

## Implicit Solvents for the Polarizable Atomic Multipole AMOEBA Force Field

Rae A. Corrigan,<sup>#</sup> Guowei Qi,<sup>#</sup> Andrew C. Thiel, Jack R. Lynn, Brandon D. Walker, Thomas L. Casavant, Louis Lagardere, Jean-Philip Piquemal, Jay W. Ponder, Pengyu Ren, and Michael J. Schnieders\*Cite This: *J. Chem. Theory Comput.* 2021, 17, 2323–2341

Read Online

ACCESS |



Metrics &amp; More



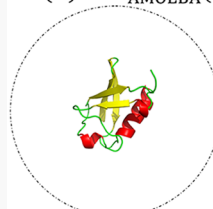
Article Recommendations



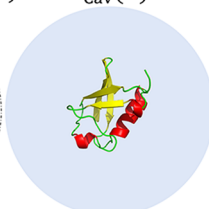
Supporting Information

**ABSTRACT:** Computational protein design, ab initio protein/RNA folding, and protein–ligand screening can be too computationally demanding for explicit treatment of solvent. For these applications, implicit solvent offers a compelling alternative, which we describe here for the polarizable atomic multipole AMOEBA force field based on three treatments of continuum electrostatics: numerical solutions to the nonlinear and linearized versions of the Poisson–Boltzmann equation (PBE), the domain-decomposition conductor-like screening model (ddCOSMO) approximation to the PBE, and the analytic generalized Kirkwood (GK) approximation. The continuum electrostatics models are combined with a nonpolar estimator based on novel cavitation and dispersion terms. Electrostatic model parameters are numerically optimized using a least-squares style target function based on a library of 103 small-molecule solvation free energy differences. Mean signed errors for the adaptive Poisson–Boltzmann solver (APBS), ddCOSMO, and GK models are 0.05, 0.00, and 0.00 kcal/mol, respectively, while the mean unsigned errors are 0.70, 0.63, and 0.58 kcal/mol, respectively. Validation of the electrostatic response of the resulting implicit solvents, which are available in the Tinker (or Tinker-HP), OpenMM, and Force Field X software packages, is based on comparisons to explicit solvent simulations for a series of proteins and nucleic acids. Overall, the emergence of performative implicit solvent models for polarizable force fields opens the door to their use for folding and design applications.

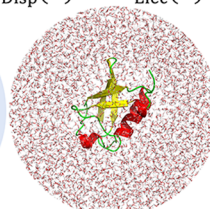
$$W(\mathbf{X}) = U_{\text{AMOEBA}}(\mathbf{X}) + \Delta W_{\text{Cav}}(\mathbf{X}) + \Delta W_{\text{Disp}}(\mathbf{X}) + \Delta W_{\text{Elec}}(\mathbf{X})$$



Vacuum

 $\mu_{\text{vacuum}}$ : 208  
(Debye)

Implicit Solvent

 $\mu_{\text{APBS}}$ : 273  
 $\mu_{\text{ddCOSMO}}$ : 274  
 $\mu_{\text{GK}}$ : 272

Explicit Solvent

 $\langle \mu_{\text{explicit}} \rangle$ : 271

## INTRODUCTION

Solvation plays a key role in accurately portraying the natural processes of molecules *in vitro* and *in vivo*.<sup>1</sup> Hydrophilic and hydrophobic interactions govern protein folding<sup>2</sup> and impact molecular recognition.<sup>3</sup> For these reasons, solvent<sup>4</sup> must be considered during computational protein design and optimization,<sup>5–9</sup> RNA folding,<sup>10,11</sup> and biocatalyst design.<sup>12</sup> While explicit solvent often provides a more complete depiction of solvation effects on molecular interactions, its use can become impractical for biomolecular folding and design applications. To help alleviate this computational expense, implicit solvation models have been developed.<sup>13,14</sup>

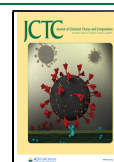
Implicit solvents are designed to replicate explicit solvent while treating water as a continuum to avoid the cost of calculating the interactions of thousands of individual water molecules. The total implicit solvent potential of mean force  $\Delta W_{\text{hydration}}(\mathbf{X})$  as a function of atomic coordinates  $\mathbf{X}$  can be formulated as a sum of cavitation, dispersion, and electrostatic contributions

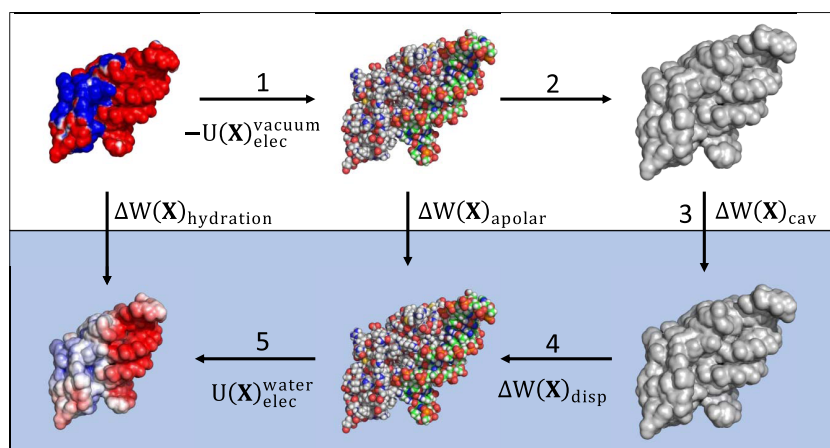
$$\Delta W_{\text{hydration}}(\mathbf{X}) = \Delta W_{\text{cav}}(\mathbf{X}) + \Delta W_{\text{disp}}(\mathbf{X}) + \Delta W_{\text{elec}}(\mathbf{X}) \quad (1)$$

where  $\Delta W_{\text{cav}}$  is the unfavorable cost of forming a solute-shaped cavity within solvent,  $\Delta W_{\text{disp}}$  is the favorable contribution of including solute–solvent van der Waals interactions, and  $\Delta W_{\text{elec}}$  captures the difference between charging the molecule in solvent and in vacuum environments (Figure 1).<sup>13,15,16</sup> Collectively, cavitation and dispersion are termed the nonpolar contribution<sup>17–21</sup> to solvation free energy, while the electrostatic term is referred to as the polar contribution.<sup>4,22–27</sup> For the latter, previous widely used implicit solvents for biomolecules include approaches based on numerical solutions to the Poisson–Boltzmann equation (PBE)<sup>26,28–31</sup> and the analytic generalized Born (GB) approximation.<sup>32–40</sup> The majority are built upon fixed partial charge force fields that maintain constant dipole moments across vacuum and aqueous environments. On the other hand, the family of implicit solvents described here is parameterized for use with the polarizable atomic multipole

Received: December 11, 2020

Published: March 26, 2021





**Figure 1.** Total implicit solvent potential of mean force  $\Delta W_{\text{hydration}}(\mathbf{X})$  can be formulated using a thermodynamic cycle composed of five steps. Step 1: the solute is discharged in vacuum  $-U_{\text{elec}}^{\text{vacuum}}(\mathbf{X})$ . Step 2: dispersion interactions are removed between the solute and the surrounding medium, which has no energetic cost in vacuum. Step 3: a solute-shaped cavity is formed in water  $\Delta W_{\text{cav}}(\mathbf{X})$ , which is unfavorable and proportional to the solvent-excluded volume for small solutes. Step 4: favorable solute-solvent dispersion interactions are added  $\Delta W_{\text{disp}}(\mathbf{X})$ . Step 5: the solute is charged in solvent  $U_{\text{elec}}^{\text{water}}(\mathbf{X})$  to yield an overall electrostatic contribution of  $\Delta W_{\text{elec}}(\mathbf{X}) = U_{\text{elec}}^{\text{water}}(\mathbf{X}) - U_{\text{elec}}^{\text{vacuum}}(\mathbf{X})$ .

AMOEBA force field.<sup>41,42</sup> These models combine intramolecular solute polarization with the electrostatic response of the dielectric continuum via a self-consistent reaction field (SCRf) that leverages numerical solutions to the PBE<sup>43,44</sup> or the much faster analytic generalized Kirkwood (GK) approximation<sup>45</sup> (GK extends GB to polarizable multipoles using work by Kirkwood<sup>46</sup>). In principle, polarizable biomolecular charge distributions (*i.e.*, induced dipoles for the AMOEBA model) are then able to respond to both low-dielectric (*e.g.*, benzene or carbon tetrachloride) and high-dielectric (*e.g.*, methanol or water) environments.

Efforts to combine polarizable biomolecular force fields with implicit solvents began in the ~2000s with the introduction of the polarizable force field (PFF) and its initial application to protein-ligand interactions.<sup>47</sup> The PFF defines solute electrostatics using permanent atomic multipoles (through dipole order) and induced dipoles, while the PBE was solved using a finite element mesh.<sup>48</sup> A more recent example combined a Drude oscillator force field<sup>49,50</sup> with numerical solutions of the PBE.<sup>51</sup> Application of this model to  $pK_a$  prediction<sup>52</sup> showed superior accuracy relative to the additive CHARMM36 force field,<sup>53,54</sup> although it increased the computational cost. A second recent example combined the bond capacity (BC) polarization model with both the generalized Born (GB) model and the conductor-like polarizable continuum model (C-PCM).<sup>55</sup> For the BC-GB model, NVE molecular dynamics (MD) was shown to conserve energy at a modest increase in cost of only 15% relative to vacuum.<sup>56</sup> Cooper et al. have developed an electrostatics solver called PyGB, which employs a tree-code accelerated boundary-element formulation. PyGB specifically addresses the multisurface problems common to boundary-element method (BEM) solvers and is able to achieve directly comparable accuracy to the adaptive Poisson-Boltzmann solver (APBS) with increased speed.<sup>57,58</sup> Although beyond our focus on implicit solvents for biomolecular polarizable force fields, there is a large body of work dedicated to quantum mechanical SCRf implicit solvents,<sup>59–61</sup> including the polarizable continuum model (PCM),<sup>62,63</sup> the solvent model (SM) series,<sup>64–66</sup> and conductor-like screening models (COSMOs).<sup>67,68</sup>

Here, we describe the theory, implementation, and parametrization of implicit solvents compatible with the polarizable

AMOEBA force field. We describe a nonpolar estimator consisting of novel cavitation and dispersion terms, which is combined with electrostatic contributions based on solving the PBE numerically with the adaptive Poisson-Boltzmann solver (APBS),<sup>43</sup> the domain-decomposition COSMO (ddCOSMO) approach,<sup>44</sup> and the analytic generalized Kirkwood (GK) model.<sup>45</sup> Model parameters are fit to experimental solvation free energy differences for a set of 103 small molecules. The resulting implicit solvent hydration free energy differences are compared to those obtained previously using explicit solvent AMOEBA free energy simulations. Furthermore, the electrostatic response of the resulting models is validated for a series of proteins and nucleic acids in continuum water compared to both explicit solvent AMOEBA simulations and to widely used fixed-charge force fields. Finally, the relative computational speed of the models is compared.

## METHODS

**AMOEBA Parameterization Using PolType2.** The PolType2 protocol was used to generate AMOEBA small-molecule parameters, beginning from an initial optimization at the MP2/6-31G\* level of theory. *Ab initio* quantum mechanics (QM) calculations were performed using Gaussian 09. All molecular mechanics (MM) force field-based calculations needed for parameterization were performed using the Tinker 8 Software.<sup>69</sup> Valence parameters were taken from the small-molecule parameter database in PolType2. Atomic multipole moments were initially assigned from the QM electron density calculated at the MP2/6-311G\*\* level via Stone's distributed multipole analysis.<sup>70</sup> Further optimization of permanent multipoles was performed using the Tinker Potential program to fit the electrostatic potential around each molecule to a QM electron density at the MP2/6-311++G(2d,2p) level. All small-molecule AMOEBA parameter files (Tinker "prm" format) are available as [Supporting Information](#).

**Small-Molecule Data Set.** A test set of 103 small molecules was used to parameterize the implicit solvent models. The experimental solvation free energy differences for neutral compounds were taken from the FreeSolv Database, version 0.51,<sup>71,72</sup> unless otherwise indicated.<sup>73,74</sup> Experimental hydra-

tion free energy differences for charged compounds ( $\Delta G_{\text{solv}}^{\text{ion}}$ ) were calculated using eq 2

$$\Delta G_{\text{solv}}^{\text{ion}} = \Delta G_{\text{solv}}^{\text{neutral}} \pm \Delta G_{\text{gas}}^{\text{basicity}} \pm \Delta G_{\text{solv}}^{\text{H}^+} \mp 2.303RT \times \text{p}K_{\text{a}} \quad (2)$$

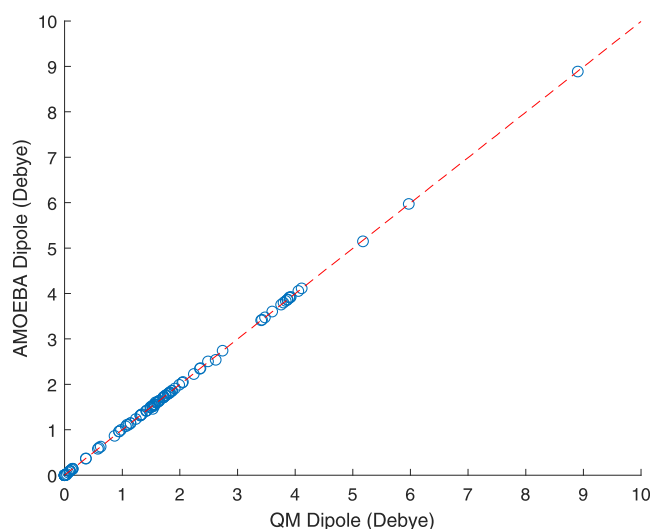
where the upper signs are used for cations, and the lower signs for anions.  $\Delta G_{\text{solv}}^{\text{neutral}}$  is the solvation free energy difference of the neutral molecule,  $\Delta G_{\text{gas}}^{\text{basicity}}$  is the gas-phase basicity from NIST,<sup>75,76</sup>  $\Delta G_{\text{solv}}^{\text{H}^+}$  is the solvation free energy difference of a proton,  $R$  is the universal gas constant in kcal/mol,  $T$  is the temperature in Kelvin, and  $\text{p}K_{\text{a}}$  is the negative decimal logarithm of the acid dissociation constant from Stewart.<sup>77</sup> For phosphate and guanidinium compounds, experimental values for the neutral solvation free energy difference and/or gas-phase basicity were not available. Due to their importance in fitting electrostatic implicit solvent parameters for proteins (*i.e.*, arginine) and nucleic acids (*i.e.*, the phosphate backbone), target solvation free energy differences for these compounds were calculated from AMOEBA explicit solvent solvation free energy simulations.

The value for the solvation free energy difference of the proton used here (−254.22 kcal/mol) was calculated as the sum of potassium ion solvation free energy (−74.32 kcal/mol<sup>78</sup>) and the experimental free energy of transfer  $\text{K}^+ \rightarrow \text{H}^+$  (−179.90 kcal/mol<sup>79</sup>). Using an intrinsic value for proton solvation, we avoid implicitly fitting to the interface potential (*i.e.*, the potential created by preferential orientation of water molecules at the vacuum–liquid interface; see Figure S1, Supporting Information), while the preferential orientation of water around uncharged solute cavities (*i.e.*, the cavity potential) is included. The potential energy contribution due to a charged molecule crossing the vacuum–liquid interface can be added to implicit solvent free energy differences in the same manner as for periodic explicit solvent simulations.<sup>80</sup> By excluding the interface potential from implicit solvent parameterization, ensemble averages, conformational distributions, and free energy differences from implicit solvent and periodic boundary explicit solvent simulations are directly comparable.<sup>78,81</sup>

Of the molecules tested, 91 were neutral and 12 were charged. Charged compounds were chosen based on their chemical similarity to charged groups in biomolecules. All starting structures were obtained from PubChem and parameterized for AMOEBA using PolType2 as described above,<sup>81</sup> then energy-minimized in vacuum. We see close agreement between the vacuum dipole moments from single point MP2/6-311G\*\*(2d,2p) calculations and those from the AMOEBA small-molecule parameters, shown in Figure 2.

**Cavitation Free Energy.** The Lum–Chandler–Weeks theory of hydrophobicity predicts contrasting behavior for the cavitation free energy of small and large solutes.<sup>1,82–84</sup> At all length scales, the driving force for phase separation is proportional to the solute volume, while the cost to form an interface is proportional to the surface area. These competing factors are exhibited in a cross-over in the dependence of the cavitation free energy change between volume scaling for small solutes and surface area scaling for large solutes, which, for a spherical cavity, occurs at a radius of approximately 10 Å.

$$\Delta W_{\text{cav}}(r) \propto \begin{cases} \text{volume,} & r \leq \sim 10 \text{ Å} \\ \text{surface area,} & r > \sim 10 \text{ Å} \end{cases} \quad (3)$$



**Figure 2.** Total dipole moments for the parameterization set of 103 molecules from QM using MP2/6-311G\*\*(2d,2p) are compared to those of the resulting AMOEBA models (red dashed line:  $Y = 0.9986 \cdot X + 0.0011$ ;  $R^2 = 0.9999$ ).

For spherical solutes with a radius below this threshold, water molecules are generally able to form a hydrogen bond network surrounding the solute that maintains a complement of hydrogen bonds similar to bulk water (*i.e.*, ~4 per water). As the solute size increases toward mimicking a flat liquid–vacuum interface, each water molecule on average sacrifices a single hydrogen bond (*i.e.*, one hydrogen from each water molecule is directed toward vacuum giving rise to a phase potential). For solutes with varying shapes, such as biomolecules, the cavitation cost is neither proportional to the volume nor the surface area but rather some local mixture of the two regimes. For example, the cost to form a cavity scales more with volume character for an extended chain than for a compact spherical conformation, where both conformations have equal surface areas. One can imagine protein conformations with both extended loops and large compact regions, suggesting that cavitation terms that do not consider local conformation are clearly an approximation. It is beyond the scope of the current work to develop a general functional form for the cavitation free energy of a biomolecular solute of arbitrary size and shape, although approaches that adjust the effective surface tension (ST) based on local curvature are promising.<sup>85</sup> Fortunately, as small-molecule cavitation free energies are in the volume-scaling regime, the magnitude of the cavitation term for the AMOEBA implicit solvents is not expected to change for the small-molecule parameterization discussed here, even if an improved cavitation model for larger biomolecules is defined in the future.

An effective radius for a nonspherical solute conformation  $\mathbf{X}$  can be determined from the calculation of either the solvent-excluded volume  $\text{SEV}(\mathbf{X})$

$$r_{\text{SEV}}(\mathbf{X}) = \sqrt[3]{3 \times \text{SEV}(\mathbf{X})/4\pi} \quad (4)$$

or the solvent-accessible surface area  $\text{SASA}(\mathbf{X})$

$$r_{\text{SASA}}(\mathbf{X}) = \sqrt{\text{SASA}(\mathbf{X})/4\pi} \quad (5)$$

using algorithms developed by Connolly<sup>86–88</sup> and implemented in Tinker<sup>69</sup> and Force Field X.<sup>89</sup> As first described by Richards, SEV and SASA are defined by rolling a spherical probe (*i.e.*, with a radius of 1.4 Å to approximate water) around the surface of a

molecule.<sup>90</sup> The cavitation free energy of an (approximately) spherical solute can then be described by a piecewise continuous function of its effective radius

$$\Delta W_{\text{sphere}}(r) = \begin{cases} \lambda \times \text{SEV}(\mathbf{X}), & r \leq \chi \\ \gamma \times \text{SASA}(\mathbf{X}), & r > \chi \end{cases} \quad (6)$$

where in the volume-scaling regime, cavitation free energy is defined by the product of SEV and solvent pressure (SP) denoted by  $\lambda$  (kcal/(mol Å<sup>3</sup>)); in the surface-area-scaling regime, cavitation free energy is defined by the product of SASA and surface tension (ST) denoted by  $\gamma$  (kcal/(mol Å<sup>2</sup>)). For our model, SP was assessed using two explicit solvent simulation approaches that are in general agreement. The first approach assumes that the SEV and SASA cavitation free energies are equal at the cross-over point  $\chi$  for a spherical solute, which yields the relationship  $\lambda = (3\gamma)/\chi$ . This defines SP to be 0.031 kcal/(mol Å<sup>3</sup>), using the experimental surface tension of water (0.103 kcal/(mol Å<sup>2</sup>)) and an approximate cross-over point of 10 Å from fixed-charge simulations.<sup>1</sup> The second approach leverages explicit solvent free energy perturbation simulations<sup>91</sup> using the AMOEBA water model<sup>92</sup> and 39 AMOEBA small molecules as described elsewhere,<sup>41</sup> which resulted in a mean SP of 0.0334 kcal/(mol Å<sup>3</sup>). Using the relationship between the experimental surface tension of water and the latter SP, the volume-to-surface-area cross-over radius is 9.251 Å. Both SP estimates are within 0.003 kcal/(mol Å<sup>3</sup>) of each other, and both define cross-over radii that differ by less than 0.75 Å. For the current model, the latter SP of 0.0334 kcal/(mol Å<sup>3</sup>) and the cross-over radius of 9.251 Å were chosen due to their consistency with the AMOEBA model.

The simple definition in eq 6 for the transition between the volume-scaling and surface-area scaling regimes is not useful for molecular dynamic simulations or optimization algorithms because it lacks continuous first and second derivatives. To address this, it is possible to introduce a simple multiplicative switch  $s_v(r)$  to smoothly turn off the volume term and a second switch  $s_{sa}(r)$  to smoothly turn on the surface area term. Each switch acts over a window of length  $w = 7$  Å centered on the cross-over point  $\chi$ , such that the switch begins at  $b = \chi - w/2$  and ends at  $e = \chi + w/2$  to give

$$\Delta W_{\text{switch}}(r) = \begin{cases} \lambda \times \text{SEV}(\mathbf{X}), & r \leq b \\ \lambda \times \text{SEV}(\mathbf{X}) \times s_v(r) + \gamma \times \text{SASA}(\mathbf{X}) \times s_{sa}(r), & b < r \leq e \\ \gamma \times \text{SASA}(\mathbf{X}), & e < r \end{cases} \quad (7)$$

The volume-scaling switch  $s_v(r)$  is a fifth-order polynomial whose six coefficients are uniquely determined by constraining its value at  $b$  to  $s_v(b) = 1$  and at  $e$  to  $s_v(e) = 0$ , as well as constraining first and second derivatives at  $b$  and  $e$  to be zero. This gives

$$s_v(r) = c_0 + c_1r + c_2r^2 + c_3r^3 + c_4r^4 + c_5r^5 \quad (8)$$

where

$$\begin{aligned} c_0 &= e^3(e^2 - 5eb + 10b^2)/d \\ c_1 &= -30e^2b^2/d \\ c_2 &= 30eb(e + b)/d \\ c_3 &= -10(e^2 + 4eb + b^2)/d \\ c_4 &= 15(e + b)/d \\ c_5 &= -6/d \\ d &= (e - b)^5 \end{aligned} \quad (9)$$

The surface area switch in this symmetric case is

$$s_{sa}(r) = 1 - s_v(r) \quad (10)$$

The behavior of the cavitation free energy using a symmetric switch showed a modest peak at the cross-over point, which is removed by shifting the center of the switching region for the SA term to larger effective radius values by a small offset  $o = 0.2$  Å. This gives the final functional form used here

$$\Delta W_{\text{cav}}(\mathbf{X}) = \begin{cases} \lambda \times \text{SEV}(\mathbf{X}), & r \leq b \\ \lambda \times \text{SEV}(\mathbf{X}) \times s_v(r), & b < r \leq e \\ 0, & e < r \end{cases} + \begin{cases} 0, & r \leq b + o \\ \gamma \times \text{SASA}(\mathbf{X}) \times s_{sa}(r - o), & b + o < r \leq e + o \\ \gamma \times \text{SASA}(\mathbf{X}), & e + o < r \end{cases} \quad (11)$$

where the surface area switch is now slightly offset from the volume switch. The smooth behavior of  $\Delta W_{\text{cav}}(\mathbf{X})$  as a function of the effective radius is shown in Figure 3. In this work, the effective radius of the molecule is determined using SASA (eq 5), rather than SEV (eq 4), due to the former being faster to compute for large solutes (*i.e.*, SEV does not need to be computed for large biomolecules). The gradient of SEV and SASA with respect to atomic coordinates has been described previously.<sup>93–95</sup>

**Dispersion Free Energy.** The pairwise dispersion energy for the AMOEBA model is given by a buffered 14-7 potential<sup>97</sup>

$$U_{14-7}(r_{ij}) = \epsilon_{ij} \left( \frac{1.07r_{0,ij}}{r_{ij} + 0.07r_{0,ij}} \right)^7 \left( \frac{1.12r_{0,ij}^7}{r_{ij}^7 + 0.12r_{0,ij}^7} - 2 \right) \quad (12)$$

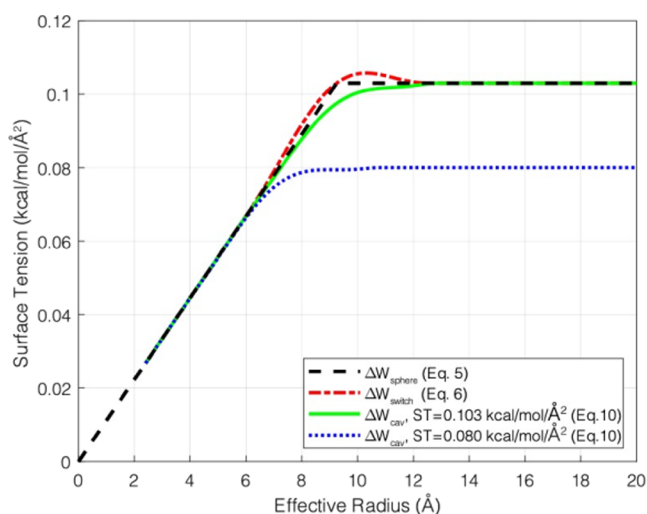
where  $r_{ij}$  is the separation distance between atoms  $i$  and  $j$ ,  $\epsilon_{ij}$  is the well depth, and  $r_{0,ij}$  is the minimum energy separation distance.<sup>41</sup> This can be used to define a purely repulsive Weeks–Chandler–Andersen (WCA) potential<sup>98,99</sup> as

$$U_{\text{rep}}(r_{ij}) = \begin{cases} U_{14-7}(r_{ij}) + \epsilon_{ij}, & r_{ij} < r_{0,ij} \\ 0, & r_{ij} \geq r_{0,ij} \end{cases} \quad (13)$$

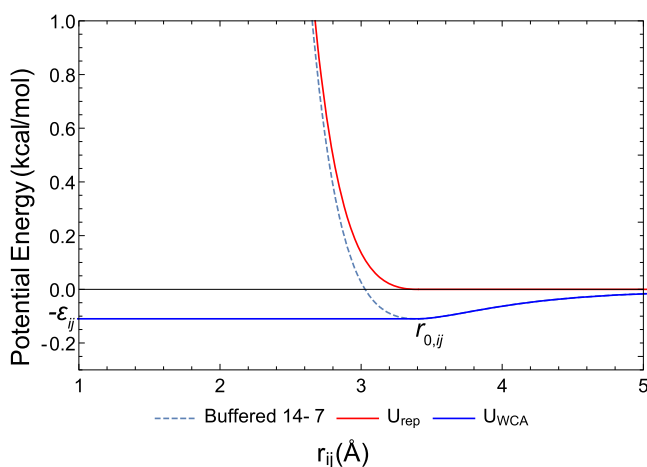
which is shown in Figure 4.

Work by Gallicchio, Kubo, and Levy (GKL) demonstrated that the free energy of adding dispersion interactions to the WCA repulsive potential, thereby restoring full van der Waals interactions between solute and solvent, is nearly equal to the change in the solute–solvent enthalpy for a series of small alkanes studied using free energy perturbation (FEP)<sup>100</sup>

$$\Delta W_{\text{disp}} \approx \langle U_{14-7} \rangle - \langle U_{\text{rep}} \rangle \quad (14)$$



**Figure 3.** Surface tension is not constant for small solutes but increases approximately linearly until the effective radius of the solute increases to beyond  $\sim 10$  Å. For large (flat) solutes, surface tension asymptotes toward the experimental value for a water-vapor interface of  $0.103$  kcal/(mol Å<sup>2</sup>).<sup>96</sup> This length scale dependence can be approximately captured by a cavitation free energy difference that switches between using SEV and SASA via either a simple, nondifferentiable form ( $\Delta W_{\text{sphere}}(r)$  given by eq 6, black dashed line) or the smooth form used in this work ( $\Delta W_{\text{cav}}(r)$  given by eq 11, green solid line). The asymptotic surface tension of  $\Delta W_{\text{cav}}(r)$  can be reduced relative to the experimental value (e.g., to  $0.08$  kcal/(mol Å<sup>2</sup>), dotted blue line) to capture cavitation for solutes with a large effective radius but which are more highly curved than a simple sphere (e.g., a DNA double helix, RNA molecule, or protein).



**Figure 4.** Pairwise buffered 14-7 potential ( $U_{14-7}$ ) can be decomposed into purely repulsive ( $U_{\text{rep}}$ , eq 13) and attractive ( $U_{\text{WCA}}$ , eq 16) contributions, which are plotted for the AMOEBA water oxygen atom ( $r_{0,ij} = 3.405$  Å,  $\epsilon_{ij} = 0.11$  kcal/mol). The cavitation free energy ( $\Delta W_{\text{cav}}$ ) represents the process of growing in the repulsive potential ( $U_{\text{rep}}$ ) for solute atoms. Subsequently, the dispersion free energy ( $\Delta W_{\text{disp}}$ ) models the process of adding the attractive  $U_{\text{WCA}}$  solute–solvent interactions to recover the full  $U_{14-7}$  potential in the context of an uncharged solute.

This led to their suggestion of a dispersion free energy estimator based on Born radii, such that the dispersion free energy of the solute is

$$\Delta W_{\text{GKL}} = \sum_{i=1}^n \frac{-16\pi\rho_w\epsilon_{iw}\sigma_{iw}^6}{3R_i^3} \quad (15)$$

where  $\rho_w$  is the number density of water ( $0.033428$  per Å<sup>3</sup>),  $\epsilon_{iw}$  and  $\sigma_{iw}$  are the well depth and sigma value of the interaction of atom  $i$  with the TIP3P water model, respectively,  $n$  is the number of solute atoms, and  $R_i$  is the Born radius.<sup>37,73</sup> In effect, the term acts like a tail correction, assuming solvent to be a continuum outside the solute and integrating the  $1/r^6$  attractive portion of a 6-12 Lennard-Jones potential. In the limit of a spherical solute, the use of the Born radii in eq 15 is exact; however, for other geometries, it is an approximation.

The goal for the dispersion free energy model is to build on the insights described above by removing the Born radii from the GKL model given in eq 15 and instead integrating the true WCA attractive potential (Figure 4) outside of the solute cavity for each atom.

$$U_{\text{WCA}}(r_{ij}) = U_{14-7}(r_{ij}) - U_{\text{rep}}(r_{ij}) = \begin{cases} -\epsilon_{ij}, & r_{ij} < r_{0,ij} \\ U_{14-7}(r_{ij}), & r_{ij} \geq r_{0,ij} \end{cases} \quad (16)$$

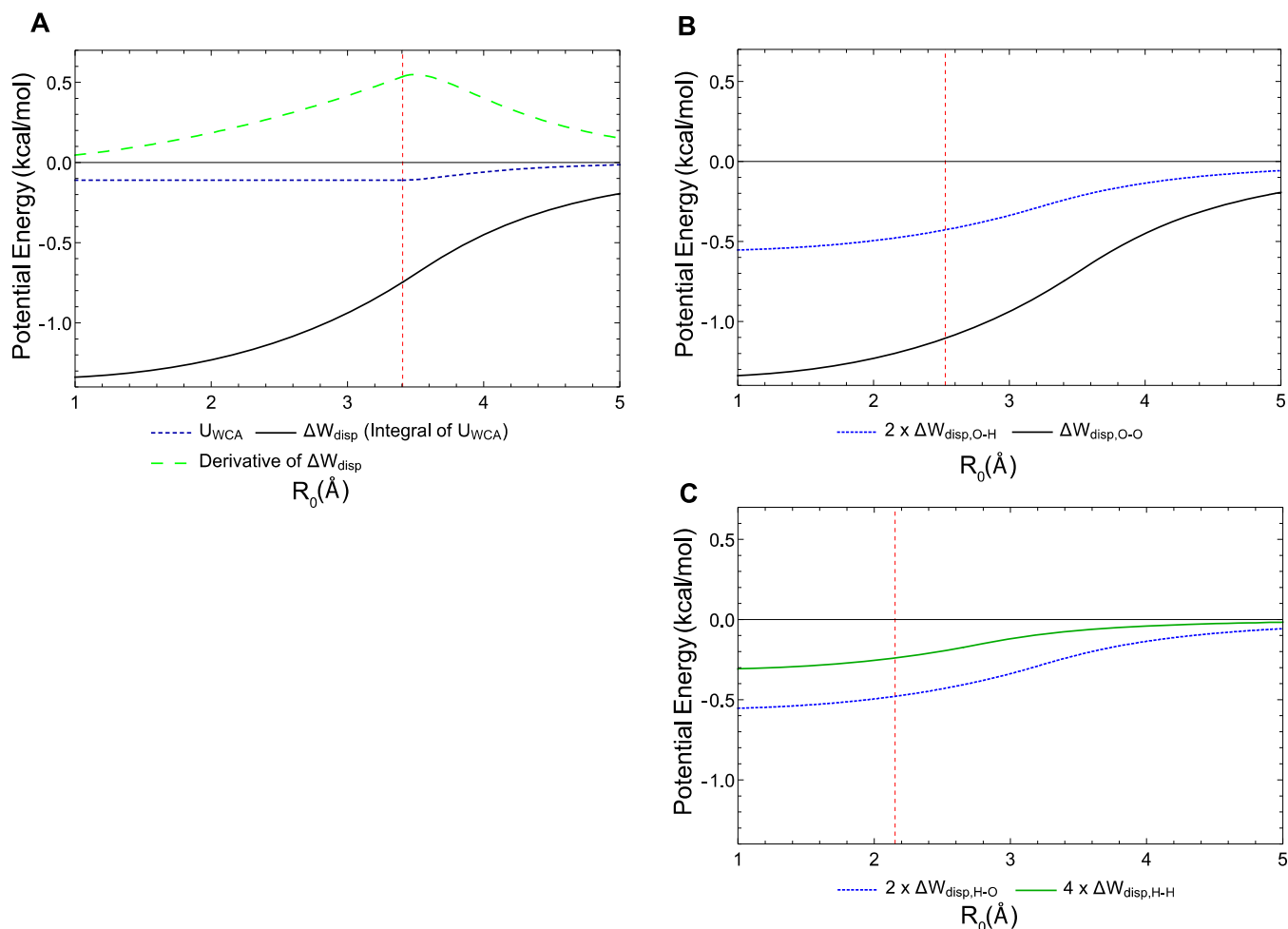
We present an analytic approach based on the Hawkins–Cramer–Truhlar (HCT) pairwise integration method also used for GK.<sup>34,101</sup> Due to the use of the buffered 14-7 potential by AMOEBA, the underlying pairwise integration machinery needs to account for the constant portion of the WCA potential for  $r < r_{0,io}$  (where in this case  $r_{0,io}$  is the minimum energy separation for solute atom  $i$  with an AMOEBA water oxygen) and integrate both  $1/r^7$  and  $1/r^{14}$  for  $r > r_{0,io}$ . The general analytic form for the dispersion free energy,  $\Delta W_{\text{disp}}(\mathbf{X})$ , of a solute with coordinates  $\mathbf{X}$  is given by

$$\Delta W_{\text{disp}}(\mathbf{X}) = \rho_w \sum_{i=1}^n \int_{R_0}^{\infty} \int_0^{\pi} \int_0^{2\pi} U_{\text{WCA}}(r) \times S(r, \theta, \phi, \mathbf{X}, \mathbf{R}) \sin \theta r^2 d\phi d\theta dr \quad (17)$$

where the solvent indicator function  $S$  is unity if the point  $(r, \theta, \phi)$  is located within the solvent but zero otherwise,  $\rho_w$  is the number density of water, and  $\mathbf{R}$  are the AMOEBA minimum energy separation distance values ( $r_{0,ij}$ ) for each atom. The radial integral for atom  $i$  with continuum water oxygen begins from half their combined  $r_{0,ij}$  (in this case,  $r_{0,io}$  for atom  $i$  with water oxygen) value plus an offset ( $d = 1.056$  Å), which is one of two free parameters in the model. The beginning of the radial integral is defined as  $R_0 = r_{0,ij}/2 + d$ . The second free parameter is a scale factor ( $s = 0.75$ ) that accounts for the overlapping volumes of neighboring atoms during evaluation of the dispersion integral over solute atoms. Both parameters, which appear below in eq 18 for  $\Delta W_{\text{disp}}(\mathbf{X})$ , were fit against dispersion enthalpies (eq 18) measured from explicit solvent simulations as described in the Supporting Information (Table S1). For a single water oxygen atom, the behavior of  $\Delta W_{\text{disp}}$  is shown in Figure 5A, and the dispersion interactions of explicit water atoms (oxygen and hydrogen) with continuum water oxygen and hydrogen are shown in Figure 5B,C.

After performing the two angular integrals in eq 16, inverting the integration domain and applying the HCT pairwise approximation<sup>34</sup> gives

$$\Delta W_{\text{disp}}(\mathbf{X}) = \rho_w \sum_{i=1}^n \left[ U_{\text{tail,water}}(d) - 4\pi \sum_{j \neq i} \int_L^U U_{\text{WCA,water}}^*(r) H(r, r_{ij}, \rho_j) r^2 dr \right] \quad (18)$$



**Figure 5.** (A) Dispersion free energy differences ( $\Delta W_{\text{disp}}$ , solid black curve) are given by the integral of the attractive WCA potential ( $U_{\text{WCA}}$ , dotted blue curve) over solvent for the interaction of two AMOEBA water oxygen atoms. The derivative of  $\Delta W_{\text{disp}}$  with respect to  $R_0$  (dashed green curve) shows a maximum slightly beyond the minimum energy separation distance (vertical red line) due to the volume element  $4\pi r^2 dr$  increasing more quickly than  $U_{\text{WCA}}$  approaches zero just beyond  $r_{0,ij}$ . (B) Dispersion interactions of an explicit water oxygen atom with continuum water oxygen (also plotted in (A)) and hydrogen. The interaction of oxygen with continuum hydrogen is multiplied by 2 (two hydrogen atoms per water molecule). (C) Same as (B) but for two explicit water hydrogen atoms interacting with continuum water.

where  $H$  is the fraction of the area of the current spherical integration shell of radius  $r$  that is covered by atom  $j$  located a distance  $r_{ij}$  from atom  $i$  and whose radius is scaled to  $\rho_j = sR_j$ , and  $H$  is given by (eq 12 in Hawkins et al.<sup>34</sup>)

$$H(r, r_{ij}, \rho_j) = \frac{1}{2} - \frac{1}{4} \left[ \frac{r_{ij}^2 + r^2 - \rho_j^2}{r_{ij}r} \right] \quad (19)$$

The integrated WCA potential  $U_{\text{WCA}}^*(r_{ij})$  uses a simplified form of the buffered 14-7 for the interaction of solute atoms with water (i.e., the buffering constants are set to zero)

$$U_{\text{WCA}}^*(r_{ij}, \varepsilon, r_{0,ij}) = \begin{cases} -\varepsilon, & r_{ij} < r_{0,ij} \\ \varepsilon r_{0,ij}^7 \left( \frac{r_{0,ij}^7}{r_{ij}^{14}} - \frac{2}{r_{ij}^7} \right), & r_{ij} \geq r_{0,ij} \end{cases} \quad (20)$$

Fortunately, the difference between this (unbuffered)  $U_{\text{WCA}}^*$  potential and the buffered 14-7 form is negligible for separations greater than the minimum energy distance. The analytic tail correction based on eq 20 is given by

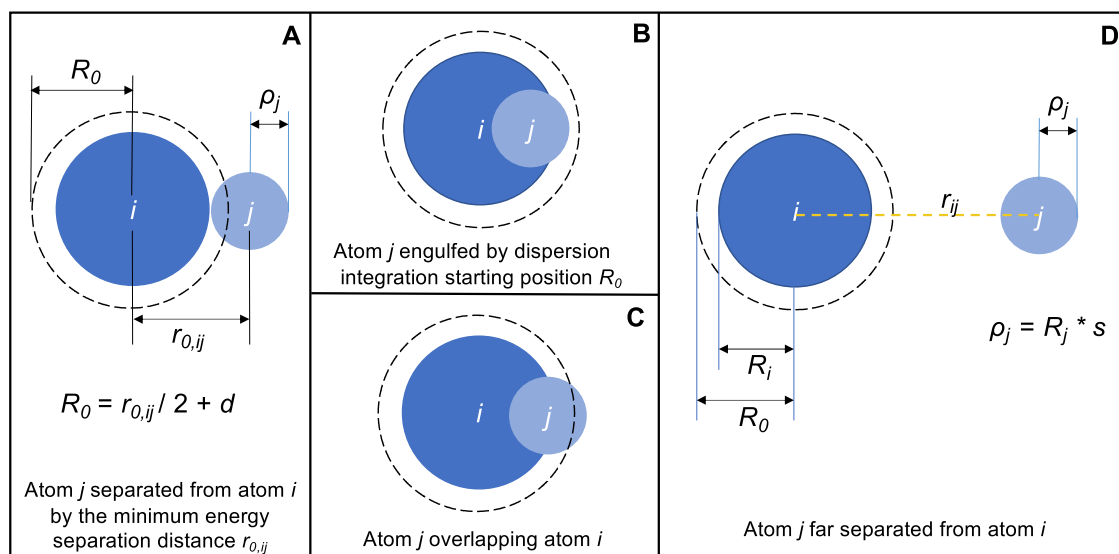
$$U_{\text{tail}}(d, \varepsilon, r_{0,ij}) = \int_{R_0=r_{0,ij}/2+d}^{\infty} U_{\text{WCA}}^*(r, \varepsilon, r_{0,ij}) 4\pi r^2 dr = \begin{cases} -\frac{4}{3}\pi\varepsilon(r_{0,ij}^3 - R_0^3) - \frac{18}{11}\pi\varepsilon r_{0,ij}^3, & R_0 < r_{0,ij} \\ \pi\varepsilon r_{0,ij}^7 \left( \frac{4r_{0,ij}^7}{11R_0^{11}} - \frac{2}{R_0^4} \right), & R_0 \geq r_{0,ij} \end{cases} \quad (21)$$

and the total tail correction for the interaction of atom  $i$  with water is then given by

$$U_{\text{tail,water}}(d) = U_{\text{tail}}(d, \varepsilon_{io}, r_{0,io}) + 2 \times U_{\text{tail}}(d, \varepsilon_{ih}, r_{0,ih}) \quad (22)$$

where the well depths ( $\varepsilon_{io}$ ,  $\varepsilon_{ih}$ ) and minimum energy distances ( $r_{0,io}$ ,  $r_{0,ih}$ ) are based on the AMOEBA mixing rules for atom  $i$  with the AMOEBA water model.<sup>92</sup> The final piece to this model is the solution to the integral in eq 18, which uses integration bounds shown in Figure 6.

If integration of the WCA dispersion begins inside the minimum energy distance  $R_0 < r_{0,ij}$ , then a contribution of



**Figure 6.** Illustration of the integration limits for the dispersion free energy based on solute van der Waals parameters and separation distance. (A) The minimum separation distance  $r_{0,ij}$  for atoms  $i$  and  $j$  is based on AMOEBA mixing rules and used to determine the beginning of the WCA dispersion integral  $R_0$ . When  $R_0 < r_{0,ij}$ , integration for the constant portion of the WCA potential begins at either  $R_0$  or  $r_{ij} - \rho_j$ , whichever is larger, and ends at  $r_{0,ij}$  or  $r_{ij} + \rho_j$ , whichever is smaller. If  $R_0 > r_{0,ij}$ , then only the variable portion of the WCA potential factors into the dispersion free energy. (B) If atom  $j$  is completely engulfed by the sphere defined by  $R_0$ , no solvent is blocked and no dispersion energy must be removed. (C) When the two atoms overlap or are close together such that  $r_{ij} - \rho_j < R_0$ , integration of the attractive WCA potential begins at  $R_0$  and ends at  $r_{ij} + \rho_j$  (the furthest edge of atom  $j$ ). (D) When atom  $j$  is outside the beginning of the integration, integration of the variable portion of the WCA potential begins at  $r_{ij} - \rho_j$  (the closest edge of atom  $j$ ) and ends at  $r_{ij} + \rho_j$ .

$$I_\varepsilon(U, L, \varepsilon) = \int_L^U \varepsilon \times H(r, r_{ij}, \rho_j) r^2 dr$$

$$= \varepsilon [r^2(3r^2 - 8r_{ij}r + 6r_{ij}^2 - 6\rho_j^2) / 48r_{ij}]_L^U \quad (23)$$

is included. The lower limit  $L$  is  $R_0$  or  $r_{ij} - \rho_j$ , whichever is greater. The upper limit  $U$  of this integral is  $r_{0,ij}$  or  $r_{ij} + \rho_j$ , whichever is smaller. If  $r_{ij} + \rho_j$  is greater than  $r_{0,ij}$ , the integration of the repulsive contribution outside  $r_{0,ij}$  is given by

$$I_{14}(L, U, \varepsilon, r_{0,ij}) = \int_L^U \frac{\varepsilon r_{0,ij}^{14}}{r^{12}} H(r, r_{ij}, \rho_j) dr$$

$$= \varepsilon r_{0,ij}^{14} [(-120r_{ij}r + 66r^2 + 55r_{ij}^2 - 55\rho_j^2) / 2640r_{ij}r^{12}]_L^U \quad (24)$$

and the attractive contribution by

$$I_7(L, U, \varepsilon, r_{0,ij}) = -2 \int_L^U \frac{\varepsilon r_{0,ij}^7}{r^5} H(r, r_{ij}, \rho_j) dr$$

$$= -2\varepsilon r_{0,ij}^7 [(-15r_{ij}r + 10r^2 + 6r_{ij}^2 - 6\rho_j^2) / 120r_{ij}r^5]_L^U \quad (25)$$

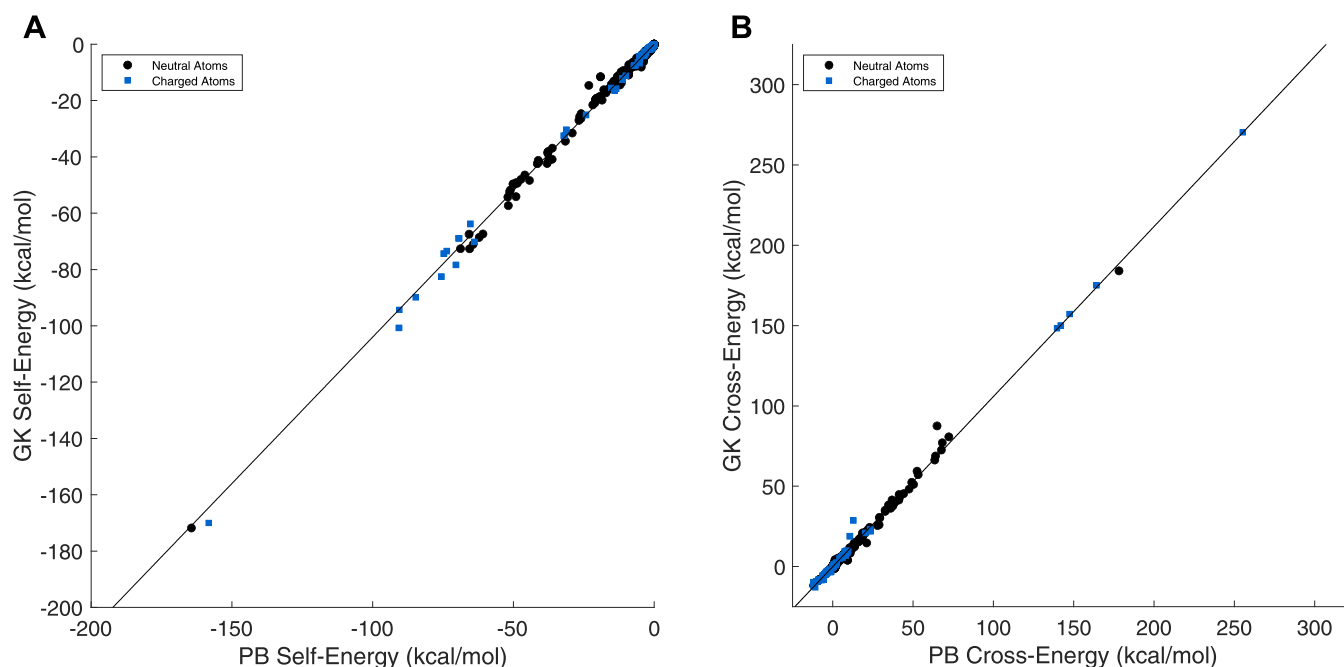
where the upper limit is always  $r_{ij} + \rho_j$ . As before, the lower limit  $L$  is  $R_0$  or  $r_{ij} - \rho_j$ , whichever is greater, unless this result is inside the minimum energy distance  $r_{0,ij}$ . In this case, a contribution up to  $r_{0,ij}$  has already been included from eq 23 and  $L$  takes the value  $r_{0,ij}$ . The distances and parameters used to define integration limits for the WCA potential are shown in Figure 6.

**Electrostatic Free Energy.** The continuum electrostatics contribution to solvation free energy of a small molecule can be determined by solving the nonlinear PBE (NPBE) or the linearized PBE (LPBE) form shown here

$$\nabla \cdot [\varepsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] - \bar{\kappa}^2(\mathbf{r}) \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \quad (26)$$

for all  $\mathbf{r}$  in a domain  $\Omega$ , where  $\varepsilon(\mathbf{r})$  is the dielectric constant,  $\phi(\mathbf{r})$  is the electrostatic potential,  $\bar{\kappa}^2(\mathbf{r})$  is the modified Debye–Hückel screening factor, and  $\rho(\mathbf{r})$  is the solute charge density. For polarizable force fields, the solute charge density  $\rho(\mathbf{r})$  responds to the reaction field of the solvent, and thus eq 26 is solved repeatedly during iterations of an SCRF solver (e.g., Jacobi over-relaxation (SOR),<sup>92</sup> conjugate gradient (CG) methods,<sup>102,103</sup> the Jacobi algorithm coupled to direct inversion in the iterative subspace (JI/DIIS),<sup>102</sup> and an optimized perturbation theory (OPT) method<sup>104</sup>). This work compares three distinct continuum electrostatics models: numerical solutions to both the NPBE and LPBE using the adaptive Poisson–Boltzmann solver (APBS),<sup>31</sup> a domain-decomposition solution of the conductor-like screening model (ddCOSMO),<sup>44,105–107</sup> and the analytic generalized Kirkwood (GK) theory.<sup>45</sup>

**Adaptive Poisson–Boltzmann Solver.** APBS determines the solution to the PBE using parallelized finite difference multigrid and finite element algebraic multigrid numerical methods. Finite difference methods subdivide the domain in which the PBE is to be solved using Taylor expansions to model the differential operators in each subdomain as difference matrices and solving them via linear algebra techniques. The final algebraic equations obtained by this discretization can be solved via a multilevel solver: iteration is used to reach solutions at varying resolutions, where long-range errors in the iterations are allowed to converge on coarser grid spacings before using a finer grid for the final solution. Though it provides one of the more accurate numerical solutions to the PBE, APBS can become computationally expensive for larger domains and finer final multigrid spacings. The combination of Tinker<sup>108</sup> with APBS to support the AMOEBA force field has been described previously,<sup>43</sup> including convergence of the SCRF and calculation of atomic forces for the LPBE. APBS was run in Tinker using a grid spacing of 129<sup>3</sup> and a probe of radius 0.0 Å to



**Figure 7.** (A) Fit of GK self-energies to *perfect* PB self-energies ( $Y = 1.040X + 0.048$ ,  $R^2 = 0.996$ ). (B) Fit of GK cross-term energies to *perfect* PB cross-term energies ( $Y = 1.059X - 0.089$ ,  $R^2 = 0.994$ ). Both self and aggregate cross-term energies are reported for 1424 atoms.

define a van der Waals solute cavity. The average grid length for the small-molecule test set was 0.105 Å. APBS was run using multiple Debye–Hückel boundary conditions, a water dielectric constant of 78.3, and a solute dielectric constant of 1.0. The APBS parallel multigrid (PMG) solver<sup>26,109,110</sup> was used for all calculations, as it is currently the only solver within APBS available for use with AMOEBA via Tinker.

**Domain-Decomposition Conductor-Like Screening Model.** The ddCOSMO electrostatics model<sup>44,105–107</sup> treats the solvent as an infinite conductor surrounding a solute-shaped cavity  $\Omega$ , determined by a union of spheres (one sphere per atom of solute)

$$\Omega = \bigcup_{j=1}^M \Omega_j(R_j, r_j) \quad (27)$$

The electrostatic interactions are calculated by integrating the charge density  $\rho$  of the solute molecule multiplied by the reaction potential  $W$  of the conductor over the molecular cavity

$$E_s = \frac{1}{2} f(\epsilon) \int_{\Omega} \rho(r) W(r) dr \quad (28)$$

where  $f(\epsilon)$  is a scaling factor used to adjust for the nonconductor nature of the solvent based on its dielectric constant  $\epsilon$ . The scaling factor  $f(\epsilon)$  is defined as

$$f(\epsilon) = \frac{\epsilon - 1}{\epsilon + x} \quad (29)$$

where  $x$  is an empirical constant.<sup>111</sup> In the original derivation of COSMO,  $x$  was set to 0.5. This value was later updated to use  $x = 0.5$  for neutral molecules and  $x = 0$  for ionic molecules.<sup>112,113</sup> The value of  $x$  used here was 0 in all cases, such that parameterized electrostatic radii described below are implicitly optimized for transferability between small molecules and charged biomolecules. The value of  $W$  in eq 28 is obtained from the solution of the following boundary value problem

$$\begin{cases} -\Delta W(r) = 0, & r \in \Omega \\ W(s) = -\Phi(s), & s \in \Gamma \end{cases} \quad (30)$$

where  $\Phi$  is the solute's electrostatic potential in vacuum and  $\Gamma$  is the boundary of the cavity. ddCOSMO uses Schwarz's domain-decomposition method to solve this boundary value problem by splitting it into a series of smaller problems, each defined on a single spherical domain. This decomposition allowed for the ddCOSMO implementation for AMOEBA (including convergence of the SCRF and calculation of atomic forces<sup>44</sup>) to be parallelized and is available in the Tinker-HP package,<sup>114</sup> which is part of the Tinker 8 distribution.

**Generalized Kirkwood.** GK is an analytic approximation to the PBE that simplifies to the generalized Born (GB) model in the absence of permanent multipoles and induced dipoles (e.g., for fixed partial charge force fields). The GB electrostatic energy<sup>115</sup> (equivalent to the GK monopole term  $G_{\text{GK}}^{(0)}$ ) is given by

$$\Delta G_{\text{GB}} = G_{\text{GK}}^{(0)} = -\frac{1}{2} \left( \frac{1}{\epsilon_h} - \frac{1}{\epsilon_s} \right) \sum_{i,j} \frac{q_i q_j}{f_{\text{GB}}} \quad (31)$$

where  $\epsilon_s$  is the permittivity of the solvent,  $\epsilon_h$  is the permittivity of a homogeneous reference state,  $q_i$  and  $q_j$  are partial charges, and the empirical generalizing function  $f_{\text{GB}}$  is given by

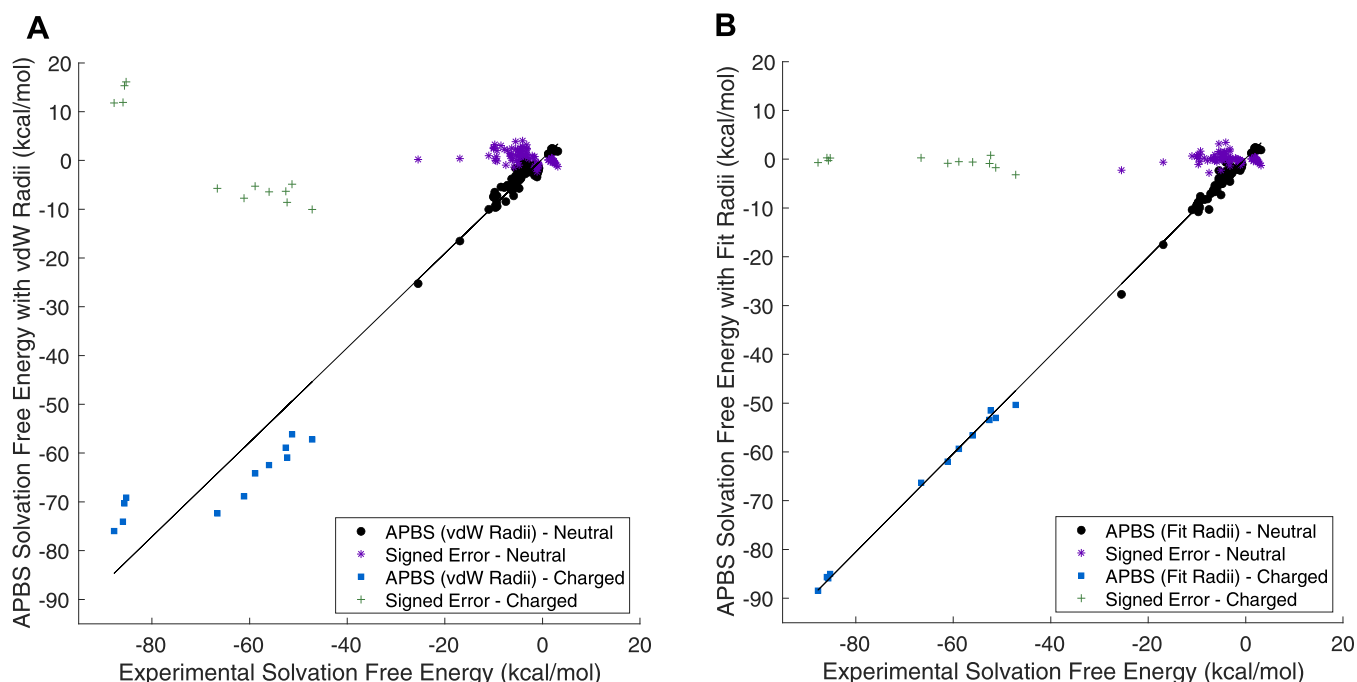
$$f_{\text{GB}} = \sqrt{r_{ij}^2 + a_i a_j} \quad (32)$$

where  $r_{ij}$  is the distance between sites  $i$  and  $j$ , effective “Born radii”  $a_i$  and  $a_j$  are given by an integral over solvent,<sup>116–118</sup>

$$\frac{1}{a_i} = \left( \frac{3}{4\pi} \int_{\text{ex}} \frac{1}{r^6} dV \right)^{1/3} \quad (33)$$

and  $f_{ij}$  is given by

$$f_{ij} = e^{-r_{ij}^2 / c_{\text{GK}} a_i a_j} \quad (34)$$



**Figure 8.** Comparison of experimental and APBS solvation free energy differences using either AMOEBA van der Waals  $R_{\min}$  radii to describe the solute–solvent boundary (A) or using fit radii is shown (B). When using AMOEBA  $R_{\min}$  radii, the linear regression gave  $Y = 0.9683 \cdot X + 0.3215$  with  $R^2 = 0.9702$ . When using fit radii, the linear regression gave  $Y = 1.0080 \cdot X + 0.1416$  with  $R^2 = 0.9979$ .

where  $c_{\text{GK}}$  is a tuning parameter that typically ranges from 2 to 4. GK extends GB methods to polarizable atomic multipole charge distributions using Kirkwood's analytic solution to the electrostatic component of solvation free energy for an arbitrary (*i.e.*, multipolar) charge distribution.<sup>46</sup> For example, the GK interaction between two permanent dipoles is expressed as

$$G_{\text{GK}}^{(1)} = \frac{1}{2} \left[ \frac{1}{\epsilon_h} \frac{2(\epsilon_h - \epsilon_s)}{2\epsilon_s + \epsilon_h} \right] \sum_{i,j} u_{i,\alpha} u_{j,\beta} \left[ \frac{3r_{\alpha\beta}(1 - f_{ij})}{f_{\text{GB}}^5} + \frac{\delta_{\alpha\beta}}{f_{\text{GB}}^3} \right] \quad (35)$$

where  $\mathbf{u}_i$  and  $\mathbf{u}_j$  are permanent dipole vectors and the subscripts  $\alpha$  and  $\beta$  indicate the use of the Einstein summation convention. GK interaction tensors up to the quadrupole–quadrupole order, as well as their inclusion in the AMOEBA SCRF calculation and calculation of atomic forces, were described previously.<sup>45</sup>

Perfect self and cross-term energies were calculated with APBS for each of the 1424 atoms in our set of 103 test molecules. The results were used to fit a unitless generalizing constant in the cross-term ( $c_{\text{GK}} = 2.455$ ) and a unitless scale factor ( $c_{\text{HCT}} = 0.72$ ) that avoids overestimation of descreening due to atomic overlaps when computing Born radii.<sup>119</sup> An initial fit based on small-molecule self-energies produced a scale factor of 0.77; however, testing with larger macromolecules led to excessive descreening. For this reason, the scale factor was reduced to 0.72. PB self and cross-term energies were calculated using Tinker with an APBS grid spacing of 129<sup>3</sup> and a van der Waals definition of the solute cavity using AMOEBA force field radii. Since Born radii are used in the calculation of cross-term energies, the HCT scale factor was chosen before finalizing the generalizing constant by minimizing the mean signed error (MSE) between PB and GK self-energy values. The final HCT scale factor of 0.72 gave an MSE of  $-0.10$ , mean unsigned error (MUE) of 0.29, and root mean square error (RMSE) of 0.87 (Figure 7A). The slightly negative MSE is compensated for implicitly during the fitting of base solute radii described below.

Testing with generalizing constants ( $c_{\text{GK}}$ ) between 2 and 4 showed little change in the MSE between PB and GK cross-term energies. The final value of 2.455 was chosen for consistency with prior work<sup>45</sup> and gave an MSE of 0.07, MUE of 0.35, and RMSE of 1.17 (Figure 7B). In both cases, GK energies are strongly correlated with PB energies, with  $R^2$  values of 0.996 and 0.994, respectively.

#### Target Function for Electrostatic Radii Optimization.

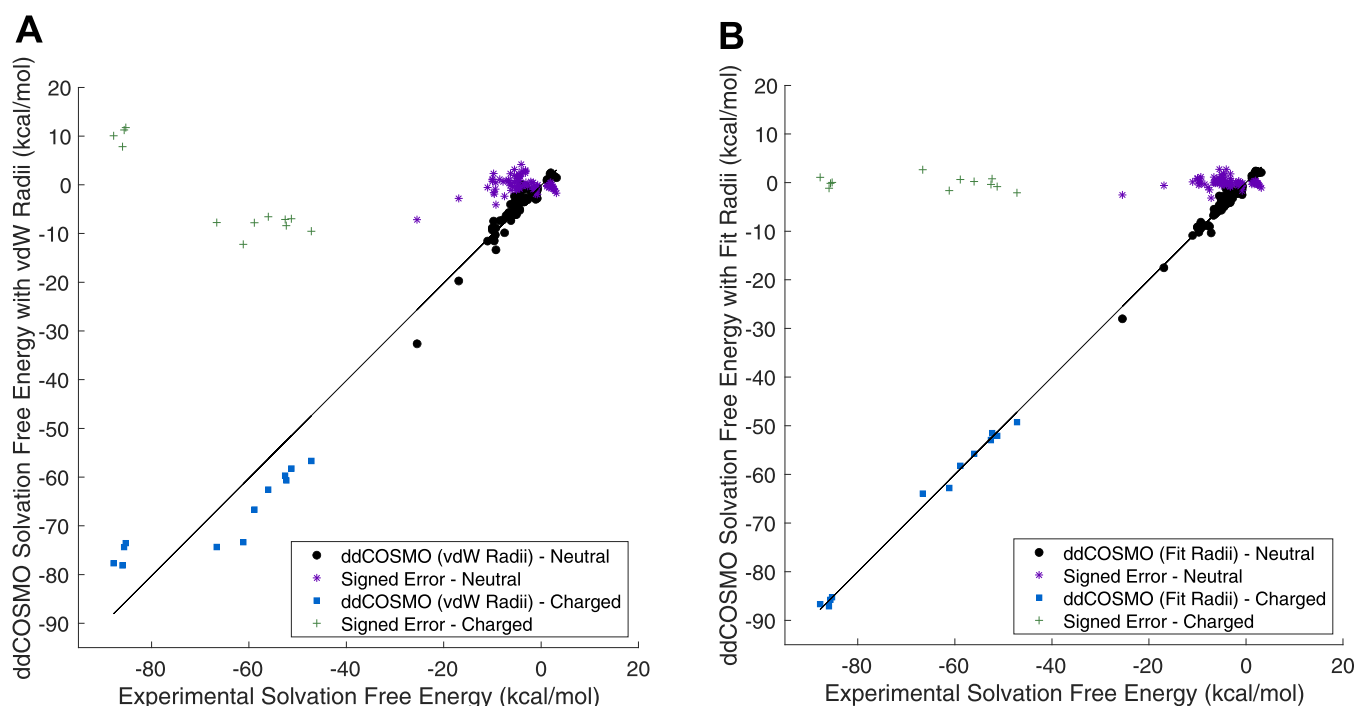
Electrostatic radii for 41 atom types were fit using the following target function

$$E(\mathbf{P}) = W_{\text{MUE}} \sum_{i=1}^n (\Delta G_i^{\text{Expt}} - \Delta G_i^{\text{Model}}(\mathbf{P}))^2 + W_{\text{MSE}} \left( \sum_{i=1}^n \Delta G_i^{\text{Expt}} - \sum_{i=1}^n \Delta G_i^{\text{Model}}(\mathbf{P}) \right)^2 + W_{\text{Regularization}} \sum_{i=1}^{N_{\text{radii}}} (R_i^{\text{Model}} - R_i^{\text{vdW}})^2 \quad (36)$$

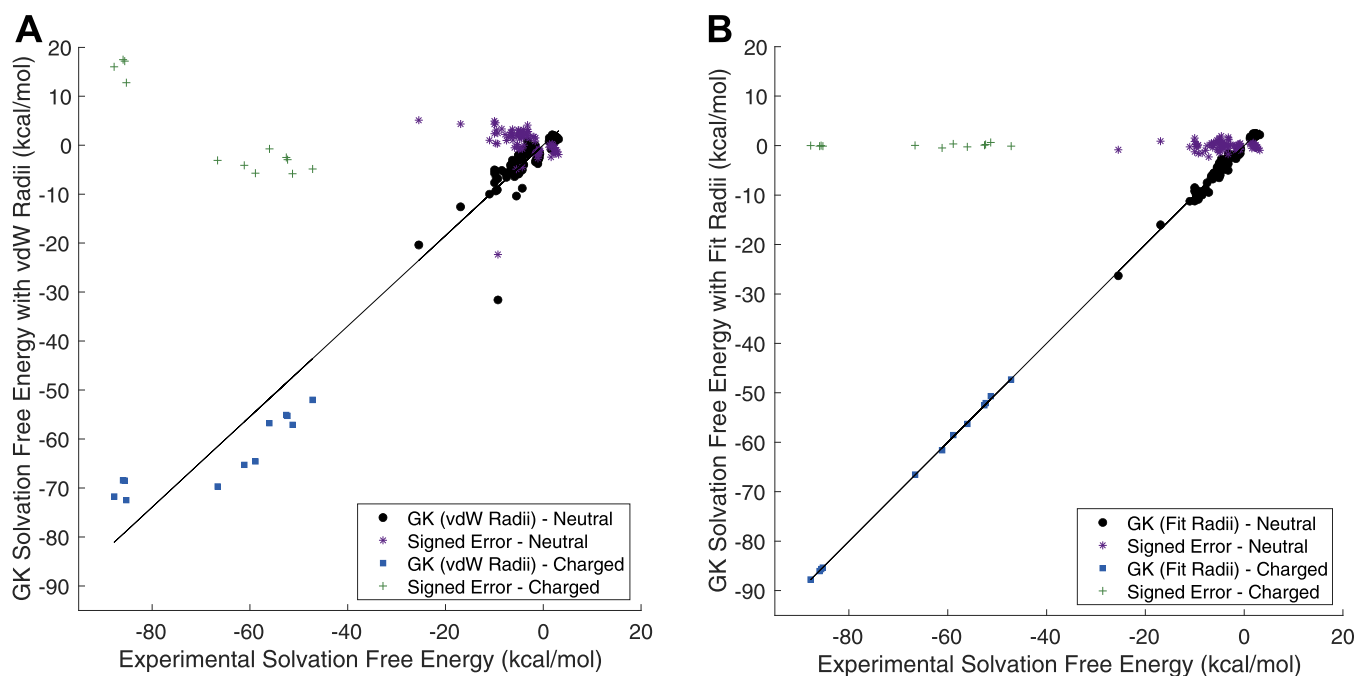
where the first term favors minimizing the unsigned error between experimental and model solvation free energy over  $n$  molecules, the second term favors minimizing the overall signed error, and the final term penalizes electrostatic radii that deviate from the AMOEBA force field definition of minimum energy van der Waals separation ( $R_{\min}$ ). The optimization was performed using an L-BFGS minimizer for each of the APBS, ddCOSMO, and GK models. The optimization was seeded with electrostatic radii based on AMOEBA van der Waals  $R_{\min}$  values, and after trial and error, all three optimization weights were set to 1.0.

## RESULTS

**Small-Molecule Hydration Free Energy.** The nonpolar portion of the model consists of an unfavorable cavitation free



**Figure 9.** Comparison of experimental and ddCOSMO solvation free energy differences using either AMOEBA van der Waals  $R_{\min}$  radii to describe the solute–solvent boundary (A) or using fit radii is shown (B). When using AMOEBA  $R_{\min}$  radii, the linear regression gave  $Y = 1.0017 \cdot X - 0.1465$  with  $R^2 = 0.9738$ . When using fit radii, the linear regression gave  $Y = 1.0001 \cdot X + 0.0015$  with  $R^2 = 0.9981$ .



**Figure 10.** Comparison of experimental and GK solvation free energy differences using either AMOEBA van der Waals  $R_{\min}$  radii to describe the solute–solvent boundary (A) or using fit radii is shown (B). When using AMOEBA  $R_{\min}$  radii, the linear regression gave  $Y = 0.9245 \cdot X + 0.0360$  with  $R^2 = 0.9567$ . When using fit radii, the linear regression gave  $Y = 0.9999 \cdot X - 0.0046$  with  $R^2 = 0.9987$ .

energy term and a favorable dispersion free energy term. One tunable parameter, solvent pressure, was used in calculating cavitation, and two tunable parameters, a dispersion offset and an atomic overlap scale factor, were used in calculating dispersion. The solvent pressure of  $0.0334 \text{ kcal}/(\text{mol } \text{\AA}^3)$  for cavitation was chosen based on previous testing. The dispersion integral offset ( $d = 1.056 \text{ \AA}$ ) and the unitless HCT dispersion

atomic overlap scale factor ( $s = 0.75$ ) were chosen to match dispersion values from solute–solvent enthalpy simulations in explicit water (Table S1, Supporting Information).

For the electrostatic portion of the model, a total of 41 solute radii classes were optimized using L-BFGS minimization as described above. Solvation free energy difference values from the FreeSolv database<sup>71,72</sup> were used as experimental targets. Each

radii class was determined based on SMARTS strings that were automatically generated for each atom type by PolType2. A total of 78 unique SMARTS strings were generated for the test set of 103 molecules. These 78 SMARTS strings were then reduced into 41 groups based on element, chemical environment, and electrostatic radii sizes from an initial optimization using all SMARTS strings under GK electrostatics (Table S2, Supporting Information). The use of fewer parameters helps to avoid overfitting and improves generalizability. Optimization was performed using the 41 radii classes to parameterize the APBS, ddCOSMO, and GK electrostatic models. Parameterized radii deviated from original AMOEBA van der Waals radii by an average of 9.3% for APBS, 9.9% for ddCOSMO, and 14.9% for GK. The quality of the resulting implicit solvent model for small molecules using APBS, ddCOSMO, and GK is shown below in Figures 8–10, respectively, and full data is available in Tables S3–S8, Supporting Information.

The data shown in Figures 8–10 is summarized in Tables 1A and 1B, which give the root mean square error (RMSE) and

**Table 1A. RMSE/MSE Values for Test Molecules Using AMOEBA van der Waals Radii Are Given by Functional Group Categories and Overall (kcal/mol)**

| functional group      | N   | AMOEBA radii |            |             |
|-----------------------|-----|--------------|------------|-------------|
|                       |     | APBS         | ddCOSMO    | GK          |
| alkanes               | 18  | 0.50/−0.24   | 0.71/−0.47 | 1.16/−0.89  |
| alcohols and phenols  | 16  | 1.48/+1.27   | 2.02/−0.19 | 2.40/+2.09  |
| amines                | 12  | 2.51/+2.40   | 2.36/+2.17 | 3.04/+1.10  |
| amides                | 8   | 2.25/+1.98   | 1.28/+0.40 | 3.18/+2.63  |
| nitrogen heterocyclic | 8   | 2.02/+0.99   | 1.93/+0.14 | 2.33/+1.60  |
| arenes                | 5   | 0.77/−0.57   | 0.95/−0.56 | 1.35/−0.54  |
| ethers                | 5   | 0.86/+0.80   | 0.74/+0.49 | 1.45/+1.41  |
| oxanes and oxines     | 4   | 1.05/+0.79   | 0.58/−0.16 | 2.29/+1.75  |
| thiols                | 4   | 1.68/−1.63   | 1.25/−1.16 | 1.93/−1.85  |
| carboxylic acids      | 3   | 1.31/+1.18   | 0.68/+0.38 | 2.58/+2.39  |
| sulfides              | 3   | 1.20/−1.17   | 0.61/−0.48 | 1.49/−1.46  |
| aldehydes             | 2   | 0.61/+0.58   | 0.22/−0.16 | 1.52/+1.52  |
| other                 | 3   | 1.55/+0.49   | 3.01/−1.38 | 13.04/−6.04 |
| total neutrals        | 91  | 1.58/+0.76   | 1.58/+0.09 | 3.22/+0.62  |
| charged               | 12  | 9.90/+0.01   | 9.13/−2.12 | 9.80/+2.80  |
| total                 | 103 | 3.69/+0.67   | 3.45/−0.17 | 4.51/+0.87  |

mean signed error (MSE) between experimental and computed solvation free energy differences. Overall, parameterized solute radii resulted in RMSE/MUE/MSE values of 1.00/0.70/0.05, 0.92/0.63/0.00, and 0.75/0.58/0.00 kcal/mol for the APBS, ddCOSMO, and GK models, respectively (Table 1B).

**Comparison to Explicit Solvent Free Energy Differences.** Although the implicit solvents described here were fit to experimental data, a direct comparison to AMOEBA explicit solvent hydration free energy differences helps illuminate if the continuum models are either overfit or if they exhibit relatively higher errors. A subset of 26 neutral small molecules used to parameterize the implicit solvent model is compared to available data from a recent AMOEBA explicit solvent study<sup>120</sup> in Table 2. The explicit solvent gave an RMSE of 0.70 kcal/mol compared to the experiment, while implicit solvents using APBS, ddCOSMO, and GK electrostatics gave RMSEs of 0.91, 0.65, and 0.63 kcal/mol, respectively. The concordance between the RMSEs for the explicit and implicit hydration free energy

**Table 1B. RMSE/MSE Values for Test Molecules Using Parameterized Electrostatic Radii Are Given by Functional Group Categories and Overall (kcal/mol)**

| functional group      | N   | fit solute radii |            |            |
|-----------------------|-----|------------------|------------|------------|
|                       |     | APBS             | ddCOSMO    | GK         |
| alkanes               | 18  | 0.49/−0.24       | 0.42/−0.20 | 0.49/−0.17 |
| alcohols and phenols  | 16  | 1.01/+0.50       | 1.31/+0.65 | 1.00/+0.56 |
| amines                | 12  | 1.49/+0.92       | 1.37/+0.06 | 0.92/−0.20 |
| amides                | 8   | 0.66/+0.13       | 0.66/+0.13 | 0.76/−0.12 |
| nitrogen heterocyclic | 8   | 1.31/+0.56       | 0.50/+0.07 | 0.88/+0.27 |
| arenes                | 5   | 0.50/−0.44       | 1.00/−0.61 | 0.80/−0.20 |
| ethers                | 5   | 0.23/−0.03       | 0.35/−0.11 | 0.57/−0.36 |
| oxanes and oxines     | 4   | 1.16/−0.33       | 0.58/−0.36 | 1.05/+0.07 |
| thiols                | 4   | 0.65/−0.49       | 0.38/−0.26 | 0.32/−0.09 |
| carboxylic acids      | 3   | 0.09/−0.08       | 0.11/−0.01 | 0.24/+0.00 |
| sulfides              | 3   | 0.36/−0.33       | 0.19/+0.09 | 0.32/−0.29 |
| aldehydes             | 2   | 0.48/−0.38       | 0.39/−0.39 | 0.11/−0.03 |
| other                 | 3   | 1.96/−0.06       | 0.88/−0.56 | 0.51/−0.31 |
| total neutrals        | 91  | 0.97/+0.14       | 0.87/+0.01 | 0.76/+0.00 |
| charged               | 12  | 1.18/−0.59       | 1.25/−0.07 | 0.73/−0.03 |
| total                 | 103 | 1.00/+0.05       | 0.92/+0.00 | 0.75/+0.00 |

differences supports the conclusion that the continuum models are neither clearly overfit nor of worse quality than what is observed for AMOEBA solutes in the explicit solvent.

**Validation Simulations on Proteins, DNA, and RNA.** Nine nucleic acids ( $\leq 24$  nucleotides) and nine proteins ( $\leq 129$  residues) of modest size were used to test the electrostatic energy and polarization response of the implicit solvents. Of the nucleic acids, seven were RNA and two were DNA. Starting structures for all 18 validation set molecules are shown in Figure 11 and were obtained from the Protein Data Bank (PDB). In the case of NMR ensembles, the first conformer was used. For explicit solvent simulations, each molecule was solvated in an explicit water box with neutralizing sodium or chloride ions. With the validation set molecule fixed, minimization to an RMS gradient of 0.1 kcal/(mol Å) was performed on each system to allow for relaxation of the water and ions.

Electrostatic energies were calculated for each validation set molecule with van der Waals and fit radii using all three electrostatics models. When using van der Waals radii (Figure 12A), the mean GK electrostatic energy was 1.65% more positive than APBS ( $R^2 = 0.9999$ ), while that for ddCOSMO was 6.71% more negative ( $R^2 = 0.9990$ ). The relatively large disparity between ddCOSMO and APBS when using identical van der Waals radii suggests that the COSMO approximation is not entirely ameliorated by its empirical scaling function. When using fit radii (Figure 12A), the mean GK electrostatic energy was 0.01% more negative than APBS ( $R^2 = 0.9991$ ), while that for ddCOSMO was 0.11% more positive ( $R^2 = 0.9991$ ). This shows that the COSMO approximation is compensated for to a large degree implicitly during the fitting of electrostatic radii. The slight reduction in  $R^2$  between APBS and GK electrostatic energies moving from van der Waals radii to fit radii suggests that the use of a single HCT overlap scale factor is not perfectly transferable across the range of atomic sizes found in proteins and nucleic acids. The optimization of element-specific HCT scale factors for GK electrostatics will be explored in future work that also focuses on handling interstitial spaces too small to accommodate water molecules (*i.e.*, calculation of Born radii based on integration over a molecular volume rather than a van der Waals volume).

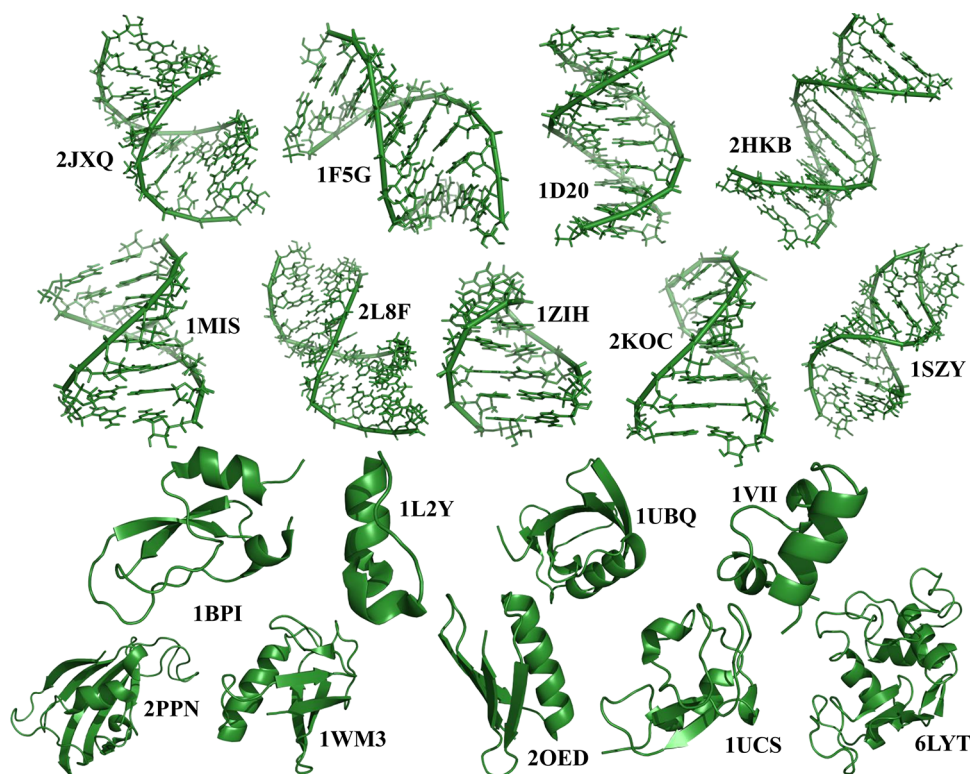
**Table 2.** Comparison of Solvation Free Energy Differences in AMOEBA Explicit and Implicit Solvents to Experimental Solvation Free Energy Differences (All Values in kcal/mol)

| molecule                  | $\Delta G_{\text{expt}}$ | explicit solvent |       | $\Delta G_{\text{implicit}}$ |        |        | signed error |       |       |
|---------------------------|--------------------------|------------------|-------|------------------------------|--------|--------|--------------|-------|-------|
|                           |                          | $\Delta G$       | error | APBS                         | COSMO  | GK     | APBS         | COSMO | GK    |
| isopropanol               | −4.74                    | −4.21            | 0.28  | −3.77                        | −3.66  | −3.92  | 0.97         | 1.08  | 0.82  |
| hydrogen sulfide          | −0.70                    | −0.41            | 0.08  | −1.06                        | −0.77  | −0.80  | −0.36        | −0.07 | −0.10 |
| <i>p</i> -cresol          | −6.13                    | −5.6             | 0.28  | −6.22                        | −5.37  | −4.99  | −0.09        | 0.76  | 1.14  |
| dimethylsulfide           | −1.61                    | −1.85            | 0.06  | −2.11                        | −1.56  | −1.99  | −0.50        | 0.05  | −0.38 |
| phenol                    | −6.60                    | −5.05            | 2.40  | −6.23                        | −5.48  | −5.49  | 0.37         | 1.12  | 1.11  |
| benzene                   | −0.90                    | −1.23            | 0.11  | −1.30                        | −1.70  | −1.54  | −0.40        | −0.80 | −0.64 |
| ethanol                   | −5.00                    | −4.69            | 0.10  | −4.09                        | −3.89  | −3.57  | 0.91         | 1.11  | 1.43  |
| ethane                    | 1.83                     | 1.73             | 0.01  | 2.31                         | 2.10   | 2.54   | 0.48         | 0.27  | 0.71  |
| <i>n</i> -butane          | 2.10                     | 1.11             | 0.98  | 2.09                         | 2.21   | 1.80   | −0.01        | 0.11  | −0.30 |
| methylamine               | −4.55                    | −5.46            | 0.83  | −4.00                        | −4.53  | −4.71  | 0.55         | 0.02  | −0.16 |
| dimethylamine             | −4.29                    | −3.04            | 1.56  | −2.72                        | −4.03  | −5.48  | 1.57         | 0.26  | −1.19 |
| trimethylamine            | −3.20                    | −2.09            | 1.23  | −3.34                        | −3.09  | −1.78  | −0.14        | 0.11  | 1.42  |
| propane                   | 2.00                     | 1.69             | 0.10  | 2.24                         | 2.28   | 2.39   | 0.24         | 0.28  | 0.39  |
| methane                   | 2.00                     | 1.73             | 0.07  | 2.43                         | 2.39   | 1.89   | 0.43         | 0.39  | −0.11 |
| methanol                  | −5.10                    | −4.79            | 0.10  | −3.55                        | −3.51  | −3.40  | 1.55         | 1.59  | 1.70  |
| <i>n</i> -propanol        | −4.85                    | −4.85            | 0.00  | −4.33                        | −4.53  | −4.23  | 0.52         | 0.32  | 0.62  |
| toluene                   | −0.90                    | −1.53            | 0.40  | −1.41                        | −1.60  | −1.12  | −0.51        | −0.70 | −0.22 |
| ethylbenzene              | −0.79                    | −0.8             | 0.00  | −1.21                        | −1.41  | −0.87  | −0.42        | −0.62 | −0.08 |
| <i>n</i> -methylacetamide | −10.00                   | −8.66            | 1.80  | −9.50                        | −9.09  | −9.05  | 0.50         | 0.91  | 0.95  |
| water                     | −6.30                    | −5.86            | 0.19  | −5.36                        | −6.34  | −6.27  | 0.94         | −0.04 | 0.03  |
| acetic acid               | −6.69                    | −5.63            | 1.12  | −6.56                        | −6.75  | −6.37  | 0.13         | −0.06 | 0.32  |
| methylethylsulfide        | −1.50                    | −1.98            | 0.23  | −1.64                        | −1.19  | −1.89  | −0.14        | 0.31  | −0.39 |
| imidazole                 | −9.63                    | −10.25           | 0.38  | −9.43                        | −9.79  | −9.36  | 0.20         | −0.16 | 0.27  |
| acetamide                 | −9.70                    | −9.3             | 0.16  | −10.78                       | −10.22 | −10.46 | −1.08        | −0.52 | −0.76 |
| ethylamine                | −4.50                    | −4.33            | 0.03  | −4.14                        | −4.97  | −4.28  | 0.36         | −0.47 | 0.22  |
| pyrrolidine               | −5.48                    | −4.88            | 0.36  | −2.31                        | −4.61  | −4.24  | 3.17         | 0.87  | 1.24  |
| MSE                       |                          |                  | 0.19  |                              |        |        | 0.36         | 0.24  | 0.31  |
| MUE                       |                          |                  | 0.57  |                              |        |        | 0.64         | 0.50  | 0.64  |
| RMSE                      |                          |                  | 0.70  |                              |        |        | 0.91         | 0.65  | 0.63  |

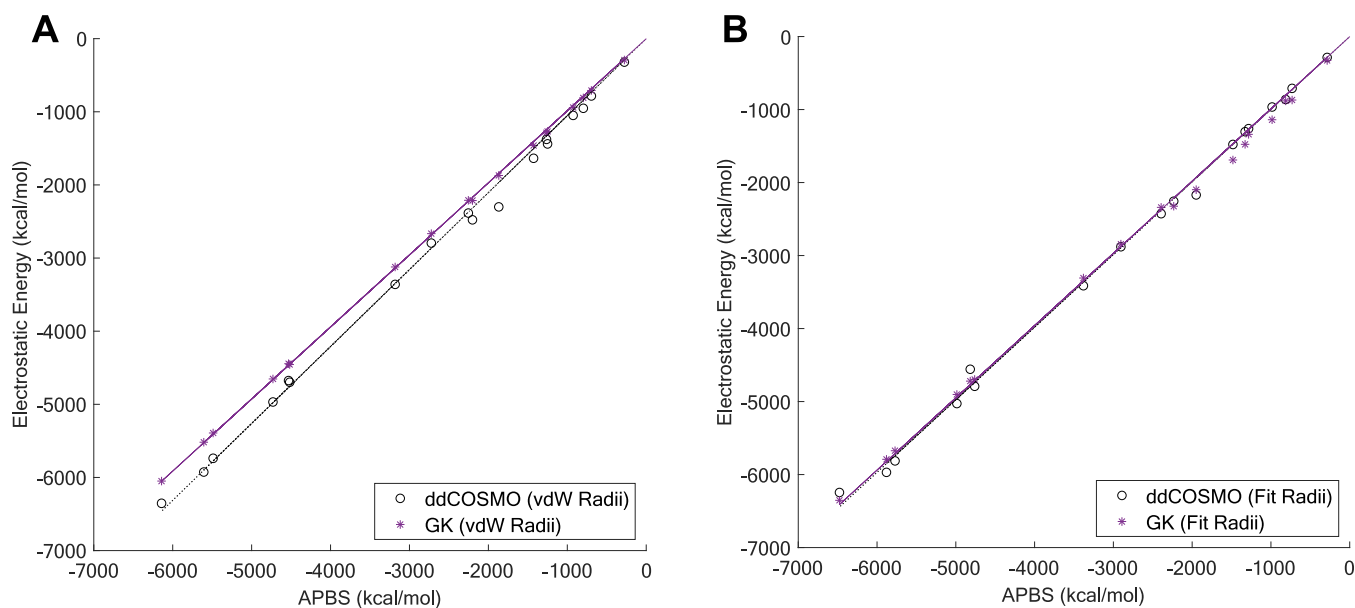
Dipole moment magnitudes were calculated for each validation set molecule in vacuum, explicit solvent, and implicit solvent. Prior to computing explicit solvent dipole moments, a series of short molecular dynamics (MD) simulations were used to equilibrate the system. Ensemble average dipole moment values were then calculated from 1 ns MD simulations with the biomolecule fixed (*i.e.*, solvent degrees of freedom were converged). For detailed simulation conditions, see Table S10, Supporting Information. In addition, dipole moment magnitudes were calculated for the validation set using three fixed-charge force fields: AMBER ff99SB,<sup>121</sup> OPLS-AA/L,<sup>122</sup> and CHARMM22/CMAP.<sup>123,124</sup> Tinker version 8.8.1 (August 2020) did not include nucleic acid force field parameters for OPLS-AA/L or CHARMM22/CMAP; therefore, only AMOEBA and AMBER ff99SB dipole moment magnitudes are reported for nucleic acids. Dipole moment magnitudes calculated using available force fields for nucleic acids and proteins are presented in Figure 13. All fixed-charge dipole moment magnitudes are plotted against ensemble average AMOEBA explicit solvent dipole moment magnitudes, as well as AMOEBA vacuum dipole moments. AMBER ff99SB, OPLS-AA/L, and CHARMM22/CMAP dipole moment magnitudes had  $R^2$  values of 0.987, 0.979, and 0.982, respectively, when compared to AMOEBA vacuum dipole moment magnitudes, and  $R^2$  values of 0.998, 0.996, and 0.996, respectively, when compared to AMOEBA explicit solvent dipole moment magnitudes. The better agreement of the fixed-charge force fields with AMOEBA in explicit (or implicit) solvent, relative to

the worse agreement with AMOEBA in vacuum, is consistent with fixed-charge biomolecular force fields being prepolarized for aqueous simulations. This also demonstrates that the AMOEBA electrostatics model produces molecular dipole moments that are consistent with previous generation force fields.

AMOEBA implicit solvent dipole moment magnitudes were calculated using each of the three parameterized electrostatics models and compared to those from explicit solvent ensemble averages, as shown in Figure 14. All three electrostatics models achieved near-perfect correlation with explicit solvent values based on  $R^2$  values for the APBS, ddCOSMO, and GK models of 0.999 in each case. Notably, each AMOEBA implicit solvent electrostatics model produces dipoles moments that agree more closely with AMOEBA in explicit solvent than does any fixed-charge force field. This supports the conclusion that if the AMOEBA biomolecular dipole moments are closer to reality than those from any of the fixed-charge models, then AMOEBA simulations in implicit solvent are in some ways more realistic than fixed-charge simulations in explicit solvent. The relative merits of fixed-charge explicit water simulations compared to polarizable implicit solvent simulations will be further explored in the future. In Figure 14, dipole moment magnitudes for the APBS electrostatics model were calculated without implicitly adding ions. A full comparison of dipole moment magnitudes and electrostatic energy components for all validation set molecules is available in Tables S11–S13, Supporting Information. This includes dipole moment magnitudes



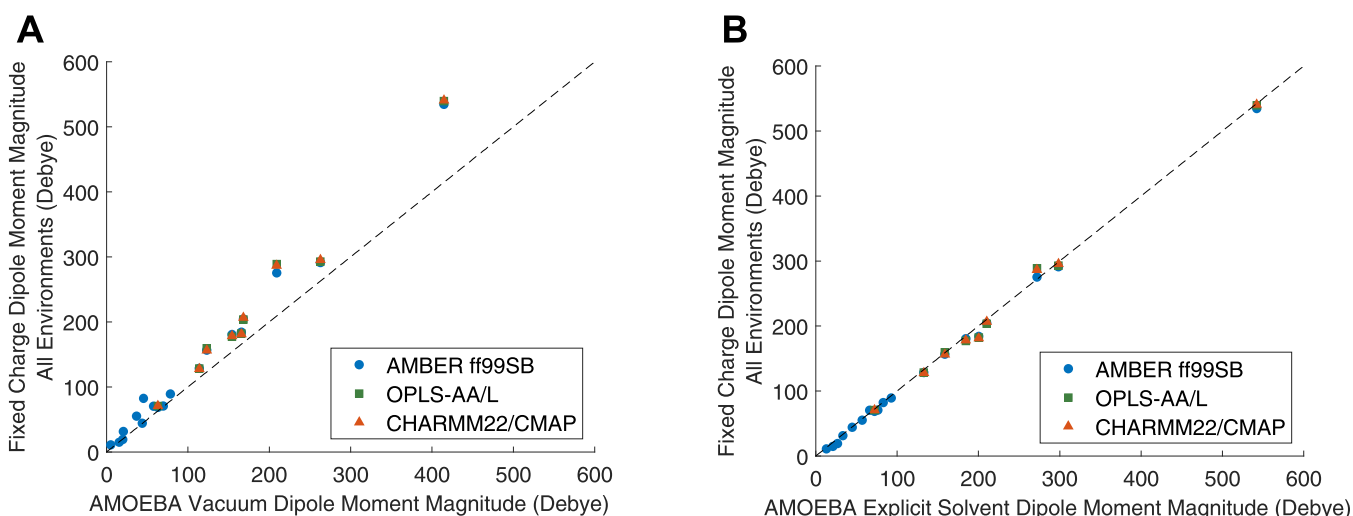
**Figure 11.** Validation set includes nine nucleic acids and nine proteins. The nucleic acid set can be further broken down into sets of four RNA helices (2JXQ, 1F5G, 1MIS, and 2L8F), three RNA hairpins (1ZIH, 2KOC, and 1SZY), and two DNA double helices (1D20 and 2HKB). Additional information on individual molecules and simulation conditions can be found in [Table S10](#), Supporting Information. Coordinate files for each molecule (Tinker “XYZ” format) are provided in the [Supporting Information](#).



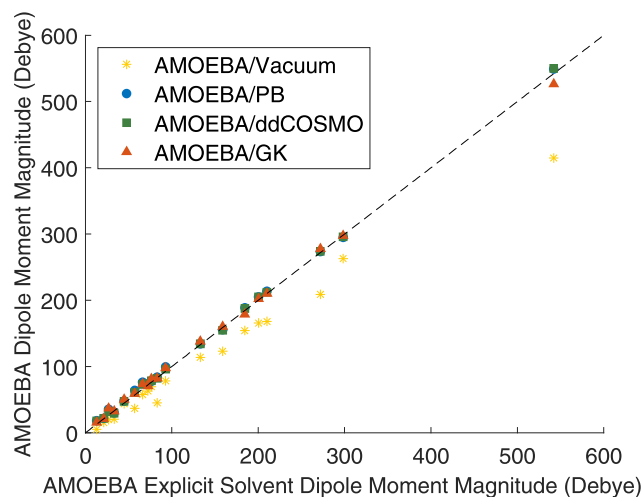
**Figure 12.** APBS energy values for the biomolecular validation set are plotted vs ddCOSMO and GK for both van der Waals radii and fit radii. (A) APBS electrostatics energy using van der Waals radii vs ddCOSMO ( $Y = 1.0523X$ ,  $R^2 = 0.9990$ ) and GK ( $Y = 0.9854X$ ,  $R^2 = 0.9999$ ). (B) APBS PBE electrostatics energy using fit radii vs ddCOSMO ( $Y = 0.9884X$ ,  $R^2 = 0.9991$ ) and GK ( $Y = 0.9833X$ ,  $R^2 = 0.9991$ ). Full data is available in [Tables S11 and S13](#), Supporting Information.

calculated in APBS using the nonlinear and linearized forms of the PBE with 150 mM salt. Overall, these results suggest the electrostatics models scale up well from initial optimization against small-molecule hydration free energy differences to applications to larger biomolecules.

The performances of the APBS, ddCOSMO, and GK electrostatics models implemented in Tinker were compared by timing an energy and gradient calculation on a single CPU core (Intel Xeon CPU E5-2680 v4 at 2.40 GHz). Calculations were performed for one nucleic acid (1ZIH) and one protein



**Figure 13.** Comparison of dipole moment magnitudes for fixed-charge force fields and AMOEBA across environments. (A) AMOEBA vacuum dipole moment magnitudes vs those for fixed-charge force fields (AMBER ff99SB  $R^2$ , 0.987; OPLS-AA/L  $R^2$ , 0.979; CHARMM22/CMAP  $R^2$ , 0.982). (B) AMOEBA explicit solvent dipole moment magnitudes vs those for fixed-charge force fields (AMBER ff99SB  $R^2$ , 0.998; OPLS-AA/L  $R^2$ , 0.996; CHARMM22/CMAP  $R^2$ , 0.996). Dashed lines at  $x = y$  are a guide to the eye.



**Figure 14.** Comparison of dipole moment magnitudes for AMOEBA in explicit solvent vs vacuum, PB, ddCOSMO, and GK environments for the validation set (vacuum  $R^2$ : 0.990, PB/ddCOSMO/GK  $R^2$ : 0.999 in each case). The dashed line at  $x = y$  is a guide to the eye.

(1VII) from the validation set. Results in Table 3 show that APBS is the costliest model, while GK is currently the most efficient. Timings for the same systems using AMOEBA/GK electrostatics implemented within FFX-OpenMM and executing on a single GPU (NVIDIA GeForce RTX 2080 Ti) were also collected. The latter performance (less than 0.005 s per time

**Table 3.** Performance of the APBS, ddCOSMO, and GK Electrostatics Models Implemented in Tinker on a Single CPU Core and for the GK Model Using FFX-OpenMM on a Single GPU

| molecule | calculation of the energy and gradient (s) |         |       |                    |
|----------|--------------------------------------------|---------|-------|--------------------|
|          | Tinker (1 CPU thread)                      |         |       | FFX-OpenMM (1 GPU) |
|          | APBS                                       | ddCOSMO | GK    | GK                 |
| 1ZIH     | 48.804                                     | 11.285  | 0.250 | 0.0042             |
| 1VII     | 53.772                                     | 21.245  | 0.533 | 0.0047             |

step) is consistent with the molecular dynamics performance of  $\sim 20$  ns/day using a conservative 1 fs integration scheme, which opens the door to tuning the AMOEBA/GK continuum model using extensive simulations of proteins and nucleic acids. As of this writing, the APBS and ddCOSMO electrostatics models are not yet available for use on GPUs.

## CONCLUSIONS

Implicit solvent models were developed for use with the polarizable AMOEBA force field. Novel cavitation and dispersion nonpolar terms were designed to replicate explicit solvent free energy differences using only three free parameters: a single cavitation parameter to describe solvent pressure for small cavities and two dispersion parameters (one to define the beginning of the dispersal integral and a second to account for atomic overlaps during integration). Based on these nonpolar terms, the solute–solvent electrostatic boundary (*i.e.*, atomic radii) was optimized for three continuum electrostatics models, APBS, ddCOSMO, and GK, using numerical optimization against experimental small-molecule solvation free energy differences. Overall, the APBS, ddCOSMO, and GK models produced mean unsigned errors of 0.70, 0.63, and 0.58 kcal/mol, respectively, compared to the experiment. All three implicit solvent models produced hydration free energy difference RMSEs within 0.2 kcal/mol of AMOEBA explicit solvent solvation free energy difference simulations for a collection of 26 small molecules (Table 2). This supports the conclusion that the implicit solvent models presented here are of similar quality to that of explicit solvent for hydration free energy differences and are not clearly overfit to the test data (*i.e.*, overfitting might be suggested by implicit solvent RMSEs that are artificially much lower than those achieved by explicit solvent simulations).

Each small molecule used to parameterize the implicit solvent model fell within the volume-scaling regime of the cavitation model, such that the contribution to solvation was calculated using SEV. For larger proteins or nucleic acids, the cavitation free energy of the model falls within the surface-area-scaling regime. A future goal is to account for local molecular curvature to promote transferability of the cavitation free energy to biomolecules of complex shapes. The dispersion model

integrates the WCA attractive potential for each atom in the solute. This analytic, pairwise approach is well equipped to handle nonspherical solutes, which adds physical detail to the previously described Born radii-based dispersion model.<sup>73</sup> Both the cavitation and dispersion models described here are currently limited by their lack of treatment of interstitial spaces, which is elaborated on below.

To optimize the agreement of GK self-energies with the calculated *perfect* PB multipolar self-energies, it may be beneficial to use separate HCT scaling factors for each chemical element instead of a single parameter, as was done here. The precedent for this split is given in the original description of the HCT pairwise descreening approximation, where the scale factor magnitude generally decreases with the increasing atomic size.<sup>34</sup> Additionally, the agreement between GK cross-term energies and calculated *perfect* PB multipolar cross-term energies might be improved using separate generalizing function constants for monopoles, dipoles, and quadrupoles, rather than a single constant. The physical motivation is that the electrostatic potential is of a longer range for monopoles than for dipoles. Therefore, the transition between the Born ion regime (or the Kirkwood multipole regime) and the Coulomb regime, which is tuned by the constant in the generalizing function (eq 34), could in principle be optimized for each multipole order separately. For simplicity, this work used a single HCT scale factor (0.72) and a single generalizing constant (2.455) for GK.

At the length scale of small molecules, continuum electrostatics is known to be sensitive to the definition of the solute–solvent boundary,<sup>125–129</sup> and thus optimization of electrostatic radii is required to implicitly account for physical details like solute–water hydrogen bonding. Overall, the quality of the resulting models using fit solute radii for PB (RMSE 1.0, MSE 0.1), ddCOSMO (RMSE 0.9, MSE 0.0), and GK (RMSE 0.8, MSE 0.0) are comparable to that of the recent Drude/PB implicit solvent (RMSE 0.8, MSE 0.0).<sup>51</sup> The fit radii reproduce experimental solvation free energy differences better than the original van der Waals radii, which gave RMS errors of 3.69, 3.45, and 4.51 kcal/mol for APBS, ddCOSMO, and GK, respectively (Table 1A). Additionally, it may be beneficial to consider using optimized GK (or ddCOSMO) electrostatic radii as a starting point for electrostatics calculations in quantum mechanical continuum solvents.<sup>61,130</sup>

Dipole moment calculations using each AMOEBA implicit solvent for 18 protein and nucleic acid biomolecules show nearly exact agreement with explicit solvent dipole moments computed by averaging over solvent degrees of freedom (Figure 14). This suggests that all three models (APBS, ddCOSMO, and GK) successfully reproduce the polarization response observed in explicit water simulations at the resolution of overall biomolecules. Future work will focus on molecular dynamic simulations of biomolecules in implicit solvent compared to explicit solvent to access stability and the agreement of conformational ensembles. Furthermore, although the implicit solvent models discussed here have been developed for use with the AMOEBA polarizable force field, their support for polarizable atomic multipole electrostatics should permit adaptation to emerging models such as AMOEBA+ and HIPPO.<sup>131–133</sup>

An important limitation of the current models is their focus on the use of a van der Waals description of the solute for cavitation, dispersion, and electrostatic contributions, rather than a molecular surface.<sup>134,135</sup> The approximation of a van der Waals description is modest for small solutes but becomes

problematic as the molecular size and complexity increase (e.g., for biomolecules). For example, a simple van der Waals surface does not account for interstitial spaces (i.e., spaces between biomolecular residues or domains where water molecules cannot fit), and thereby allows for continuum water access to spaces not accessible to explicit water. Favorable hydration effects of continuum water in interstitial spaces promote swelling of biomolecules and oppose hydrophobic compaction forces. For this reason, future work must incorporate methods that have been proposed to account for interstitial spaces<sup>118,135–137</sup> into the AMOEBA family of implicit solvents.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.0c01286>.

Description of the electrostatic potential experienced by an empty cavity under three different simulation conditions due to the preferential orientation of water at boundaries; implicit solvent dispersion term fit to dispersion free energy differences calculated for a set of small molecules in explicit solvent; initial (van der Waals) and optimized solute diameters for APBS, ddCOSMO, and GK; solvation free energy difference components using AMOEBA van der Waals radii; solvation free energy differences using AMOEBA van der Waals radii; and signed and unsigned solvation free energy difference errors using AMOEBA van der Waals radii; solvation free energy difference components using fit electrostatic radii; solvation free energy differences using fit electrostatic radii; signed and unsigned solvation free energy difference errors using fit electrostatic radii; free energy differences for phosphate and guanidinium compounds from explicit solvent simulations; simulation conditions for validation set molecules with explicit solvent; electrostatic component of the solvation free energy and dipole moments for validation set molecules using van der Waals radii and fit radii; dipole moment magnitudes from fixed charge force fields (PDF)

AMOEBA parameters (Tinker “prm” format) and coordinate files (Tinker “XYZ” format) for all small molecules used in this work and coordinate files for all protein and nucleic acid validation molecules (ZIP)

## ■ AUTHOR INFORMATION

### Corresponding Author

Michael J. Schnieders – Roy J Carver Department of Biomedical Engineering, University of Iowa, Iowa City, Iowa 52242, United States; Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242, United States; [orcid.org/0000-0003-1260-4592](https://orcid.org/0000-0003-1260-4592); Email: [michael-schnieders@uiowa.edu](mailto:michael-schnieders@uiowa.edu)

### Authors

Rae A. Corrigan – Roy J Carver Department of Biomedical Engineering, University of Iowa, Iowa City, Iowa 52242, United States; [orcid.org/0000-0003-3628-9347](https://orcid.org/0000-0003-3628-9347)

Guowei Qi – Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242, United States

Andrew C. Thiel – Roy J Carver Department of Biomedical Engineering, University of Iowa, Iowa City, Iowa 52242, United States

Jack R. Lynn – Roy J Carver Department of Biomedical Engineering, University of Iowa, Iowa City, Iowa 52242, United States

Brandon D. Walker – Department of Biomedical Engineering, University of Texas in Austin, Austin, Texas 78712, United States

Thomas L. Casavant – Roy J Carver Department of Biomedical Engineering, University of Iowa, Iowa City, Iowa 52242, United States

Louis Lagardere – Department of Chemistry, Sorbonne Université, F-75005 Paris, France

Jean-Philip Piquemal – Department of Chemistry, Sorbonne Université, F-75005 Paris, France; [orcid.org/0000-0001-6615-9426](https://orcid.org/0000-0001-6615-9426)

Jay W. Ponder – Department of Chemistry, Washington University in St. Louis, St. Louis, Missouri 63130, United States; [orcid.org/0000-0001-5450-9230](https://orcid.org/0000-0001-5450-9230)

Pengyu Ren – Department of Biomedical Engineering, University of Texas in Austin, Austin, Texas 78712, United States; [orcid.org/0000-0002-5613-1910](https://orcid.org/0000-0002-5613-1910)

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acs.jctc.0c01286>

## Author Contributions

#R.A.C. and G.Q. contributed equally.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

All computations were performed on The University of Iowa Argon cluster with support and guidance from Danny Tang, Joe Hetrick, Glenn Johnson, and John Saxton. M.J.S. was supported by NIH R01DK110023, NIH R01DC012049, the Donald D. Harrington Fellows Program and NSF CHE-1751688. J.W.P. and P.R. were supported by R01GM106137 and R01GM114237. J.-P.P. and L.L. were supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program No 810367. R.A.C. and A.C.T. were supported by the NSF Graduate Research Fellowship Program NSF DGE-1945994, and G.Q. was supported by the Goldwater Foundation and the Iowa Center for Research by Undergraduates (ICRU).

## REFERENCES

- (1) Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **2005**, *437*, 640–647.
- (2) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338*, 1042–1046.
- (3) Zhang, Q. C.; Petrey, D.; Deng, L.; Qiang, L.; Shi, Y.; Thu, C. A.; Bisikirskaya, B.; Lefebvre, C.; Accili, D.; Hunter, T.; Maniatis, T.; Califano, A.; Honig, B. Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* **2012**, *490*, 556–560.
- (4) Ren, P.; Chun, J.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J.; Baker, N. A. Biomolecular electrostatics and solvation: a computational perspective. *Q. Rev. Biophys.* **2012**, *45*, 427–491.
- (5) LuCore, S. D.; Litman, J. M.; Powers, K. T.; Gao, S.; Lynn, A. M.; Tollefson, W. T. A.; Fenn, T. D.; Washington, M. T.; Schnieders, M. J. Dead-end elimination with a polarizable force field repacks PCNA structures. *Biophys. J.* **2015**, *109*, 816–826.
- (6) Huang, P.-S.; Boyken, S. E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016**, *537*, 320.
- (7) Hallen, M. A.; Martin, J. W.; Ojewole, A.; Jou, J. D.; Lowegard, A. U.; Frenkel, M. S.; Gainza, P.; Nisonoff, H. M.; Mukund, A.; Wang, S.; Holt, G. T.; Zhou, D.; Dowd, E.; Donald, B. R. OSPREY 3.0: Open-

source protein redesign for you, with powerful new features. *J. Comput. Chem.* **2018**, *39*, 2494–2507.

(8) Villa, F.; Panel, N.; Chen, X. Y.; Simonson, T. Adaptive landscape flattening in amino acid sequence space for the computational design of protein:peptide binding. *J. Chem. Phys.* **2018**, *149*, No. 072302.

(9) Tollefson, M. R.; Litman, J. M.; Qi, G.; O'Connell, C. E.; Wipfler, M. J.; Marini, R. J.; Bernabe, H. V.; Tollefson, W. T. A.; Braun, T. A.; Casavant, T. L.; Smith, R. J. H.; Schnieders, M. J. Structural insights into hearing loss genetics from polarizable protein repacking. *Biophys. J.* **2019**, *117*, 602–612.

(10) Cheatham, T. E.; Case, D. A. Twenty-five years of nucleic acid simulations. *Biopolymers* **2013**, *99*, 969–977.

(11) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA* **2015**, *21*, 1578–1590.

(12) Musil, M.; Konegger, H.; Hong, J.; Bednar, D.; Damborsky, J. Computational Design of Stable and Soluble Biocatalysts. *ACS Catal.* **2019**, *9*, 1033–1054.

(13) Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999**, *78*, 1–20.

(14) Onufriev, A. V.; Izadi, S. Water models for biomolecular simulations. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1347.

(15) Tan, C.; Tan, Y. H.; Luo, R. Implicit nonpolar solvent models. *J. Phys. Chem. B* **2007**, *111*, 12263–12274.

(16) Decherchi, S.; Masetti, M.; Vyalov, I.; Rocchia, W. Implicit solvent methods for free energy estimation. *Eur. J. Med. Chem.* **2015**, *91*, 27–42.

(17) Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. On the nonpolar hydration free energy of proteins: Surface area and continuum solvent models for the solute-solvent interaction energy. *J. Am. Chem. Soc.* **2003**, *125*, 9523–9530.

(18) Wagoner, J. A.; Baker, N. A. Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8331–8336.

(19) Chen, Z.; Zhao, S.; Chun, J.; Thomas, D. G.; Baker, N. A.; Bates, P. W.; Wei, G. W. Variational approach for nonpolar solvation analysis. *J. Chem. Phys.* **2012**, *137*, No. 084101.

(20) Harris, R. C.; Pettitt, B. M. Effects of geometry and chemistry on hydrophobic solvation. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 14681–14686.

(21) Michael, E.; Polydorides, S.; Simonson, T.; Archontis, G. Simple models for nonpolar solvation: parameterization and testing. *J. Comput. Chem.* **2017**, *38*, 2509–2519.

(22) Gilson, M. K.; Honig, B. H. Calculation of electrostatic potentials in an enzyme active-site. *Nature* **1987**, *330*, 84–86.

(23) Gilson, M. K.; Honig, B. Calculation of the total electrostatic energy of a macromolecular system - solvation energies, binding-energies, and conformational-analysis. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 7–18.

(24) Jeanschales, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempczyk, A.; Hendrickson, T. F.; Still, W. C. Electrostatic contributions to solvation energies - comparison of free-energy perturbation and continuum calculations. *J. Am. Chem. Soc.* **1991**, *113*, 1454–1455.

(25) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144–1149.

(26) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.

(27) Simonson, T. Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243–252.

(28) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

(29) Tan, Y. H.; Luo, R. Continuum treatment of electronic polarization effect. *J. Chem. Phys.* **2007**, *126*, No. 094103.

- (30) Tan, Y. H.; Tan, C. H.; Wang, J.; Luo, R. Continuum polarizable force field within the Poisson-Boltzmann framework. *J. Phys. Chem. B* **2008**, *112*, 7675–7688.
- (31) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **2018**, *27*, 112–128.
- (32) Rashin, A. A.; Honig, B. Reevaluation of the Born model of ion hydration. *J. Phys. Chem. A* **1985**, *89*, 5588–5593.
- (33) Roux, B.; Yu, H. A.; Karplus, M. Molecular-basis for the Born model of ion solvation. *J. Phys. Chem. B* **1990**, *94*, 4683–4688.
- (34) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246*, 122–129.
- (35) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA continuum model for solvation: a fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **1997**, *101*, 3005–3014.
- (36) Nina, M.; Beglov, D.; Roux, B. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. B* **1997**, *101*, 5239–5248.
- (37) Gallicchio, E.; Levy, R. M. AGBNP: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* **2004**, *25*, 479–499.
- (38) Gallicchio, E.; Paris, K.; Levy, R. M. The AGBNP2 implicit solvation model. *J. Chem. Theory Comput.* **2009**, *9*, 2544–2564.
- (39) Mukhopadhyay, A.; Aguilar, B. H.; Tolokh, I. S.; Onufriev, A. V. Introducing charge hydration asymmetry into the generalized Born model. *J. Chem. Theory Comput.* **2014**, *10*, 1788–1794.
- (40) Onufriev, A. V.; Case, D. A. Generalized Born implicit solvent models for biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 275–296.
- (41) Ren, P.; Wu, C.; Ponder, J. W. Polarizable atomic multipole-based molecular mechanics for organic molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143–3161.
- (42) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA polarizable atomic multipole force field for nucleic acids. *J. Chem. Theory Comput.* **2018**, *14*, 2084–2108.
- (43) Schnieders, M. J.; Baker, N. A.; Ren, P. Y.; Ponder, J. W. Polarizable atomic multipole solutes in a Poisson-Boltzmann continuum. *J. Chem. Phys.* **2007**, *126*, No. 124114.
- (44) Lipparini, F.; Lagardère, L.; Raynaud, C.; Stamm, B.; Cancès, E.; Mennucci, B.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. Polarizable Molecular Dynamics in a Polarizable Continuum Solvent. *J. Chem. Theory Comput.* **2015**, *11*, 623–634.
- (45) Schnieders, M. J.; Ponder, J. W. Polarizable atomic multipole solutes in a generalized Kirkwood continuum. *J. Chem. Theory Comput.* **2007**, *3*, 2083–2097.
- (46) Kirkwood, J. G. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J. Chem. Phys.* **1934**, *2*, 351–361.
- (47) Maple, J. R.; Cao, Y. X.; Damm, W. G.; Halgren, T. A.; Kaminski, G. A.; Zhang, L. Y.; Friesner, R. A. A polarizable force field and continuum solvation methodology for modeling of protein-ligand interactions. *J. Chem. Theory Comput.* **2005**, *1*, 694–715.
- (48) Cortis, C. M.; Friesner, R. A. Numerical solution of the Poisson-Boltzmann equation using tetrahedral finite-element meshes. *J. Comput. Chem.* **1997**, *18*, 1591–1608.
- (49) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **2016**, *116*, 4983–5013.
- (50) Lemkul, J. A.; MacKerell, A. D., Jr. Polarizable force field for RNA based on the classical drude oscillator. *J. Comput. Chem.* **2018**, *39*, 2624–2646.
- (51) Aleksandrov, A.; Lin, F. Y.; Roux, B.; MacKerell, A. D. Combining the polarizable Drude force field with a continuum electrostatic Poisson-Boltzmann implicit solvation model. *J. Comput. Chem.* **2018**, *39*, 1707–1719.
- (52) Aleksandrov, A.; Roux, B.; MacKerell, A. D. pKa Calculations with the Polarizable Drude Force Field and Poisson–Boltzmann Solvation Model. *J. Chem. Theory Comput.* **2020**, *16*, 4655–4668.
- (53) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kucsera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (54) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (55) Poier, P. P.; Jensen, F. Including implicit solvation in the bond capacity polarization model. *J. Chem. Phys.* **2019**, *151*, No. 114118.
- (56) Poier, P. P.; Jensen, F. Polarizable charges in a generalized Born reaction potential. *J. Chem. Phys.* **2020**, *153*, No. 024111.
- (57) Cooper, C. D. A boundary-integral approach for the poisson–boltzmann equation with polarizable force fields. *J. Comput. Chem.* **2019**, *40*, 1680–1692.
- (58) Cooper, C. D.; Bardhan, J. P.; Barba, L. A. A biomolecular electrostatics solver using Python, GPUs and boundary elements that can handle solvent-filled cavities and Stern layers. *Comput. Phys. Commun.* **2014**, *185*, 720–729.
- (59) Cramer, C. J.; Truhlar, D. G. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.
- (60) Tomasi, J. Thirty years of continuum solvation chemistry: a review, and prospects for the near future. *Theor. Chem. Acc.* **2004**, *112*, 184–203.
- (61) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3093.
- (62) Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117–129.
- (63) Cancès, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032–3041.
- (64) Cramer, C. J.; Truhlar, D. G. An SCF solvation model for the hydrophobic effect and absolute free energies of aqueous solvation. *Science* **1992**, *256*, 213–217.
- (65) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. SM6: A density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters. *J. Chem. Theory Comput.* **2005**, *1*, 1133–1152.
- (66) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- (67) Klamt, A.; Schuurmann, G. COSMO - a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799–805.
- (68) Klamt, A. Conductor-like screening model for real solvents - a new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem. C* **1995**, *99*, 2224–2235.
- (69) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289.
- (70) Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.

- (71) Mobley, D. L.; Guthrie, J. P. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 711–720.
- (72) Matos, G. D. R.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; Mobley, D. L. Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the FreeSolv database. *J. Chem. Eng. Data* **2017**, *62*, 1559–1569.
- (73) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. The SGB/NP hydration free energy model based on the surface Generalized Born solvent reaction field and novel nonpolar hydration free energy estimators. *J. Comput. Chem.* **2002**, *23*, 517–529.
- (74) Rizzo, R. C.; Aynechi, T.; Case, D. A.; Kuntz, I. D. Estimation of absolute free energies of hydration using continuum methods: accuracy of partial charge models and optimization of nonpolar contributions. *J. Chem. Theory Comput.* **2006**, *2*, 128–139.
- (75) Bartmess, J. E. Negative Ion Energetics Data. *NIST Chemistry WebBook*; National Institute of Standards and Technology: Gaithersburg, MD, p 20899.
- (76) Hunter, E. P.; Lias, S. G. Proton Affinity Evaluation. *NIST Chemistry WebBook*; National Institute of Standards and Technology: Gaithersburg, MD, p 20899.
- (77) Stewart, R. *The Proton: Applications to Organic Chemistry*; Academic Press, Inc.: New York, 1985; Vol. 46.
- (78) Wang, Z. Polarizable Force Field Development, and Applications to Conformational Sampling and Free Energy Calculation. Dissertation, Washington University in St. Louis: St. Louis, MO, 2018.
- (79) Fawcett, W. R. Thermodynamic parameters for the solvation of monatomic ions in water. *J. Phys. Chem. B* **1999**, *103*, 11181–11185.
- (80) Ashbaugh, H. S. Convergence of molecular and macroscopic continuum descriptions of ion hydration. *J. Phys. Chem. B* **2000**, *104*, 7235–7238.
- (81) Wu, J. C.; Chattree, G.; Ren, P. Y. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012**, *131*, No. 1138.
- (82) Huang, D. M.; Chandler, D. Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8324–8327.
- (83) Huang, D. M.; Geissler, P. L.; Chandler, D. Scaling of hydrophobic solvation free energies. *J. Phys. Chem. B* **2001**, *105*, 6704–6709.
- (84) Huang, D. M.; Chandler, D. The hydrophobic effect and the influence of solute-solvent attractions. *J. Phys. Chem. B* **2002**, *106*, 2047–2053.
- (85) Chen, J. H.; Brooks, C. L. Critical importance of length-scale dependence in implicit modeling of hydrophobic interactions. *J. Am. Chem. Soc.* **2007**, *129*, 2444.
- (86) Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.
- (87) Connolly, M. L. Computation of molecular volume. *J. Am. Chem. Soc.* **1985**, *107*, 1118–1124.
- (88) Connolly, M. L. The molecular-surface package. *J. Mol. Graphics* **1993**, *11*, 139–143.
- (89) Schnieders, M. J. *Force Field X*, version 1.0; University of Iowa, 2021. <https://fx.biochem.uiowa.edu>.
- (90) Richards, F. M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176.
- (91) Schnieders, M. J. The Theory and Effect of Solvation on Biomolecules. Dissertation, Washington University in St. Louis: St. Louis, MO, 2007.
- (92) Ren, P.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.
- (93) Richmond, T. J. Solvent accessible surface area and excluded volume in proteins: Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.* **1984**, *178*, 63–89.
- (94) Kundrot, C. E.; Ponder, J. W.; Richards, F. M. Algorithms for calculating excluded volume and its derivatives as a function of molecule-conformation and their use in energy minimization. *J. Comput. Chem.* **1991**, *12*, 402–409.
- (95) Wesson, L.; Eisenberg, D. Atomic solvation parameters applied to molecular-dynamics of proteins in solution. *Protein Sci.* **1992**, *1*, 227–235.
- (96) C.R.C. *Handbook of Chemistry and Physics*, 58th ed.; Weast, R. C., Ed.; CRC Press, 1977.
- (97) Halgren, T. A. The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.* **1992**, *114*, 7827–7843.
- (98) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* **1971**, *54*, 5237–5247.
- (99) Chandler, D.; Weeks, J. D.; Andersen, H. C. Van der Waals picture of liquids, solids, and phase transformations. *Science* **1983**, *220*, 787–794.
- (100) Gallicchio, E.; Kubo, M. M.; Levy, R. M. Enthalpy-entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation. *J. Phys. Chem. B* **2000**, *104*, 6271–6285.
- (101) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem. D* **1996**, *100*, 19824–19839.
- (102) Lipparini, F.; Lagardère, L.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: I. Toward Massively Parallel Direct Space Computations. *J. Chem. Theory Comput.* **2014**, *10*, 1638–1651.
- (103) Aviat, F.; Levitt, A.; Stamm, B.; Maday, Y.; Ren, P.; Ponder, J. W.; Lagardère, L.; Piquemal, J.-P. Truncated Conjugate Gradient: An Optimal Strategy for the Analytical Evaluation of the Many-Body Polarization Energy and Forces in Molecular Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 180–190.
- (104) Simmonett, A. C.; Pickard, F. C., IV; Ponder, J. W.; Brooks, B. R. An empirical extrapolation scheme for efficient treatment of induced dipoles. *J. Chem. Phys.* **2016**, *145*, No. 164101.
- (105) Lipparini, F.; Lagardère, L.; Scalmani, G.; Stamm, B.; Cancès, E.; Maday, Y.; Piquemal, J. P.; Frisch, M. J.; Mennucci, B. Quantum calculations in solution for large to very large molecules: a new linear scaling QM/continuum approach. *J. Phys. Chem. Lett.* **2014**, *5*, 953–958.
- (106) Lipparini, F.; Scalmani, G.; Lagardère, L.; Stamm, B.; Cancès, E.; Maday, Y.; Piquemal, J. P.; Frisch, M. J.; Mennucci, B. Quantum, classical, and hybrid QM/MM calculations in solution: general implementation of the ddCOSMO linear scaling strategy. *J. Chem. Phys.* **2014**, *141*, No. 184108.
- (107) Stamm, B.; Lagardère, L.; Scalmani, G.; Gatto, P.; Cancès, E.; Piquemal, J. P.; Maday, Y.; Mennucci, B.; Lipparini, F. How to make continuum solvation incredibly fast in a few simple steps: a practical guide to the domain decomposition paradigm for the conductor-like screening model. *Int. J. Quantum Chem.* **2019**, *119*, No. e25669.
- (108) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: software tools for molecular design. *J. Chem. Theory Comput.* **2018**, *5273*.
- (109) Holst, M.; Saied, F. Multigrid solution of the Poisson-Boltzmann equation. *J. Comput. Chem.* **1993**, *14*, 105–113.
- (110) Holst, M. J.; Saied, F. Numerical solution of the nonlinear Poisson-Boltzmann equation: developing more robust and efficient methods. *J. Comput. Chem.* **1995**, *16*, 337–364.
- (111) Klamt, A.; Moya, C.; Palomar, J. A comprehensive comparison of the IEFPCM and SS(V)PE continuum solvation methods with the COSMO approach. *J. Chem. Theory Comput.* **2015**, *11*, 4220–4225.
- (112) Baldridge, K.; Klamt, A. First principles implementation of solvent effects without outlying charge error. *J. Chem. Phys.* **1997**, *106*, 6622–6633.
- (113) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669–681.

- (114) Lagardère, L.; Jolly, L. H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y. Y.; Ponder, J. W.; Piquemal, J. P. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, *9*, 956–972.
- (115) Bashford, D.; Case, D. A. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.
- (116) Grycuk, T. Deficiency of the Coulomb-field approximation in the Generalized Born model: An improved formula for Born radii evaluation. *J. Chem. Phys.* **2003**, *119*, 4817–4826.
- (117) Mongan, J.; Svrcek-Seiler, W. A.; Onufriev, A. Analysis of integral expressions for effective Born radii. *J. Chem. Phys.* **2007**, *127*, No. 185101.
- (118) Aguilar, B.; Shadrach, R.; Onufriev, A. V. Reducing the secondary structure bias in the generalized Born model via R6 effective radii. *J. Chem. Theory Comput.* **2010**, *6*, 3613–3630.
- (119) Onufriev, A.; Case, D. A.; Bashford, D. Effective Born radii in the Generalized Born approximation: The importance of being perfect. *J. Comput. Chem.* **2002**, *23*, 1297–1304.
- (120) Ponder, J.; Wu, C.; Ren, P.; Pande, V.; Chodera, J.; Schnieders, M.; Haque, I.; Mobley, D.; Lambrecht, D.; DiStasio, R.; Head-Gordon, M.; Clark, G.; Johnson, M.; Head-Gordon, T. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- (121) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–725.
- (122) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- (123) Foloppe, N.; MacKerell, A. D. Conformational properties of the deoxyribose and ribose moieties of nucleic acids: a quantum mechanical study. *J. Phys. Chem. B* **1998**, *102*, 6669–6678.
- (124) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D. Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme. *Biophys. J.* **2006**, *90*, L36–L38.
- (125) Nina, M.; Im, W.; Roux, B. Optimized atomic radii for protein continuum electrostatics solvation forces. *Biophys. Chem.* **1999**, *78*, 89–96.
- (126) Swanson, J. M. J.; Adcock, S. A.; McCammon, J. A. Optimized radii for Poisson-Boltzmann calculations with the AMBER force field. *J. Chem. Theory Comput.* **2005**, *1*, 484–493.
- (127) Swanson, J. M. J.; Wagoner, J. A.; Baker, N. A.; McCammon, J. A. Optimizing the Poisson dielectric boundary with explicit solvent forces and energies: lessons learned with atom-centered dielectric functions. *J. Chem. Theory Comput.* **2007**, *3*, 170–183.
- (128) Green, D. F. Optimized parameters for continuum solvation calculations with carbohydrates. *J. Phys. Chem. B* **2008**, *112*, 5238–5249.
- (129) Yamagishi, J.; Okimoto, N.; Morimoto, G.; Taiji, M. A New Set of Atomic Radii for Accurate Estimation of Solvation Free Energy by Poisson-Boltzmann Solvent Model. *J. Comput. Chem.* **2014**, *35*, 2132–2139.
- (130) Bernales, V. S.; Marenich, A. V.; Contreras, R.; Cramer, C. J.; Truhlar, D. G. Quantum mechanical continuum solvation models for ionic liquids. *J. Phys. Chem. B* **2012**, *116*, 9122–9129.
- (131) Liu, C. W.; Piquemal, J. P.; Ren, P. Y. AMOEBA plus classical potential for modeling molecular interactions. *J. Chem. Theory Comput.* **2019**, *15*, 4122–4139.
- (132) Rackers, J. A.; Ponder, J. W. Classical Pauli repulsion: an anisotropic, atomic multipole model. *J. Chem. Phys.* **2019**, *150*, No. 084104.
- (133) Liu, C.; Piquemal, J.-P.; Ren, P. Implementation of geometry-dependent charge flux into the polarizable AMOEBA+ potential. *J. Phys. Chem. Lett.* **2020**, *11*, 419–426.
- (134) Swanson, J. M. J.; Mongan, J.; McCammon, J. A. Limitations of atom-centered dielectric functions in implicit solvent models. *J. Phys. Chem. B* **2005**, *109*, 14769–14772.
- (135) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born model with a simple, robust molecular volume correction. *J. Chem. Theory Comput.* **2007**, *3*, 156–169.
- (136) Onufriev, A.; Bashford, D.; Case, D. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383–394.
- (137) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved generalized Born solvent model parameters for protein simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2020–2034.