial x_i \partial y_j} & \frac{\partial^2}{\partial x_i \partial z_j} & \cdots \\ \frac{\partial}{\partial y_i} & \frac{\partial^2}{\partial y_i \partial x_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \frac{\partial^2}{\partial y_i \partial z_j} & \cdots \\ \frac{\partial}{\partial z_i} & \frac{\partial^2}{\partial z_i \partial x_j} & \frac{\partial^2}{\partial z_i \partial y_j} & \frac{\partial^2}{\partial z_i \partial z_j} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \frac{1}{\varepsilon_h r_{ij}} \begin{bmatrix} q_j \\ d_{j,x} \\ d_{j,y} \\ d_{j,z} \\ \Theta_{j,xx} \\ \vdots \end{bmatrix} \quad (45)$$

Similarly, the GK energy for two multipoles (self or cross-term) is given by

$$\Delta G_{ij}(\mathbf{r}_{ij}, a_i, a_j) = \frac{1}{2} \mathbf{M}_i^t \mathbf{K}_{ij} \mathbf{M}_j \quad (46)$$

where the factor of ½ accounts for the cost of charging the continuum and the GK interaction matrix $\mathbf{K}_{ij}$ depends on the coordinates of all atoms via the effective radii a_i and a_j. As introduced above, GK requires averaging of the auxiliary reaction potentials and their respective gradients to obtain a consistent interaction matrix

$$\mathbf{K}_{ij} = \frac{1}{2} \left[\mathbf{K}^{(i)} + \mathbf{K}^{(j)} \right]$$

$$\mathbf{K}^{(i)}(\mathbf{r}_{ij}, a_i, a_j) = \begin{bmatrix} A & \frac{\partial A}{\partial x} & \frac{\partial A}{\partial y} & \frac{\partial A}{\partial z} & \cdots \\ A_x & \frac{\partial A_x}{\partial x} & \frac{\partial A_x}{\partial y} & \frac{\partial A_x}{\partial z} & \cdots \\ A_y & \frac{\partial A_y}{\partial x} & \frac{\partial A_y}{\partial y} & \frac{\partial A_y}{\partial z} & \cdots \\ A_z & \frac{\partial A_z}{\partial x} & \frac{\partial A_z}{\partial y} & \frac{\partial A_z}{\partial z} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (47)$$

$$\mathbf{K}^{(j)}\left(\mathbf{r}_{ji}, a_i, a_j\right) = \left(\mathbf{K}^{(i)}\left(\mathbf{r}_{ji}, a_i, a_j\right)\right)^t$$

Each site may also be polarizable, such that an induced dipole is formed in vacuum $\boldsymbol{\mu}_i^v$ proportional to the strength of the local field

$$\boldsymbol{\mu}_i^v = \alpha_i \mathbf{E}_i^v = \alpha_i \left(\sum_{j \neq i} \mathbf{T}_{d,ij}^{(1)} \mathbf{M}_j + \sum_{k \neq i} \mathbf{T}_{ik}^{(11)} \boldsymbol{\mu}_k \right) \quad (48)$$

Here α_i is an isotropic atomic polarizability and $\mathbf{E}_i^v$ is the total vacuum field, which can be decomposed into contributions from permanent multipole sites and induced dipoles, and the summations run over all multipole sites. The interaction tensors $\mathbf{T}_{d,ij}^{(1)}$ and $\mathbf{T}_{ik}^{(11)}$ are, respectively,

$$\mathbf{T}_{d,ij}^{(1)} = \begin{bmatrix} \frac{\partial}{\partial x_i} & \frac{\partial^2}{\partial x_i \partial x_j} & \frac{\partial^2}{\partial x_i \partial y_j} & \frac{\partial^2}{\partial x_i \partial z_j} & \cdots \\ \frac{\partial}{\partial y_i} & \frac{\partial^2}{\partial y_i \partial x_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \frac{\partial^2}{\partial y_i \partial z_j} & \cdots \\ \frac{\partial}{\partial z_i} & \frac{\partial^2}{\partial z_i \partial x_j} & \frac{\partial^2}{\partial z_i \partial y_j} & \frac{\partial^2}{\partial z_i \partial z_j} & \cdots \end{bmatrix} \frac{1}{\varepsilon_h r_{ij}} \quad (49)$$

and

$$\mathbf{T}_{ik}^{(11)} = \begin{bmatrix} \frac{\partial^2}{\partial x_i \partial x_k} & \frac{\partial^2}{\partial x_i \partial y_k} & \frac{\partial^2}{\partial x_i \partial z_k} \\ \frac{\partial^2}{\partial y_i \partial x_k} & \frac{\partial^2}{\partial y_i \partial y_k} & \frac{\partial^2}{\partial y_i \partial z_k} \\ \frac{\partial^2}{\partial z_i \partial x_k} & \frac{\partial^2}{\partial z_i \partial y_k} & \frac{\partial^2}{\partial z_i \partial z_k} \end{bmatrix} \frac{1}{\varepsilon_h r_{ik}} \quad (50)$$

where the d in $\mathbf{T}_{d,ij}^{(1)}$ denotes that masking rules for the AMOEBA group-based polarization model are applied. Upon adding the GK reaction field due to the permanent multipoles and induced dipoles, the self-consistent induced dipoles are proportional to the self-consistent reaction field

$$\boldsymbol{\mu}_i = \alpha_i \mathbf{E}_i = \alpha_i \left[\sum_j \left[(1 - \delta_{ij}) \mathbf{T}_{d,ij}^{(1)} + \mathbf{K}_{ij}^{(1)} \right] \mathbf{M}_j + \sum_k \left[(1 - \delta_{ik}) \mathbf{T}_{ik}^{(11)} + \mathbf{K}_{ik}^{(11)} \right] \boldsymbol{\mu}_k \right] \quad (51)$$

where the sums now include self-contributions to the reaction field, but exclude Coulomb self-interactions via Kronecker delta functions. The GK interaction matrices $\mathbf{K}_{ij}^{(1)}$ and $\mathbf{K}_{ik}^{(11)}$ are, respectively,

$$\mathbf{K}_{ij}^{(1)} = \frac{1}{2} \left(\mathbf{K}_{ij}^{(1,i)}\left(\mathbf{r}_{ij}, a_i, a_j\right) + \mathbf{K}_{ij}^{(1,j)}\left(\mathbf{r}_{ji}, a_i, a_j\right) \right) \quad (52)$$

where

$$\mathbf{K}_{ij}^{(1,i)}\left(\mathbf{r}_{ij},a_i,a_j\right)=\begin{bmatrix} A_x & \frac{\partial A_x}{\partial x} & \frac{\partial A_x}{\partial y} & \frac{\partial A_x}{\partial z} & \cdots \\ A_y & \frac{\partial A_y}{\partial x} & \frac{\partial A_y}{\partial y} & \frac{\partial A_y}{\partial z} & \cdots \\ A_z & \frac{\partial A_z}{\partial x} & \frac{\partial A_z}{\partial y} & \frac{\partial A_z}{\partial z} & \cdots \end{bmatrix} \quad (53)$$

$$\mathbf{K}_{ij}^{(1,j)}\left(\mathbf{r}_{ji},a_i,a_j\right)=\begin{bmatrix} \frac{\partial A}{\partial x} & \frac{\partial A_x}{\partial x} & \frac{\partial A_y}{\partial x} & \frac{\partial A_z}{\partial x} & \cdots \\ \frac{\partial A}{\partial y} & \frac{\partial A_x}{\partial y} & \frac{\partial A_y}{\partial y} & \frac{\partial A_z}{\partial y} & \cdots \\ \frac{\partial A}{\partial z} & \frac{\partial A_x}{\partial z} & \frac{\partial A_y}{\partial z} & \frac{\partial A_z}{\partial z} & \cdots \end{bmatrix} \quad (54)$$

and

$$\mathbf{K}_{ik}^{(11)}=\begin{bmatrix} \frac{\partial A_x}{\partial x} & \frac{\partial A_x}{\partial y} & \frac{\partial A_x}{\partial z} \\ \frac{\partial A_y}{\partial x} & \frac{\partial A_y}{\partial y} & \frac{\partial A_y}{\partial z} \\ \frac{\partial A_z}{\partial x} & \frac{\partial A_z}{\partial y} & \frac{\partial A_z}{\partial z} \end{bmatrix} \quad (55)$$

where averaging cancels for the matrix $\mathbf{K}_{ik}^{(11)}$ that produces the field at site *i* due to the induced dipole at site *k* as a result of symmetry.

The linear system of equations, both for the vacuum and solvated systems, can be solved via a number of approaches, including direct matrix inversion or iterative schemes such as successive over-relaxation (SOR). The total vacuum electrostatic energy $\mathrm{U}_{\mathrm{elec}}^{\mathrm{v}}$ includes pairwise permanent multipole interactions and many-body polarization

$$\mathrm{U}_{\mathrm{elec}}^{\mathrm{v}}=\frac{1}{2}\left[\mathbf{M}^t\mathbf{T}-\left(\boldsymbol{\mu}^{\mathrm{v}}\right)^t\mathbf{T}_{\mathrm{p}}^{(1)}\right]\mathbf{M} \quad (56)$$

where the factor of ½ avoids double-counting of permanent multipole interactions in the first term and accounts for the cost of polarizing the system in the second term. Furthermore, **M** is a column vector of 13N multipole components

$$\mathbf{M}=\begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \vdots \\ \mathbf{M}_{\mathrm{N}} \end{bmatrix} \quad (57)$$

T is a N × N supermatrix with $\mathbf{T}_{ij}$ off-diagonal elements

$$\mathbf{T}=\begin{bmatrix} 0 & \mathbf{T}_{12} & \mathbf{T}_{13} & \cdots \\ \mathbf{T}_{21} & 0 & \mathbf{T}_{23} & \cdots \\ \mathbf{T}_{31} & \mathbf{T}_{32} & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (58)$$

$\boldsymbol{\mu}^{\mathrm{v}}$ is a 3N column vector of converged induced dipole components in vacuum

$$\boldsymbol{\mu}^{\mathrm{v}}=\begin{bmatrix}\mu^{v}_{1,x}\\ \mu^{v}_{1,y}\\ \mu^{v}_{1,z}\\ \vdots\\ \mu^{v}_{N,z}\end{bmatrix} \quad (59)$$

and $\mathbf{T}^{(1)}_{\mathrm{p}}$ is a 3N × 13N supermatrix with $\mathbf{T}^{(1)}_{\mathrm{p},ij}$ as off-diagonal elements

$$\mathbf{T}^{(1)}_{\mathrm{p}}=\begin{bmatrix}0 & \mathbf{T}^{(1)}_{\mathrm{p},12} & \mathbf{T}^{(1)}_{\mathrm{p},13} & \cdots\\ \mathbf{T}^{(1)}_{\mathrm{p},21} & 0 & \mathbf{T}^{(1)}_{\mathrm{p},23} & \cdots\\ \mathbf{T}^{(1)}_{\mathrm{p},31} & \mathbf{T}^{(1)}_{\mathrm{p},32} & 0 & \cdots\\ \vdots & \vdots & \vdots & \ddots\end{bmatrix}. \quad (60)$$

The subscript p denotes a tensor matrix that operates on the permanent multipoles to produce the electric field in which the polarization energy is evaluated, while the subscript d is used to specify an analogous tensor matrix that produces the field that induces dipoles. The differences between the two are masking rules that leave out 1–2, 1–3, and 1–4 interactions in the former case and use the AMOEBA group based polarization scheme for the later.[35]

For the solvated system, the total electrostatic energy is similar to the vacuum case

$$\mathrm{U}_{\mathrm{elec}}=\frac{1}{2}\left[\mathbf{M}^{t}\left(\mathbf{T}+\mathbf{K}\right)-\boldsymbol{\mu}^{t}\left(\mathbf{T}^{(1)}_{\mathrm{p}}+\mathbf{K}^{(1)}\right)\right]\mathbf{M} \quad (61)$$

where the GK matrices are

$$\mathbf{K}=\begin{bmatrix}\mathbf{K}_{11} & \mathbf{K}_{12} & \mathbf{K}_{13} & \cdots\\ \mathbf{K}_{21} & \mathbf{K}_{22} & \mathbf{K}_{23} & \cdots\\ \mathbf{K}_{31} & \mathbf{K}_{32} & \mathbf{K}_{33} & \cdots\\ \vdots & \vdots & \vdots & \ddots\end{bmatrix} \quad (62)$$

and

$$\mathbf{K}^{(1)}=\begin{bmatrix}\mathbf{K}^{(1)}_{11} & \mathbf{K}^{(1)}_{12} & \mathbf{K}^{(1)}_{13} & \cdots\\ \mathbf{K}^{(1)}_{21} & \mathbf{K}^{(1)}_{22} & \mathbf{K}^{(1)}_{23} & \cdots\\ \mathbf{K}^{(1)}_{31} & \mathbf{K}^{(1)}_{32} & \mathbf{K}^{(1)}_{33} & \cdots\\ \vdots & \vdots & \vdots & \ddots\end{bmatrix} \quad (63)$$

The total electrostatic solvation free energy is determined as the difference between the vacuum electrostatic energy and total electrostatic energy in solvent as

$$\mathrm{U}_{\mathrm{solv}}=\frac{1}{2}\left(\mathbf{M}^{t}\mathbf{K}-\boldsymbol{\mu}^{\Delta}\mathbf{T}^{(1)}_{\mathrm{p}}-\boldsymbol{\mu}\mathbf{K}^{(1)}\right)\mathbf{M} \quad (64)$$

where $\boldsymbol{\mu}^{\Delta}$ represents the change in the induced dipoles upon solvation

$$\boldsymbol{\mu}^{\Delta}=\boldsymbol{\mu}-\boldsymbol{\mu}^{\mathrm{v}}. \quad (65)$$

4.2 Permanent multipole energy gradient

The permanent multipole electrostatic solvation energy gradient between sites *i* and *j* only depends on the gradient of the GK interaction tensor

$$\frac{\partial \mathbf{K}_{ij}}{\partial r_{i,\sigma}}=\frac{1}{2}\left[\left(\frac{\partial \mathbf{K}^{(i)}}{\partial r_{i,\sigma}}+\frac{\partial \mathbf{K}^{(j)}}{\partial r_{i,\sigma}}\right)_{a_i,a_j}+\left(\frac{\partial \mathbf{K}^{(i)}}{\partial a_i}+\frac{\partial \mathbf{K}^{(j)}}{\partial a_i}\right)\frac{\partial a_i}{\partial r_{i,\sigma}}+\left(\frac{\partial \mathbf{K}^{(i)}}{\partial a_j}+\frac{\partial \mathbf{K}^{(j)}}{\partial a_j}\right)\frac{\partial a_j}{\partial r_{i,\sigma}}\right] \quad (66)$$

and subscript a_i and a_j denote keeping the effective radii fixed in this case. Generation of the GK interaction tensors that make up $\frac{\partial \mathbf{K}^{(i)}}{\partial r_{i,\sigma}}$, $\frac{\partial \mathbf{K}^{(i)}}{\partial a_i}$, $\frac{\partial \mathbf{K}^{(i)}}{\partial a_j}$, $\frac{\partial \mathbf{K}^{(j)}}{\partial r_{i,\sigma}}$, $\frac{\partial \mathbf{K}^{(j)}}{\partial a_i}$ and $\frac{\partial \mathbf{K}^{(j)}}{\partial a_j}$ are described in Appendix B of the Supporting Information. The derivatives of the effective radii with respect to an atomic displacement follow from the pairwise descreening implementation of the RPA and will not be discussed here.[8,9,38] We also point out that there is a torque on the permanent dipoles due to the permanent reaction field and also on the permanent quadrupoles due to the permanent reaction field gradient. All torques, including contributions from the polarization energy gradient discussed below, are converted to forces on adjacent atoms that define the local coordinate frame of the multipole.

4.3 Polarization energy gradient

The polarization energy gradient when using either the “direct” or “mutual” polarization models within the GK continuum will now be derived. The definition of the starting point for the iterative convergence of the self-consistent reaction field (SCRF) is the total “direct” field $\mathbf{E}_{\mathrm{direct}}$ at each polarizable site. This field is the sum of the permanent atomic multipoles (PAM) intramolecular field

$$\mathbf{E}_{\mathrm{d}}=\mathbf{T}_{\mathrm{d}}^{(1)}\mathbf{M} \quad (67)$$

where $\mathbf{T}_{\mathrm{d}}^{(1)}$ is analogous to the tensor matrix defined in deriving the AMOEBA vacuum energy in Eq. (56), and the PAM GK reaction field

$$\mathbf{E}_{\mathrm{RF}}=\mathbf{K}^{(1)}\mathbf{M} \quad (68)$$

The product of the direct field $\mathbf{E}_{\mathrm{direct}}$ with a vector of atomic polarizabilities determines the initial induced dipoles $\boldsymbol{\mu}_{\mathrm{direct}}$

$$\boldsymbol{\mu}_{\mathrm{direct}}=\boldsymbol{\alpha}\mathbf{E}_{\mathrm{direct}}=\boldsymbol{\alpha}\left(\mathbf{T}_{\mathrm{d}}^{(1)}+\mathbf{K}^{(1)}\right)\mathbf{M} \quad (69)$$

At this point the induced dipoles do not act upon each other nor do they elicit a reaction field. This is defined as the direct model of polarization.

In contrast to the direct polarization model, the total SCRF **E** has two additional contributions due to the induced dipoles and their reaction field,

$$\mathbf{E}=\left(\mathbf{T}_{\mathrm{d}}^{(1)}+\mathbf{K}^{(1)}\right)\mathbf{M}+\left(\mathbf{T}^{(11)}+\mathbf{K}^{(11)}\right)\boldsymbol{\mu} \quad (70)$$

for a sum of 4 contributions. The induced dipoles

$$\boldsymbol{\mu}=\boldsymbol{\alpha}\left[\left(\mathbf{T}_{\mathrm{d}}^{(1)}+\mathbf{K}^{(1)}\right)\mathbf{M}+\left(\mathbf{T}^{(11)}+\mathbf{K}^{(11)}\right)\boldsymbol{\mu}\right] \quad (71)$$

can be solved for in an iterative fashion using successive over-relaxation (SOR) to accelerate convergence.[52] Alternatively, the induced dipoles can be solved for directly as a mechanism for deriving the polarization energy gradient with respect to an atomic displacement. Moving all terms containing the induced dipoles to the LHS allows their isolation

$$\left(\boldsymbol{\alpha}^{-1}-\mathbf{T}^{(11)}-\mathbf{K}^{(11)}\right)\boldsymbol{\mu}=\left(\mathbf{T}_{\mathrm{d}}^{(1)}+\mathbf{K}^{(1)}\right)\mathbf{M} \quad (72)$$

For convenience, a matrix C is defined as

$$\mathbf{C}=\boldsymbol{\alpha}^{-1}-\mathbf{T}^{(11)}-\mathbf{K}^{(11)} \quad (73)$$

which is substituted into Eq. (72) above to show the induced dipoles are a linear function of the PAM **M**, directly via the intramolecular interaction tensor $\mathbf{T}_{\mathrm{d}}^{(1)}$ that implicitly contains the AMOEBA group based polarization scheme, and also through their reaction field

$$\boldsymbol{\mu}=\mathbf{C}^{-1}\left(\mathbf{T}_{\mathrm{d}}^{(1)}+\mathbf{K}^{(1)}\right)\mathbf{M}=\mathbf{C}^{-1}\left(\mathbf{E}_{\mathrm{d}}+\mathbf{E}_{\mathrm{RF}}\right) \quad (74)$$

The polarization energy can now be described in terms the permanent reaction field and solute field $\mathbf{E}_{\mathrm{p}}$

$$\mathrm{U}_{\mu}=-\frac{1}{2}(\mathbf{E}_{\mathrm{p}}+\mathbf{E}_{\mathrm{RF}})^{t}\boldsymbol{\mu} \quad (75)$$

To find the polarization energy gradient, we wish to avoid terms that rely on the change in induced dipoles with respect to an atomic displacement. Therefore, the induced dipoles in Eq. (75) are substituted for using Eq. (74) to yield

$$\mathrm{U}_{\mu}=-\frac{1}{2}(\mathbf{E}_{\mathrm{p}}+\mathbf{E}_{\mathrm{RF}})^{t}\mathbf{C}^{-1}\left(\mathbf{E}_{\mathrm{d}}+\mathbf{E}_{\mathrm{RF}}\right) \quad (76)$$

By the chain rule, the polarization energy gradient is

$$\frac{\partial \mathrm{U}_{\mu}}{\partial r_{i,\sigma}}=-\frac{1}{2}\left[\left(\frac{\partial \mathbf{E}_{\mathrm{p}}}{\partial r_{i,\sigma}}+\frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}}\right)^{t}\mathbf{C}^{-1}\left(\mathbf{E}_{\mathrm{d}}+\mathbf{E}_{\mathrm{RF}}\right)+(\mathbf{E}_{\mathrm{p}}+\mathbf{E}_{\mathrm{RF}})^{t}\frac{\partial \mathbf{C}^{-1}}{\partial r_{i,\sigma}}\left(\mathbf{E}_{\mathrm{d}}+\mathbf{E}_{\mathrm{RF}}\right)+(\mathbf{E}_{\mathrm{p}}+\mathbf{E}_{\mathrm{RF}})^{t}\mathbf{C}^{-1}\left(\frac{\partial \mathbf{E}_{\mathrm{d}}}{\partial r_{i,\sigma}}+\frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}}\right)\right] \quad (77)$$

For convenience a mathematical quantity **ν** is defined, which is similar to **μ**, as

$$\boldsymbol{\nu} = (\mathbf{E}_{\mathrm{p}} + \mathbf{E}_{\mathrm{RF}})\,\mathbf{C}^{-1} \quad (78)$$

We can now greatly simplify Eq. (77) above using Eqs. (74) and (78) along with the identity $\frac{\partial \mathbf{C}^{-1}}{\partial r_{i,\sigma}} = -\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial r_{i,\sigma}}\mathbf{C}^{-1}$ to give

$$\frac{\partial \mathrm{U}_{\mu}}{\partial r_{i,\sigma}} = -\frac{1}{2}\left[\left(\frac{\partial \mathbf{E}_{\mathrm{p}}}{\partial r_{i,\sigma}}\right)^{t}\boldsymbol{\mu} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{E}_{\mathrm{d}}}{\partial r_{i,\sigma}} + \left(\frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}}\right)^{t}\boldsymbol{\mu} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}} - \boldsymbol{\nu}^{t}\frac{\partial \mathbf{C}}{\partial r_{i,\sigma}}\boldsymbol{\mu}\right] \quad (79)$$

Under the direct polarization model, **C** is an identity matrix whose derivative is zero, and therefore Eq. (79) simplifies to

$$\frac{\partial \mathrm{U}_{\mu_{\mathrm{direct}}}}{\partial r_{i,\sigma}} = -\frac{1}{2}\left[\left(\frac{\partial \mathbf{E}_{\mathrm{p}}}{\partial r_{i,\sigma}}\right)^{t}\boldsymbol{\mu} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{E}_{\mathrm{d}}}{\partial r_{i,\sigma}} + \left(\frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}}\right)^{t}\boldsymbol{\mu} + \boldsymbol{\nu}^{t}\frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}}\right] \quad (80)$$

The first two terms on the RHS appear in the polarization energy gradient even in the absence of a continuum reaction field and are described elsewhere.[35] The third and fourth terms are specific to GK and can be combined. We require the derivative of the GK reaction field due to permanent multipoles with respect to movement of any atom

$$\frac{\partial \mathbf{E}_{\mathrm{RF}}}{\partial r_{i,\sigma}} = \frac{\partial \mathbf{K}^{(1)}}{\partial r_{i,\sigma}}\mathbf{M} \quad (81)$$

It is therefore sufficient to describe the gradient of any $\mathbf{K}_{ij}^{(1)}$ sub-matrix of $\mathbf{K}^{(1)}$ as

$$\frac{\partial \mathbf{K}_{ij}^{(1)}}{\partial r_{i,\sigma}} = \frac{1}{2}\left[\left(\frac{\partial \mathbf{K}_{ij}^{(1,i)}}{\partial r_{i,\sigma}} + \frac{\partial \mathbf{K}_{ij}^{(1,j)}}{\partial r_{i,\sigma}}\right)_{a_i,a_j} + \left(\frac{\partial \mathbf{K}_{ij}^{(1,i)}}{\partial a_i} + \frac{\partial \mathbf{K}_{ij}^{(1,j)}}{\partial a_i}\right)\frac{\partial a_i}{\partial r_{i,\sigma}} + \left(\frac{\partial \mathbf{K}_{ij}^{(1,i)}}{\partial a_j} + \frac{\partial \mathbf{K}_{ij}^{(1,j)}}{\partial a_j}\right)\frac{\partial a_j}{\partial r_{i,\sigma}}\right] \quad (82)$$

The tensors that make up $\frac{\partial \mathbf{K}_{ij}^{(1,i)}}{\partial r_{i,\sigma}}$, $\frac{\partial \mathbf{K}_{ij}^{(1,i)}}{\partial a_i}$, $\frac{\partial \mathbf{K}_{ij}^{(1,i)}}{\partial a_j}$, $\frac{\partial \mathbf{K}_{ij}^{(1,j)}}{\partial r_{i,\sigma}}$, $\frac{\partial \mathbf{K}_{ij}^{(1,j)}}{\partial a_i}$ and $\frac{\partial \mathbf{K}_{ij}^{(1,j)}}{\partial a_j}$ are described in Appendix B. In this case there is a torque on the permanent dipoles and quadrupoles due to the reaction field and reaction field gradient of (μ + ν)/2, respectively.

The full mutual polarization gradient has an additional term compared to the direct polarization gradient, in addition to the implicit difference due to the induced dipoles being converged self-consistently. Specifically, the derivative of the matrix **C** leads to two terms

$$\boldsymbol{\nu}^{t}\frac{\partial \mathbf{C}}{\partial r_{i,\sigma}}\boldsymbol{\mu} = -\boldsymbol{\nu}^{t}\left(\frac{\partial \mathbf{T}^{(11)}}{\partial r_{i,\sigma}} + \frac{\partial \mathbf{K}^{(11)}}{\partial r_{i,\sigma}}\right)\boldsymbol{\mu} \quad (83)$$

The first term on the RHS occurs in vacuum and is described elsewhere[35], however, the final term is specific to GK. The gradient of one sub-matrix of the $\frac{\partial \mathbf{K}^{(11)}}{\partial r_{i,\sigma}}$ supermatrix is

$$\frac{\partial \mathbf{K}_{ij}^{(11)}}{\partial r_{i,\sigma}} = \left(\frac{\partial \mathbf{K}_{ij}^{(11)}}{\partial r_{i,\sigma}} \right)_{a_i, a_j} + \frac{\partial \mathbf{K}_{ij}^{(11)}}{\partial a_i} \frac{\partial a_i}{\partial r_{i,\sigma}} + \frac{\partial \mathbf{K}_{ij}^{(11)}}{\partial a_j} \frac{\partial a_j}{\partial r_{i,\sigma}} \quad (84)$$

The expression for the gradient of $\mathbf{K}_{ij}^{(11)}$ is simpler than those for the other GK interaction matrices because it is symmetric.

The veracity of the AMOEBA/GK energy gradients was checked using finite-differences of the energy, optimization of proteins to an RMS convergence criterion of 10^{-4} kcal/mole/Å and constant energy molecular dynamics. For example, at a mean temperature of 300 K the protein 1ETL showed a mean total energy of −361.20 kcal/mole with a standard deviation of just 0.25 kcal/mole over 1 nsec. The simulation started with a total energy of −361.25 and finished at −365.28 kcal/mole.

5 VALIDATION AND APPLICATION

GK is an approximation to the Poisson solution that extends GB to arbitrary order polarizable atomic multipoles. Here we test GK by comparing to numerical PMPB solutions in the limit of using a van der Waals definition of the solute-solvent interface parameterized using the Bondi radii set.[51] Specifically, the electrostatic solvation free energy and total solvated dipole moment for a series of 55 proteins was compared using the PMPB and GK continuums. This test set based on PDB entries[45] was recently proposed by Tjong and Zhou for studying the accuracy of analytic solvation models and is characterized by structures with less than 10% sequence identity, resolution better than 1.0 Å and less than 250 residues.[44] Amino acids with missing side-chains were changed to alanine if the C_β carbon was present and glycine if not. The TINKER[53] pdbxyz program added missing hydrogen atoms. Histidine residues were made neutral with the δ-nitrogren protonated. All structures were optimized in vacuum to an RMS gradient of 5.0 kcal/mole/Å, with the goal being to remove bad contacts. The average heavy atom RMS distance from the crystal structure was 0.07 Å after optimization.

5.1 Electrostatic solvation free energy of proteins

Previous studies have shown that given accurate effective radii, GB predicts the electrostatic solvation energy of proteins to a mean unsigned relative difference of approximately 1% relative to numerical Poisson calculations.[16] In this section we investigate whether it is reasonable to expect similar performance from GK by comparing the electrostatic solvation free energy for a series of folded proteins to values computed using the PMPB model.

The PMPB calculations used a grid spacing of 0.31 Å and at least 10 Å between the edge of the solute-solvent boundary and the grid boundary. A finer grid spacing of 0.23 Å was also tried, which lowered the PMPB energy by approximately 2%, but did not change the quality of the agreement between the two models. The interior of the protein was assigned a

permittivity of 1.0 while the solvent was set to 78.3. The induced dipoles were deemed to have converged at a tolerance of 0.01 RMS Debye. Converging to a tighter tolerance of 10^{-6} RMS Debye only changed the electrostatic solvation free energy by 0.1% relative to the loser criteria, and was therefore deemed unnecessary. The constant in generalizing function c_f was optimized by hand to eliminate systematic error, which was found to occur at a value of 2.455.

The results are shown in Table 4. The mean signed relative difference is 0.0% a result of tuning the cross-term parameter. The mean unsigned relative difference is 0.9%, which is comparable to the most accurate GB methods.[13,15,17,34,44,54] We anticipate using a different cross-term parameter when optimizing GK to reproduce PMPB calculations based on a molecular surface definition.

5.2 Dipole moment of solvated proteins

The change in dipole moment as a function of environment for a polarizable solute is a relevant observable in terms of validating GK because it indicates whether or not the reaction field strength is consistent. The PMPB calculations are exactly equivalent to those described in the previous section. Furthermore, the same constant was used in the GK cross-term. In Table 5 it is observed that the total dipole moment of proteins within the GK continuum achieve a mean signed relative difference of −2.7% and a mean unsigned percent difference of 2.7%. This indicates a small, but systematic underestimation of the reaction field. In all cases, for both PMPB and GK models, the reaction field factor was greater than one, except for 1P9G. In this case, the vacuum dipole moment decreased from 18 to 15 and 13 Debye in the PMPB and GK models, respectively. Overall, the mean reaction field factor for the 55 proteins was 1.28 in the PMPB model and 1.24 in GK.

6 CONCLUSIONS

Over the course of the past several years GB has been shown to be capable of capturing the electrostatic response of the solvent environment to solutes. It has been successfully applied to molecular dynamics simulations, scoring protein conformations and the prediction of binding affinities.[24] However, GB models are generally limited to use with fixed atomic partial charge electrostatic representations. Applications of recent interest, including high-resolution homology modeling, design of protein-protein interactions and design of proteins with enzymatic activity may require improved accuracy in force field electrostatics.[28,29,55] We suggest that the AMOEBA force field coupled with the GK continuum model is a promising improvement.

There are two main differences between GB and GK. First, the GK self-energy of a permanent multipole site depends on Kirkwood's solution for the electrostatic solvation energy of a spherical particle with arbitrary charge distribution, which is reduced to Born's formula in the case of a monopole. Second, the GK cross-term is formulated by averaging a simple auxiliary potential for each multipole site, which reduces to the GB cross-term for monopole interactions.

We have implemented GK for the AMOEBA force field, including energy gradients, within the TINKER package.[53] The model was tested against numerical PMPB calculations of the electrostatic solvation free energy for a series of 55 diverse proteins and showed a mean unsigned relative difference of 0.9%. The fidelity of the reaction field of GK relative to PMPB can be inferred from the total solvated dipole moment of each protein, which showed GK to have a mean unsigned relative difference of 2.7%.

The next step in the implementation of GK for AMOEBA solutes will be parameterization of a complete implicit solvent model by addition of an apolar term.[26] The overall model will be parameterized against small molecules solvation free energies, which has been a successful approach in the past.[9,10,25,56,57] Alternatively, the electrostatic, dispersion and cavitation components of solvation can be matched to explicit solvent free energy perturbation results.[58–61]

GK may be useful for developing new continuum models based on electron densities derived from electronic structure calculations. For example, Cramer and Truhlar have successfully employed GB in their SMX series of solvation models.[8,9,62,63] GK would also offer an analytic alternative to the numerical distributed multipole solvation model of Rinaldi *et al.*[64,65]

Further improvements in both the PMPB and GK continuum electrostatics models may depend on reconciling deficiencies that emerge in treating local, specific molecular interactions. For example, both the Clausius-Mossotti[40] and Onsager[66] theories for the permittivity of a liquid break down for substances that “associate”, such as water. Here association is defined as short range ordering that leads to correlations in the orientations and positions of neighboring groups, such as hydrogen bonding pairs. Formalisms introduced by Kirkwood[67] and Fröhlich[68] include a correction factor to explicitly account for this deviation from continuum behavior. More recently, Rick and Berne showed that no parameterization of the dielectric boundary for a water molecule in liquid water could simultaneously fit the electrostatic free energy and reaction potential to within 20%, mainly due to nonlinear electrostriction.[69] This effect, inherent in both numerical and analytic continuum electrostatic models, may be a limiting factor in the accuracy of current implicit solvation models.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Dr. Pengyu Ren, Dr. Nathan Baker, Dr. Chuanjie Wu, Dr. Sergio Urahata and Justin Xiang for helpful discussions. MJS was partially supported by a Computational Biology Training Grant from the NIH and a Grace Norman Scholarship. JWP was supported by NSF grants CHE-0535675 and MCB-0344670, and NIH grant GM069553.

REFERENCES

1. Kirkwood JG. J. Chem. Phys. 1934; 2:351–361.

2. Schnieders MJ, Baker NA, Ren PY, Ponder JW. J. Chem. Phys. 2007:126.
3. Flory, PJ. Statistical Mechanics Of Chain Molecules. Butterworth-Heinemann Ltd; 1969.
4. Born M. Z. Phys. 1920; 1:45–48.
5. Kong Y, Ponder JW. J. Chem. Phys. 1997; 107:481–492.
6. Schaefer M, Karplus M. J. Phys. Chem. 1996; 100:1578–1599.
7. Schaefer M, Froemmel C. J. Mol. Biol. 1990; 216:1045–1066. [PubMed: 2266555]
8. Hawkins GD, Cramer CJ, Truhlar DG. Chem. Phys. Lett. 1995; 246:122–129.
9. Hawkins GD, Cramer CJ, Truhlar DG. J. Phys. Chem. 1996; 100:19824–19839.
10. Still WC, Tempczyk A, Hawley RC, Hendrickson T. J. Am. Chem. Soc. 1990; 112:6127–6129.
11. Qiu D, Shenkin PS, Hollinger FP, Still WC. J. Phys. Chem. A. 1997; 101:3005–3014.
12. Feig M, Im W, Brooks CL. J. Chem. Phys. 2004; 120:903–911. [PubMed: 15267926]
13. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL. J. Comput. Chem. 2004; 25:265–284. [PubMed: 14648625]
14. Onufriev A, Bashford D, Case DA. J. Phys. Chem. B. 2000; 104:3712–3720.
15. Onufriev A, Bashford D, Case DA. Proteins. 2004; 55:383–394. [PubMed: 15048829]
16. Onufriev A, Case DA, Bashford D. J. Comput. Chem. 2002; 23:1297–1304. [PubMed: 12214312]
17. Sigalov G, Fenley A, Onufriev A. J. Chem. Phys. 2006:124.
18. Sigalov G, Scheffel P, Onufriev A. J. Chem. Phys. 2005:122.
19. Baker NA. Methods Enzymol. 2004; 383:94–118. [PubMed: 15063648]
20. Baker NA. Curr. Opin. Struc. Biol. 2005; 15:137–143.
21. Holst M, Saied F. J. Comput. Chem. 1993; 14:105–113.
22. Nicholls A, Honig B. J. Comput. Chem. 1991; 12:435–445.
23. Lee MS, Salsbury FR, Brooks CL. J. Chem. Phys. 2002; 116:10606–10614.
24. Feig M, Brooks CL 3rd. Curr. Opin. Struc. Biol. 2004; 14:217–224.
25. Gallicchio E, Zhang LY, Levy RM. J. Comput. Chem. 2002; 23:517–529. [PubMed: 11948578]
26. Roux B, Simonson T. Biophys. Chem. 1999; 78:1–20. [PubMed: 17030302]
27. Maple JR, Cao YX, Damm WG, Halgren TA, Kaminski GA, Zhang LY, Friesner RA. J. Chem. Theory Comput. 2005; 1:694–715. [PubMed: 26641692]
28. Jaramillo A, Wodak SJ. Biophys. J. 2005; 88:156–171. [PubMed: 15377512]
29. Marshall SA, Vizcarra CL, Mayo SL. Protein Sci. 2005; 14:1293–1304. [PubMed: 15802649]
30. Piquemal JP, Cisneros GA, Reinhardt P, Gresh N, Darden TA. J. Chem. Phys. 2006:124.
31. Friesner, RA.; Robert, L.; Baldwin, aDB. Advances in Protein Chemistry. Vol. 72. Academic Press; 2005. p. 79-104.
32. Ponder, JW.; Case, DA. Advances in Protein Chemistry. Vol. 66. Academic Press; 2003. p. 27-85.
33. Bashford D, Case DA. Annu. Rev. Phys. Chem. 2000; 51:129–152. [PubMed: 11031278]
34. Gallicchio E, Levy RM. J. Comput. Chem. 2004; 25:479–499. [PubMed: 14735568]
35. Ren PY, Ponder JW. J. Comput. Chem. 2002; 23:1497–1506. [PubMed: 12395419]
36. Ren PY, Ponder JW. J. Phys. Chem. B. 2003; 107:5933–5947.
37. Ren PY, Ponder JW. J. Phys. Chem. B. 2004; 108:13427–13437.
38. Grycuk T. J. Chem. Phys. 2003; 119:4817–4826.
39. Jackson, JD. Classical Electrodynamics. 3rd. New York: John Wiley & Sons, Inc.; 1998.
40. Böttcher, CJF. Dielectrics in Static Fields. 2. Vol. 1. Amsterdam: Elsevier Pub. Co.; 1993.
41. Böttcher, CJF. Dielectrics in Static Fields. 1. Vol. 1. Amsterdam: Elsevier Pub. Co.; 1952.
42. Stone, AJ. The Theory of Intermolecular Forces. Vol. 32. Oxford: Clarendon Press; 1996.
43. Tanford C, Kirkwood JG. J. Am. Chem. Soc. 1957; 79:5333–5339.
44. Tjong H, Zhou HX. J. Phys. Chem. B. 2007
45. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]
46. Teeter MM. Proc. Natl. Acad. Sci. 1984; 81:6014–6018. [PubMed: 16593516]

47. Clarke ND, Kissinger CR, Desjarlais J, Gilliland GL, Pabo CO. Protein Sci. 1994; 3:1779–1787. [PubMed: 7849596]
48. Dahiyat BI, Mayo SL. Science. 1997; 278:82–87. [PubMed: 9311930]
49. Gallagher T, Alexander P, Bryan P, Gilliland GL. Biochemistry. 1994; 33:4721–4729. [PubMed: 8161530]
50. McKnight CJ, Matsudaira PT, Kim PS. Nat. Struct. Biol. 1997; 4:180–184. [PubMed: 9164455]
51. Bondi A. J. Phys. Chem. 1964; 68:441–451.
52. Young, DM. Iterative Solutions of Large Linear Systems. New York: Acedemic Press; 1971.
53. Ponder, JW. TINKER: Software Tools for Molecular Design. 4.2. Saint Louis, MO: 2004.
54. Lee MS, Feig M, Salsbury FR, Brooks CL. J. Comput. Chem. 2003; 24:1348–1356. [PubMed: 12827676]
55. Vizcarra CL, Mayo SL. Curr. Opin. Chem. Biol. 2005; 9:622–626. [PubMed: 16257567]
56. Jorgensen WL, Ulmschneider JP, Tirado-Rives J. J. Phys. Chem. B. 2004; 108:16264–16270.
57. Barone V, Cossi M, Tomasi J. J. Chem. Phys. 1997; 107:3210–3221.
58. Banavali NK, Im W, Roux B. J. Chem. Phys. 2002; 117:7381–7388.
59. Gallicchio E, Kubo MM, Levy RM. J. Phys. Chem. B. 2000; 104:6271–6285.
60. Nina M, Beglov D, Roux B. J. Phys. Chem. B. 1997; 101:5239–5248.
61. Nina M, Im W, Roux B. Biophys. Chem. 1999; 78:89–96. [PubMed: 17030305]
62. Hawkins GD, Cramer CJ, Truhlar DG. J. Phys. Chem. B. 1998; 102:3257–3271.
63. Kelly CP, Cramer CJ, Truhlar DG. J. Chem. Theory Comput. 2005; 1:1133–1152. [PubMed: 26631657]
64. Rinaldi D, Bouchy A, Rivail JL. Theor. Chem. Acc. 2006; 116:664–669.
65. Rinaldi D, Bouchy A, Rivail JL, Dillet V. J. Chem. Phys. 2004; 120:2343–2350. [PubMed: 15268373]
66. Onsager L. J. Am. Chem. Soc. 1936; 58:1486–1493.
67. Kirkwood JG. J. Chem. Phys. 1939; 7:911–919.
68. Fröhlich, H. Theory of Dielectrics. 2. London: Oxford University Press; 1958.
69. Rick SW, Berne BJ. J. Am. Chem. Soc. 1994; 116:3949–3954.
70. Challacombe M, Schwegler E, Almlof J. Chem. Phys. Lett. 1995; 241:67–72.
71. McMurchie LE, Davidson ER. J. Comput. Phys. 1978; 26:218–231.
72. Applequist J. J. Phys. A: Math. Gen. 1989; 22:4303–4330.
73. Applequist J. Theor. Chem. Acc. 2002; 107:103–115.

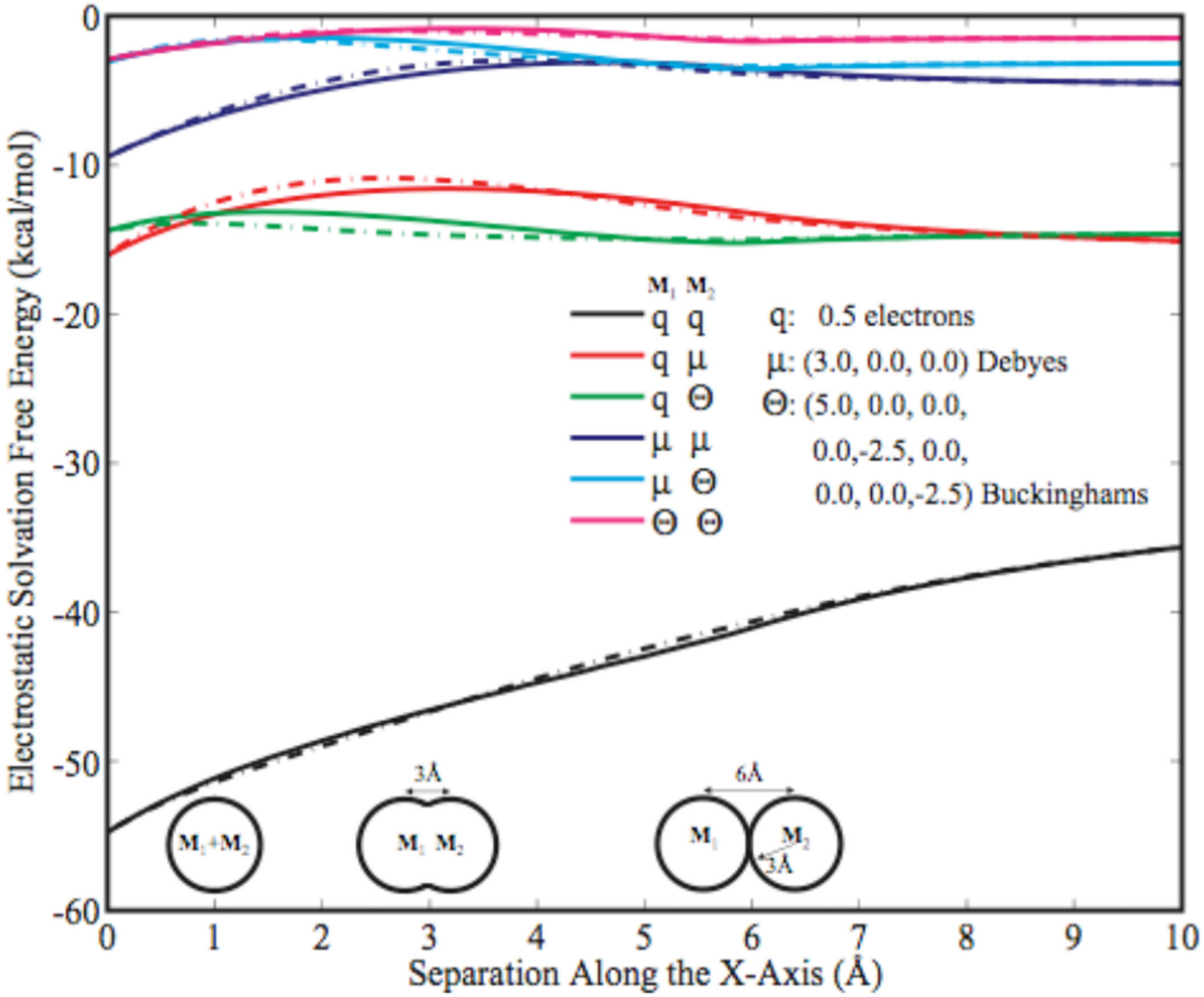

Figure 1.

The solvation energy for a system composed two spheres, each with a radius of 3 Å and permittivity of 1, and a variety of multipole combinations are computed as a function of separation along the x-axis using numerical Poisson solutions (solid lines) and generalized Kirkwood (dashed lines). The solvent permittivity was 78.3. The limiting cases of wide separation and superimposition are reproduced in all cases, while intermediate separations are seen to be a reasonable approximation.

Table 1

Indefinite integrals for pairwise descreening of multipoles through quadrupole

$(l, m)_1$	$(l, m)_2$	$D_{(l, m)_i, (l, m)_i}(r_{ij}, R_j)$
(0, 0)	(0, 0)	$-(2\ \ln(r)r^2+4r_{ij}r - r_{ij}^2+R_j^2)/16r_{ij}r^2$
	(1, 0)	$-(4r^4\ \ln(r)+4r_{ij}^2r^2+4r^2R_j^2 - r_{ij}^4+2r_{ij}^2R_j^2 - R_j^4)/64r^4r_{ij}^2$
	(2, 0)	$-(12r^6\ \ln(r)+6r_{ij}^2r^4+18r^4R_j^2+3r^2r_{ij}^4\ +6r^2r_{ij}^2R_j^2-9r^2R_j^4-2r_{ij}^6+6r_{ij}^4R_j^2-6r_{ij}^2R_j^4+2R_j^6)/256r_{ij}^3r^6$
(1, 0)	(1, 0)	$-(12r^6\ \ln(r)-42r_{ij}^2r^4+18r^4R_j^2+64r^3r_{ij}^3-21r^2r_{ij}^4\ +30r^2r_{ij}^2R_j^2-9r^2R_j^4-2r_{ij}^6+6r_{ij}^4R_j^2-6r_{ij}^2R_j^4+2R_j^6)/384r^6r_{ij}^3$
	(2, 0)	$-(24r^8\ \ln(r) - 48r^6r_{ij}^2+48r^6R_j^2+60r_{ij}^4r^4$ $+72r_{ij}^2r^4R_j^2 - 36r^4R_j^4 - 16r^2r_{ij}^6+48r^2r_{ij}^4R_j^2 - 48r^2r_{ij}^2R_j^4$ $+16r^2R_j^6 - 3r_{ij}^8+12r_{ij}^6R_j^2 - 18r_{ij}^4R_j^4+12r_{ij}^2R_j^6 - 3R_j^8)/1024r^8r_{ij}^4$
(1, 1) (1, −1)	(1, 1) (1, −1)	$(12r^6\ \ln(r)+102r_{ij}^2r^4+18r^4R_j^2-128r^3r_{ij}^3+51r^2r_{ij}^4\ -42r^2r_{ij}^2R_j^2-9r^2R_j^4-2r_{ij}^6+6r_{ij}^4R_j^2-6r_{ij}^2R_j^4+2R_j^6)/768r^6r_{ij}^3$
	(2, 1) (2, −1)	$\sqrt{3}(24r^8\ \ln(r)+96r^6r_{ij}^2+48r^6R_j^2 - 84r_{ij}^4r^4$ $- 72r_{ij}^2r^4R_j^2 - 36r^4R_j^4+32r^2r_{ij}^6 - 48r^2r_{ij}^4R_j^2$ $+16r^2R_j^6 - 3r_{ij}^8+12r_{ij}^6R_j^2 - 18r_{ij}^4R_j^4+12r_{ij}^2R_j^6 - 3R_j^8)/3072r^8r_{ij}^4$
(2, 0)	(2, 0)	$- 3(120r^{10}\ \ln(r) - 140r^8r_{ij}^2+300r^8R_j^2 - 540r^6r_{ij}^4+360r^6r_{ij}^2R_j^2$ $- 300r^6R_j^4+1024r^5r_{ij}^5 - 360r^4r_{ij}^6+600r^4r_{ij}^4R_j^2 - 440r^4r_{jj}^2R_j^4$ $+200r^4R_j^6-35r^2r_{ij}^8+180r^2r_{ij}^6R_j^2-330r^2r_{ij}^4R_j^4+260r^2r_{ij}^2R_j^6-75r^2R_j^8$ $- 12r_{ij}^{10}+60r_{ij}^8R_j^2 - 120r_{ij}^6R_j^4+120r_{ij}^4R_j^6 - 60r_{ij}^2R_j^8+12R_j^{10})/20480r^{10}r_{ij}^5$
(2, 1) (2, −1)	(2, 1) (2, −1)	$(120r^{10}\ \ln(r)+180r^8r_{ij}^2+300r^8R_j^2+900r^6r_{ij}^4 - 120r^6r_{ij}^2R_j^2$ $- 300r^6R_j^4 - 1536r^5r_{ij}^5+600r^4r_{ij}^6 - 680r^4r_{ij}^4R_j^2 - 120r^4r_{ij}^2R_j^4$ $+200r^4R_j^6+45r^2r_{ij}^8 - 60r^2r_{ij}^6R_j^2 - 90r^2r_{ij}^4R_j^4+180r^2r_{ij}^2R_j^6 - 75r^2R_j^8$ $- 12r_{ij}^{10}+60r_{ij}^8R_j^2 - 120r_{ij}^6R_j^4+120r_{ij}^4R_j^6 - 60r_{ij}^2R_j^8+12R_j^{10})/10240r^{10}r_{ij}^5$
(2, 2) (2, −2)	(2, 2) (2, −1)	$- (120r^{10}\ \ln(r)+1140r^8r_{ij}^2+300r^8R_j^2 - 4380r^6r_{ij}^4 - 1560r^6r_{ij}^2R_j^2$ $- 300r^6R_j^4+6144r^5r_{ij}^5 - 2920r^4r_{ij}^6+1880r^4r_{ij}^4R_j^2+840r^4r_{ij}^2R_j^4$ $+200r^4R_j^6+285r^2r_{ij}^8-780r^2r_{ij}^6R_j^2+630r^2r_{ij}^4R_j^4-60r^2r_{ij}^2R_j^6-75r^2R_j^8$ $- 12r_{ij}^{10}+60r_{ij}^8R_j^2 - 120r_{ij}^6R_j^4+120r_{ij}^4R_j^6 - 60r_{ij}^2R_j^8+12R_j^{10})/40960r^{10}r_{ij}^5$

Table 2

Indefinite integrals for pairwise descreening of multipoles through quadrupole when $\xi_{ij} = \pi$.

l_i	D_{l_i}
0	$-\frac{1}{2r}$
1	$-\frac{1}{3r^3}$
2	$-\frac{3}{10r^5}$

Table 3

Shown is a comparison of the performance of the SFA and RPA in determining the perfect self-energy (kcal/mole) for a series of five folded proteins. Optimization of a single HCT scale factor for each method removes systematic error as shown by the mean signed percent differences. However, the mean RPA unsigned percent difference of 0.5 is smaller than that of the SFA.

	Self-Energy			Signed % Difference		Unsigned % Difference	
	PMPB	**SFA**	**RPA**	**SFA**	**RPA**	**SFA**	**RPA**
CRN	−8141	−8191	−8196	−0.6	−0.7	0.6	0.7
ENH	−11919	−11852	−11878	0.6	0.3	0.6	0.3
FSV	−6254	−6341	−6287	−1.4	−0.5	1.4	0.5
PGB	−11794	−11743	−11803	0.4	−0.1	0.4	0.1
VII	−7206	−7132	−7133	1.0	1.0	1.0	1.0
Mean				0.0	0.0	0.8	0.5

J Chem Theory Comput. Author manuscript; available in PMC 2016 February 25.

Table 4

The electrostatic solvation free energy (kcal/mole) for 55 proteins within the PMPB and GK continuum models. The number of atoms and total charge of each protein is listed along with the signed and unsigned relative difference of the GK model to PMPB.

			Energy		% Difference	
	N_{atoms}	Q	PMPB	GK	Signed	Unsigned
1A6M	2435	2	−2831	−2765	2.3	2.3
1AHO	936	0	−1161	−1158	0.3	0.3
1BYI	3383	−4	−3861	−3873	−0.3	0.3
1C75	985	−6	−1733	−1742	−0.5	0.5
1C7K	1927	−5	−2523	−2481	1.7	1.7
1CEX	2867	1	−3161	−3212	−1.6	1.6
1EB6	2566	−15	−5044	−5042	0.1	0.1
1EJG	642	0	−580	−614	−6.0	6.0
1ETL	140	−1	−246	−247	−0.5	0.5
1EXR	2240	−25	−8656	−8620	0.4	0.4
1F94	967	2	−1240	−1226	1.1	1.1
1F9Y	2535	−5	−2964	−2968	−0.2	0.2
1G4I	1842	−1	−2356	−2345	0.5	0.5
1G66	2794	−2	−2826	−2824	0.1	0.1
1GQV	2135	7	−2708	−2723	−0.6	0.6
1HJE	175	1	−264	−269	−2.0	2.0
1IQZ	1171	−17	−4663	−4729	−1.4	1.4
1IUA	1207	−1	−1400	−1419	−1.4	1.4
1J0P	1597	8	−2975	−2934	1.4	1.4
1K4I	3253	−6	−4085	−4099	−0.3	0.3
1KTH	885	0	−1469	−1448	1.4	1.4
1L9L	1226	11	−3182	−3150	1.0	1.0
1M1Q	1236	−4	−2084	−2077	0.3	0.3
1NLS	3564	−7	−4743	−4756	−0.3	0.3
1NWZ	1912	−6	−2768	−2760	0.3	0.3
1OD3	1893	−3	−2105	−2104	0.0	0.0

J Chem Theory Comput. Author manuscript; available in PMC 2016 February 25.

			Energy		% Difference	
	N_{atoms}	Q	PMPB	GK	Signed	Unsigned
1OK0	1076	−5	−1578	−1571	0.5	0.5
1P9G	519	4	−814	−817	−0.4	0.4
1PQ7	3065	4	−2946	−2942	0.1	0.1
1R6J	1230	0	−1486	−1477	0.6	0.6
1SSX	2755	8	−3000	−2980	0.7	0.7
1TG0	1029	−12	−3017	−3014	0.1	0.1
1TQG	1660	−7	−2920	−2900	0.7	0.7
1TT8	2676	1	−2762	−2758	0.1	0.1
1U2H	1495	2	−2038	−2002	1.8	1.8
1UCS	997	0	−1027	−1042	−1.4	1.4
1UFY	1911	0	−2130	−2145	−0.7	0.7
1UNQ	1947	−1	−3217	−3155	1.9	1.9
1VB0	913	3	−1246	−1232	1.1	1.1
1VBW	1056	8	−1931	−1927	0.2	0.2
1W0N	1756	−5	−2380	−2356	1.0	1.0
1WY3	560	1	−750	−747	0.3	0.3
1X6Z	1720	−1	−2170	−2198	−1.3	1.3
1X8Q	2815	−1	−3739	−3714	0.7	0.7
1XMK	1268	1	−1723	−1724	0.0	0.0
1YK4	774	−8	−1893	−1920	−1.4	1.4
1ZZK	1243	1	−1730	−1699	1.8	1.8
2A6Z	3430	−3	−4203	−4186	0.4	0.4
2BF9	560	−2	−933	−940	−0.8	0.8
2CHH	1624	−3	−2128	−2131	−0.1	0.1
2CWS	3400	−3	−3651	−3616	1.0	1.0
2ERL	567	−6	−1178	−1179	0.0	0.0
2FDN	731	−8	−1746	−1796	−2.9	2.9
2FWH	1830	−6	−2495	−2502	−0.3	0.3
3LZT	1960	8	−2754	−2723	1.1	1.1

			Energy		% Difference	
	N_{atoms}	Q	PMPB	GK	Signed	Unsigned
Mean	1692	−1.9	−2458	−2454	0.0	0.9

Table 5

The total dipole moment (Debye) for 55 proteins in vacuum and within the PMPB and GK continuum models. The signed and unsigned relative difference of the GK model to PMPB is given along with their reaction field factors.

	Dipole Moment			% Difference		Reaction Field Factor	
	Vacuum	PMPB	GK	Signed	Unsigned	PMPB	GK
1A6M	191.5	252.1	242.6	−3.7	3.7	1.32	1.27
1AHO	119.3	143.6	142.6	−0.7	0.7	1.20	1.20
1BYI	295.8	357.4	343.1	−4.0	4.0	1.21	1.16
1C75	125.0	167.2	165.7	−0.9	0.9	1.34	1.33
1C7K	229.3	310.3	302.7	−2.4	2.4	1.35	1.32
1CEX	451.0	599.7	574.3	−4.2	4.2	1.33	1.27
1EB6	217.9	281.0	274.6	−2.3	2.3	1.29	1.26
1EJG	37.4	49.0	48.9	−0.3	0.3	1.31	1.31
1ETL	29.3	42.9	41.2	−3.8	3.8	1.46	1.41
1EXR	352.5	395.6	384.2	−2.9	2.9	1.12	1.09
1F94	90.7	116.7	113.0	−3.2	3.2	1.29	1.25
1F9Y	138.4	166.0	161.9	−2.5	2.5	1.20	1.17
1G4I	87.9	102.1	97.9	−4.1	4.1	1.16	1.11
1G66	226.5	279.9	273.5	−2.3	2.3	1.24	1.21
1GQV	314.6	394.5	385.2	−2.4	2.4	1.25	1.22
1HJE	48.3	61.2	60.4	−1.4	1.4	1.27	1.25
1IQZ	86.1	110.7	107.2	−3.1	3.1	1.29	1.25
1IUA	107.5	146.1	141.5	−3.2	3.2	1.36	1.32
1J0P	105.2	148.7	142.3	−4.3	4.3	1.41	1.35
1K4I	130.1	163.0	159.6	−2.1	2.1	1.25	1.23
1KTH	117.1	152.1	148.9	−2.1	2.1	1.30	1.27
1L9L	422.8	525.9	517.0	−1.7	1.7	1.24	1.22
1M1Q	261.7	318.1	311.2	−2.2	2.2	1.22	1.19
1NLS	244.9	331.8	313.0	−5.7	5.7	1.35	1.28
1NWZ	83.2	130.2	126.9	−2.5	2.5	1.56	1.53
1OD3	115.2	165.9	160.9	−3.0	3.0	1.44	1.40

J Chem Theory Comput. Author manuscript; available in PMC 2016 February 25.

	Dipole Moment			% Difference		Reaction Field Factor	
	Vacuum	PMPB	GK	Signed	Unsigned	PMPB	GK
1OK0	149.4	193.7	189.1	−2.4	2.4	1.30	1.27
1P9G	17.7	14.6	13.0	−10.7	10.7	0.82	0.74
1PQ7	46.4	49.6	49.1	−1.1	1.1	1.07	1.06
1R6J	86.8	108.8	106.7	−1.9	1.9	1.25	1.23
1SSX	66.0	93.8	89.9	−4.2	4.2	1.42	1.36
1TG0	236.9	316.8	311.1	−1.8	1.8	1.34	1.31
1TQG	355.4	489.5	477.3	−2.5	2.5	1.38	1.34
1TT8	339.6	450.3	434.3	−3.6	3.6	1.33	1.28
1U2H	157.1	206.0	200.6	−2.6	2.6	1.31	1.28
1UCS	111.1	133.0	132.9	0.0	0.0	1.20	1.20
1UFY	94.0	105.9	102.3	−3.4	3.4	1.13	1.09
1UNQ	601.1	735.2	718.8	−2.2	2.2	1.22	1.20
1VB0	132.2	158.2	155.0	−2.0	2.0	1.20	1.17
1VBW	94.4	117.0	114.0	−2.6	2.6	1.24	1.21
1W0N	114.9	155.4	150.0	−3.5	3.5	1.35	1.31
1WY3	63.7	96.4	93.6	−3.0	3.0	1.51	1.47
1X6Z	294.2	366.7	355.9	−2.9	2.9	1.25	1.21
1X8Q	183.8	244.2	237.6	−2.7	2.7	1.33	1.29
1XMK	272.8	356.1	347.0	−2.6	2.6	1.31	1.27
1YK4	66.1	83.7	83.6	−0.2	0.2	1.27	1.26
1ZZK	195.2	246.5	241.6	−2.0	2.0	1.26	1.24
2A6Z	84.1	105.0	101.4	−3.4	3.4	1.25	1.21
2BF9	255.7	290.6	288.4	−0.7	0.7	1.14	1.13
2CHH	267.4	335.7	329.2	−1.9	1.9	1.26	1.23
2CWS	168.6	220.5	211.0	−4.3	4.3	1.31	1.25
2ERL	81.2	108.1	105.6	−2.3	2.3	1.33	1.30
2FDN	78.3	93.2	93.4	0.3	0.3	1.19	1.19
2FWH	104.9	146.3	142.9	−2.3	2.3	1.39	1.36
3LZT	178.5	214.6	209.8	−2.3	2.3	1.20	1.18

	Dipole Moment			% Difference		Reaction Field Factor	
	Vacuum	PMPB	GK	Signed	Unsigned	PMPB	GK
Mean	173.2	220.9	215.0	−2.7	2.7	1.28	1.24