Implicit Solvents for the Polarizable Atomic Multipole AMOEBA Force Field

Rae A. Corrigan^{1,*}, Guowei Qi^{2,*}, Andrew C. Thiel¹, Jack R. Lynn¹, Brandon D. Walker⁵, Thomas L. Casavant¹, Louis Lagardere³, Jean-Philip Piquemal³, Jay W. Ponder⁴, Pengyu Ren⁵, and Michael J. Schnieders^{1,2,#}

¹Roy J Carver Department of Biomedical Engineering, University of Iowa, Iowa City, IA

²Department of Biochemistry, University of Iowa, Iowa City, IA

³Department of Chemistry, Sorbonne Université, Paris, France

⁴Department of Chemistry, Washington University in St. Louis, St. Louis, MO

⁵Department of Biomedical Engineering, University of Texas in Austin, Austin, TX

*The authors contributed equally.

Corresponding Author
michael-schnieders@uiowa.edu

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 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 electrostatic 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 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 0.51 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 will open the door to their use for folding and design applications.

Introduction

Solvation plays a key role in accurately portraying the natural processes of molecules *in vitro* and *in vivo*.¹ Hydrophilic and hydrophobic interactions govern protein folding² and impact molecular recognition³. For these reasons, solvent⁴ must be considered during computational protein design⁵⁻⁷, RNA folding^{8, 9}, and biocatalyst design¹⁰. 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.^{11, 12}

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 function of atomic coordinates $\mathbf{X}$ can be formulated as a sum of cavitation, dispersion, and electrostatic contributions

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

Equation 1.

where ΔW_{cav} is the unfavorable cost of forming a solute shaped cavity within solvent, ΔW_{disp} is the favorable contribution of including solute-solvent van der Waals interactions, and ΔW_{elec} captures the difference between charging the molecule in solvent and in vacuum environments (Figure 1)^{11, 13, 14}. Collectively, cavitation and dispersion are termed the nonpolar contribution¹⁵⁻¹⁹ to solvation free energy, while the electrostatics term is referred to as the polar contribution^{4, 20-25}. For the latter, previous widely used implicit solvents for biomolecules include approaches based on numerical solutions to the Poisson-Boltzmann Equation (PBE)^{24, 26-29} and the analytic generalized Born approximation³⁰⁻³⁸. 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 are parameterized for use with the polarizable

atomic multipole AMOEBA force field^{39, 40}. 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^{41, 42} or the much faster analytic generalized Kirkwood (GK) approximation⁴³. 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.

Figure 1. The 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 decharged in vacuum $-U_{\text{elec}}^{\text{vacuum}}(\mathbf{X})$. Step 2: dispersion interactions are removed between the solute and 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 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 $\Delta W_{\text{elec}}(\mathbf{X}) = U_{\text{elec}}^{\text{water}}(\mathbf{X}) - U_{\text{elec}}^{\text{vacuum}}(\mathbf{X})$.

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 interactions⁴⁴. The PFF defines solute electrostatics using permanent atomic multipoles (through dipole order) and induced dipoles, while the PBE was solved using a finite element mesh⁴⁵. A more recent example combined a Drude oscillator force field^{46, 47} with

numerical solutions of the PBE⁴⁸. Application of this model to pK_a prediction⁴⁹ showed superior accuracy relative to the additive CHARMM36 force field^{50, 51}, although at increased computational cost. A second recent example combined the Bond Capacity (BC) polarization model with both the generalized Born model (GB) and conductor-like polarizable continuum model (C-PCM)⁵². For the BC-GB model, NVE molecular dynamics was shown to conserve energy at a modest increase in cost of only 15% relative to vacuum⁵³. 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 solvents⁵⁴⁻⁵⁶, including the Polarizable Continuum Model (PCM)^{57, 58}, the Solvent Model (SM) series⁵⁹⁻⁶¹, and Conductor-like Screening Models (COSMO)^{62, 63}.

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)⁴¹, the domain decomposition COSMO (ddCOSMO) approach⁴², and the analytic generalized Kirkwood (GK) model⁴³. 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 calculations (QM) were performed using Gaussian 09. All molecular mechanics (MM) force field-based calculations needed for parameterization were performed using the Tinker 8 Software⁶⁴. Valence parameters were taken from the small molecule parameter database in PolType2. Atomic multipole moments were initially assigned from QM electron density calculated at the MP2/6-311G** level via Stone's distributed multipole analysis⁶⁵. Further optimization of permanent multipoles was performed using Tinker's Potential program to fit the electrostatic potential around each molecule to a QM electron density at the MP2/6-311++G(2d,2p) level.

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^{66,67}, unless otherwise indicated^{68,69}. Experimental hydration free energy differences for charged compounds (ΔG_{solv}^{ion}) were calculated using equation 2:

$$\Delta G_{solv}^{ion} = \Delta G_{solv}^{neutral} \pm \Delta G_{gas}^{basicity} \pm \Delta G_{solv}^{H+} \mp 2.303RT * pK_a$$

Equation 2

where the upper signs are used for cations and the lower signs for anions. $\Delta G_{solv}^{neutral}$ is the solvation free energy difference of the neutral molecule, $\Delta G_{gas}^{basicity}$ is the gas phase basicity from NIST^{70,71}, ΔG_{solv}^{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 pK_a is the negative decimal logarithm of the acid

dissociation constant from Stewart⁷². 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 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/mol⁷³) and the experimental free energy of transfer $K^+ \rightarrow H^+$ (-179.90 kcal/mol⁷⁴). By 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 Supplementary Figure S1), 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 simulations⁷⁵. 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^{73, 76}.

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 above⁷⁷, then energy minimized in vacuum. We see close agreement between the dipole moments from single point MP2/6-311G**(2d,2p) calculations and those from the AMOEBA small molecule parameters, shown in Figure 2.

Figure 2. The total dipole moment 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: $Y = 0.9986 \cdot X + 0.0011$; $R^2 = 0.9999$).

Cavitation Free Energy

The Lum-Chandler-Weeks theory of hydrophobicity predicts contrasting behavior for the cavitation free energy of small and large solutes.^{1, 78-80} At all length scales, the driving force for phase separation is proportional to solute volume, while the cost to form an interface is proportional to surface area. These competing factors manifest 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.

$$\Delta W_{\text{cav}}(r) \propto \begin{cases} \text{Volume,} & r \leq \sim 10 \text{ Å} \\ \text{Surface Area,} & r > \sim 10 \text{ Å} \end{cases}$$

Equation 3.

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 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 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 volume nor 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 effective surface tension based on local curvature are promising.⁸¹ 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 non-spherical solute conformation $\mathbf{X}$ can be determined from calculation of either solvent excluded volume $\text{SEV}(\mathbf{X})$

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

Equation 4.

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

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

Equation 5.

using algorithms developed by Connolly⁸²⁻⁸⁴ and implemented in Tinker⁶⁴ and Force Field X⁸⁵. 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 molecule⁸⁶. The cavitation free energy of an (approximately) spherical solute can then be described by a piecewise continuous function of its effective radius

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

Equation 6.

where in the volume scaling regime, cavitation free energy is defined by the product of SEV with solvent pressure (SP) denoted by λ (kcal/mole/Å³); in the surface area scaling regime, cavitation free energy is defined by the product of SASA with surface tension (ST) denoted by γ (kcal/mole/Å²). For our model, SP was assessed using two explicit solvent simulation approaches that are in general agreement. The first approach assumes the SEV and SASA cavitation free energies are equal at the cross-over point χ for a spherical solute, which yields the relationship $\lambda = 3 \cdot \gamma / \chi$. This defines SP to be 0.031 kcal/mole/Å³, using the experimental surface tension of water (0.103 kcal/mole/Å²) and an approximate cross-over point of 10 Å from fixed charge simulations¹. The second approach leverages explicit solvent free energy perturbation simulations⁸⁷ using the AMOEBA water model⁸⁸ and 39 AMOEBA small molecules as described elsewhere³⁹, which resulted in a mean SP of 0.0334 kcal/mole/Å³. 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/mole/Å³ 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/mole/Å³ and cross-over radius of 9.251 Å were chosen due to their consistency with the AMOEBA model.

The simple definition in Equation 6 for the transition between the volume scaling and surface area scaling regimes is not useful for molecular dynamics 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 $w = 7 \text{ \AA}$ centered on the cross-over point χ , such that the switch begins at $b = \chi - w/2$ and ends at $e = \chi + w/2$ to give

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

Equation 7.

The volume scaling switch $s_v(r)$ is a 5th order polynomial whose 6 coefficients are uniquely determined by constraining its value at b to $s_v(b) = 1$ and its value at e to $s_v(e) = 0$, as well as constraining first and second derivatives at b and e to be zero. This gives

$$s_v(r) = c_0 + c_1 r + c_2 r^2 + c_3 r^3 + c_4 r^4 + c_5 r^5$$

Equation 8.

where

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

Equation 9.

The surface area switch in this symmetric case is

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

Equation 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 $o = 0.2 \text{ \AA}$. This gives the final functional form used here

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

Equation 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 effective radius is shown Figure 3.

Figure 3. Surface tension is not constant for small solutes, but increases approximately linearly until the effective radius of the solute grows to beyond ~ 10 Å. For large (flat) solutes, surface tension asymptotes toward the experimental value for a water-vapor interface of $0.103 \text{ kcal/mol/Å}^2$.⁸⁹ 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, non-differentiable form ($\Delta W_{\text{sphere}}(r)$) given by Equation 6, black dashes) or the smooth form used in this work ($\Delta W_{\text{cav}}(r)$ given by Equation 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 $0.08 \text{ kcal/mol/Å}^2$, dashed 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).

Dispersion Free Energy

The pairwise dispersion energy for the AMOEBA model is given by a buffered-14-7 potential⁹⁰

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

Equation 12.

where r_{ij} is the separation distance between atoms i and j , ε_{ij} is the well depth, and $r_{0,ij}$ is the minimum energy separation distance³⁹. This can be used to define a purely repulsive Weeks-Chandler-Andersen (WCA) potential^{91, 92} as

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

Equation 13.

which is shown in Figure 4.

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

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 solute-solvent enthalpy for a series of small alkanes studied using free energy perturbation (FEP)⁹³

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

Equation 14.

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

$$\Delta W_{GKL} = \sum_{i=1}^n \frac{-16\pi\rho_w\varepsilon_{iw}\sigma_{iw}^6}{3R_i^3}$$

Equation 15.

where ρ_w is the number density of water (0.033428 per $\text{\AA}^3$), ε_{iw} and σ_{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.^{35, 68} 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, use of the Born radii in Equation 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 Equation 15 and instead integrating the true WCA attractive potential (Figure 4) outside of the solute cavity for each atom.

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

Equation 16.

We present an analytic approach based on the HCT pairwise integration method also used for GK.^{32, 94} Due to 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

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

Equation 17.

where the solvent indicator function S is unity if the point (r, θ, ϕ) is located within the solvent, but zero otherwise, ρ_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 ($d = 1.056 \text{ \AA}$), 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 Equation 18 for $\Delta W_{\text{disp}}(\mathbf{X})$, were fit against dispersion enthalpies (Equation 14) measured from explicit solvent simulations as described in the Supplementary Materials (Table S1). For a single water oxygen atom, the behavior of ΔW_{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 Figures 5B and 5C.

After performing the two angular integrals in Equation 16, inverting the integration domain and applying the HCT pairwise approximation³² gives

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

Equation 18.

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 (Equation 12 in Hawkins *et al.*³²)

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

Equation 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)

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

Equation 20.

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

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

Equation 21.

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

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

Equation 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 model⁸⁸. The final piece to this model is the solution to the integral in Equation 18, which uses integration bounds shown in Figure 6.

Figure 6. Illustration of the integration limits for the dispersion free energy based on solute van der Waals parameters and separation distance. **Panel 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 $r_{ij} + \rho_i$, whichever is smaller. If $R_0 > r_{0,ij}$, then only variable portion of WCA potential factors into dispersion free energy. **Panel 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. **Panel 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). **Panel 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 $r_{ij} + \rho_j$.

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

then a contribution of

$$I_\varepsilon(U, L, \varepsilon) = \int_L^U \varepsilon \times H(r, r_{ij}, \rho_j) r^2 dr = \varepsilon \left[r^2 (3r^2 - 8r_{ij}r + 6r_{ij}^2 - 6\rho_j^2) / 48r_{ij} \right]_L^U$$

Equation 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

$$\begin{aligned} I_{14}(L, U, \varepsilon, r_{0,ij}) &= \int_L^U \frac{\varepsilon r_{0,ij}^{14}}{r^{12}} H(r, r_{ij}, \rho_j) dr \\ &= \varepsilon r_{0,ij}^{14} \left[(-120r_{ij}r + 66r^2 + 55r_{ij}^2 - 55\rho_j^2) / 2640r_{ij}r^{12} \right]_L^U \end{aligned}$$

Equation 24.

and the attractive contribution by

$$I_7(L, U, \varepsilon, r_{0,ij}) = -2 \int_L^U \frac{\varepsilon r_{0,ij}^7}{r^5} H(r, r_{ij}, \rho_j) dr = -2\varepsilon r_{0,ij}^7 \left[\left(-15r_{ij}r + 10r^2 + 6r_{ij}^2 - 6\rho_j^2 \right) / 120r_{ij}r^5 \right]_L^U$$

Equation 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 Equation 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 linearized Poisson-Boltzmann Equation (PBE)

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

Equation 26.

for all $\mathbf{r}$ in a domain Ω , 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 Equation 26 is solved repeatedly during iterations of an SCRF solver (*e.g.* Jacobi Over-Relaxation (SOR)⁸⁸, Conjugate Gradient (CG) methods^{95,96}, the Jacobi algorithm coupled to Direct Inversion in the Iterative Subspace (JI/DIIS)⁹⁵ and an optimized perturbation theory (OPT) method⁹⁷). This work compares three distinct continuum electrostatics models: the numerical solutions to the linearized PBE using the Adaptive Poisson-Boltzmann Solver (APBS)²⁹, a domain decomposition solution of the Conductor-like Screening Model (ddCOSMO)^{42, 98-100}, and the analytic Generalized Kirkwood (GK) theory⁴³.

Adaptive Poisson-Boltzmann Solver

APBS determines the solution to the PBE using parallelized finite difference multi-grid and finite element algebraic multi-grid 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 multi-level 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 multi-grid spacings. The combination of Tinker¹⁰¹ with APBS to support the AMOEBA force field has been described previously⁴¹. APBS was run in Tinker using a grid spacing of 129³ and a probe of radius 0.0 Å to define a van der Waals solute cavity. The APBS parallel multigrid solver (PMG)^{24, 102, 103} was used for all calculations as it is currently the only APBS solver available for use with Amoebea via Tinker.

Domain Decomposition Conductor-like Screening Model

The ddCOSMO electrostatics model^{42, 98-100} treats the solvent as an infinite conductor surrounding a solute-shaped cavity Ω , determined by a union of spheres (one sphere per atom of solute)

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

Equation 27.

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

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

Equation 28.

where $f(\epsilon)$ is a scaling factor used to adjust for the non-conductor nature of the solvent based on its dielectric constant ϵ . W is obtained from the solution of the following boundary value problem

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

Equation 29.

where Φ is the solute's electrostatic potential in vacuum and Γ 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 the ddCOSMO implementation to be parallelized and is available in the Tinker-HP package¹⁰⁴, 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 energy¹⁰⁵ (equivalent to the GK monopole term $G_{\text{GK}}^{(0)}$) is given by

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

Equation 30.

where ϵ_s is the permittivity of the solvent, ϵ_h is the permittivity of a homogeneous reference state, q_i and q_j are partial charges, and the empirical generalizing function f_{GB} is given by

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

Equation 31.

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 solvent¹⁰⁶⁻¹⁰⁸

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

Equation 32.

and

$$f_{ij} = e^{-r_{ij}^2 / c_{GK} a_i a_j}$$

Equation 33.

where c_{GK} is a tuning parameter that typically ranges from 2 to 4. GK extends GB methods to polarizable atomic multipole charge distributions by using Kirkwood’s analytic solution to the electrostatic component of solvation free energy for an arbitrary (*i.e.* multipolar) charge distribution¹⁰⁹. For example, the interaction between two permanent dipoles is expressed as

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

Equation 34.

Where $\mathbf{u}_i$ and $\mathbf{u}_j$ are permanent dipole vectors and the subscripts α and β indicate use of the Einstein summation convention. GK interaction tensors up to quadrupole-quadrupole order, as well as their inclusion in the AMOEBA SCRF calculation, have been described previously⁴³.

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 ($c_{GK} = 2.455$) and a unitless scale factor ($c_{HCT} = 0.72$) that avoids overestimation of descreening due to atomic overlaps when computing Born radii¹¹⁰. 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. The scale factor was then reduced to 0.72. PB self and cross-term energies were calculated using Tinker with an APBS grid spacing of 129³ and a

van der Waals definition of solute cavity using Amoeba 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 energies. The final HCT scale factor of 0.720 gave an MSE of -0.10, MUE of 0.29, and RMSE of 0.87 (Figure 7, Panel A). The slightly negative MSE is compensated for during fitting of base solute radii described below. Testing with generalizing constants (c_{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 work⁴³ and gave an MSE of 0.07, MUE of 0.35, and RMSE of 1.17 (Figure 7, Panel B) and led to the fewest outliers during electrostatic radii fitting. In both cases, GK energies are strongly correlated with PB energies, with R-squared values of 0.996 and 0.994, respectively.

Figure 7. Panel A: Fit of GK self energies to *perfect* PB self energies ($Y = 1.040X + 0.048$, $R^2 = 0.996$). **Panel 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 cross term energies are reported for 1424 atoms.

Target Function for Electrostatic Radii Optimization

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

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

Equation 35.

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 nonpolar parameters determined for GK were transferred to APBS and ddCOSMO without being re-optimized. The optimization was seeded with electrostatic radii based on AMOEBA van der Waals R_{min} values, and after trial and error optimization weights in all cases were set to 1.0.

Results

Small Molecule Hydration Free Energy

The nonpolar portion of the model consists of an unfavorable cavitation free 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 a scale factor – were used in calculating dispersion. A solvent pressure of 0.0334 kcal/mol/ Å³ for cavitation was chosen based on previous testing. A dispersion offset ($d = 1.056$ Å) and unitless HCT dispersion scale factor ($s = 0.75$) were chosen to match dispersion values from solute-solvent enthalpy simulations in explicit water (Supplementary Table S1).

For the electrostatic portion of the model, a total of 41 solute radii classes were optimized using L-BFGS minimization as described above, starting with electrostatic radii based on AMOEBA van der Waals R_{min} values. Solvation free energy difference values from the FreeSolv database^{66, 67} were used as benchmarks. Each radii class was determined based on SMARTS strings, automatically generated for each atom by PolType2. A total of 78 unique SMARTS strings were generated for the test set of 103 molecules. These 78 SMARTS strings were then collected into 41 groups based on element (*i.e.* hydrogen SMARTS strings were only grouped with other hydrogen SMARTS strings), chemical environment, and electrostatic radii sizes from an initial optimization using all SMARTS strings under GK electrostatics (Supplementary Table S2) – a 52% reduction. The use of fewer parameters helps to avoid overfitting and improve generalizability. Optimization was performed using the 41 radii classes to individually parameterize the APBS, ddCOSMO, and GK electrostatic models¹¹¹. Fit radii deviated from original van der Waals radii by an average of 9.3% for APBS, 9.9% for ddCOSMO, and 14.0% for GK. The quality of the resulting implicit solvent model for small molecules using APBS, ddCOSMO, and GK is shown below in Figures 8, 9, and 10, respectively, and full data is available in Supplementary Tables S3 – S6.

Figure 8. Shown is a comparison of experimental and APBS solvation free energy differences using either AMOEBA van der Waals $R_{\min}$ radii to describe the solute-solvent boundary (Panel A) or using fit radii (Panel B). When using $R_{\min}$ 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 $R^2 = 0.9979$.

Figure 9. Shown is a comparison of experimental and ddCOSMO solvation free energy differences using either AMOEBA van der Waals $R_{\min}$ radii to describe the solute-solvent boundary (Panel A) or using fit radii (Panel B). When using $R_{\min}$ 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. Shown is a comparison of experimental and GK solvation free energy differences using either AMOEBA van der Waals $R_{\min}$ radii to describe the solute-solvent boundary (Panel A) or using fit radii (Panel B). When using $R_{\min}$ 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 $Y = 0.9998 \cdot X - 0.0068$ with $R^2 = 0.9989$.

Implicit solvation energy calculations performed using APBS, ddCOSMO, and GK with parameterized electrostatics radii gave mean unsigned errors (MUE) of 0.70, 0.63, and 0.51 kcal/mol from experimental values, respectively, as shown in Supplementary Table S6. Tables 2A and 2B show the mean signed error (MSE) and root mean square error (RMSE) between experimental and computed solvation free energies. Fit solute radii resulted in RMSE/MSE values of 1.00/0.05, 0.92/0.00, and 0.71/0.00 kcal/mol for the APBS, ddCOSMO, and GK models, respectively (Table 1B).

Table 1A. RMSE/MSE 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

Table 1B. RMSE/MSE for test molecules using fit solute 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	0.45/-0.15
Alcohols and Phenols	16	1.01/+0.50	1.31/+0.65	0.87/+0.47
Amines	12	1.49/+0.92	1.37/+0.06	1.12/-0.15
Amides	8	0.66/+0.13	0.66/+0.13	0.88/-0.05
Nitrogen Heterocyclic	8	1.31/+0.56	0.50/+0.07	0.65/+0.10
Arenes	5	0.50/-0.44	1.00/-0.61	0.52/+0.07
Ethers	5	0.23/-0.03	0.35/-0.11	0.68/-0.48
Oxanes and Oxines	4	1.16/-0.33	0.58/-0.36	0.95/+0.02
Thiols	4	0.65/-0.49	0.38/-0.26	0.42/-0.08
Carboxylic Acids	3	0.09/-0.08	0.11/-0.01	0.19/-0.01
Sulfides	3	0.36/-0.33	0.19/+0.09	0.13/-0.12
Aldehydes	2	0.48/-0.38	0.39/-0.39	0.18/+0.09
Other	3	1.96/-0.06	0.88/-0.56	0.92/-0.53
Total Neutrals	91	0.97/+0.14	0.87/+0.01	0.75/-0.01
Charged	12	1.18/-0.59	1.25/-0.07	0.27/+0.02
Total	103	1.00/+0.05	0.92/+0.00	0.71/+0.00

Comparison to Explicit Solvent Free Energy Differences

Although the implicit solvents described here were fit to experimental data, direct comparison to AMOEBA explicit solvent hydration free energy differences helps illuminate if the continuum models are either overfit or exhibit relatively higher errors. A subset of 26 neutral small molecules used to parameterize the implicit solvent model are compared to available data from a recent AMOEBA explicit solvent study¹¹² in Table 2. Explicit solvent gave an RMSE of 0.70 kcal/mol compared to 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 differences support the conclusion that the continuum models are neither clearly overfit nor of worse quality than what is observed for AMOEBA solutes in explicit solvent.

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	ΔG_{expt}	Explicit		$\Delta G_{\text{implicit}}$			Signed Error		
		ΔG	Error	APBS	COSMO	GK	APBS	COSMO	GK
isopropanol	-4.74	-4.21	0.28	-3.77	-3.66	-4.38	0.97	1.08	0.36
hydrogen sulfide	-0.70	-0.41	0.08	-1.06	-0.77	-0.77	-0.36	-0.07	-0.07
p-cresol	-6.13	-5.6	0.28	-6.22	-5.37	-5.52	-0.09	0.76	0.61
dimethylsulfide	-1.61	-1.85	0.06	-2.11	-1.56	-1.77	-0.50	0.05	-0.16
phenol	-6.60	-5.05	2.40	-6.23	-5.48	-6.11	0.37	1.12	0.49
benzene	-0.90	-1.23	0.11	-1.30	-1.70	-1.13	-0.40	-0.80	-0.23
ethanol	-5.00	-4.69	0.10	-4.09	-3.89	-3.48	0.91	1.11	1.52
ethane	1.83	1.73	0.01	2.31	2.10	2.42	0.48	0.27	0.59
n-butane	2.10	1.11	0.98	2.09	2.21	1.68	-0.01	0.11	-0.42
methylamine	-4.55	-5.46	0.83	-4.00	-4.53	-4.21	0.55	0.02	0.34
dimethylamine	-4.29	-3.04	1.56	-2.72	-4.03	-4.57	1.57	0.26	-0.28
trimethylamine	-3.20	-2.09	1.23	-3.34	-3.09	-1.54	-0.14	0.11	1.66
propane	2.00	1.69	0.10	2.24	2.28	2.36	0.24	0.28	0.36
methane	2.00	1.73	0.07	2.43	2.39	1.92	0.43	0.39	-0.08
methanol	-5.10	-4.79	0.10	-3.55	-3.51	-3.84	1.55	1.59	1.26
n-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	-0.58	-0.51	-0.70	0.32
ethylbenzene	-0.79	-0.8	0.00	-1.21	-1.41	-0.35	-0.42	-0.62	0.44
n-methylacetamide	-10.00	-8.66	1.80	-9.50	-9.09	-9.19	0.50	0.91	0.81
water	-6.30	-5.86	0.19	-5.36	-6.34	-6.28	0.94	-0.04	0.02
acetic acid	-6.69	-5.63	1.12	-6.56	-6.75	-6.53	0.13	-0.06	0.16
methylethylsulfide	-1.50	-1.98	0.23	-1.64	-1.19	-1.62	-0.14	0.31	-0.12
imidazole	-9.63	-10.25	0.38	-9.43	-9.79	-9.68	0.20	-0.16	-0.05
acetamide	-9.70	-9.3	0.16	-10.78	-10.22	-9.91	-1.08	-0.52	-0.21
ethylamine	-4.50	-4.33	0.03	-4.14	-4.97	-3.88	0.36	-0.47	0.62
pyrrolidine	-5.48	-4.88	0.36	-2.31	-4.61	-4.93	3.17	0.87	0.55
MSE			0.19				0.36	0.24	0.35
MUE			0.57				0.64	0.50	0.48
RMSE			0.70				0.91	0.65	0.63

Validation Simulations on Proteins, DNA and RNA

Nine nucleic acids (≤ 24 nucleotides) and nine proteins (≤ 129 residues) of modest size were used to test the polarization response of the implicit solvent electrostatics. 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 relaxation of the water and ions.

Figure 11. Validation molecule set consisting of 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 hair pins (1ZIH, 2KOC, and 1SZY), and two DNA double helices (1D20 and 2HKB). Additional information on individual molecules and simulation conditions can be found in Supplementary Table S7.

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 nsec MD simulations with the biomolecule fixed (*i.e.* solvent degrees of freedom were converged). For detailed simulation conditions, see Supplementary Table S7. In addition, dipole moment magnitudes were calculated

for the validation set using three fixed charge force fields – AMBER ff99SB¹¹³, OPLS-AA/L¹¹⁴, and CHARMM22/CMAP^{115, 116}. Tinker version 8.8.1 (August 2020) does not include nucleic acid force field parameters for OPLS-AA/L or CHARMM22/CMAP, so 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 12. 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-squared values of 0.987, 0.979, and 0.982, respectively, when compared to AMOEBA vacuum dipole moment magnitudes, and R-squared 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 worse agreement with AMOEBA in vacuum, is consistent with fixed charge biomolecular force fields being *pre-polarized* for aqueous simulations. This also demonstrates that the AMOEBA electrostatics model produces molecular dipole moments that are consistent with previous generation force fields.

Figure 12. Comparison of dipole moment magnitudes for fixed charge force fields and AMOEBA across environments. **Panel 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). **Panel 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). Dotted lines at $x = y$ are to guide the eye.

AMOEBA implicit solvent dipole moment magnitudes were calculated using each of the three electrostatics models and compared to those from explicit solvent ensemble averages as shown in Figure 13. All three electrostatics models showed near perfect correlation with explicit solvent values based on R-squared values for the APBS, ddCOSMO, and GK models of 0.999 in each case. Notably, each AMOEBA implicit solvent electrostatic 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 model, 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 explored more in the future. Overall, these results suggest the electrostatics models scale up well from initial optimization against small molecule hydration free energy differences to applications on larger biomolecules.

Figure 13 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 dotted line at $x = y$ is to guide the eye.

The performance 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 (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 single GPU (NVIDIA® GeForce® RTX 2080 Ti) were also collected. The latter performance (less than 0.005 seconds per time step) is consistent with molecular dynamics performance of ~20 nsec per day using a conservative 1 fsec 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.

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 (seconds)			
	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

Conclusions

Implicit solvent models were developed for use with the polarizable AMOEBA force field. Novel cavitation and dispersion non-polar 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 non-polar terms, the solute-solvent electrostatic boundary (*i.e.* atomic radii) was optimized for three continuum electrostatics models – ABPS, ddCOSMO, and GK – using numerical optimization against experimental solvation free energy differences. Overall, the APBS, ddCOSMO, and GK models produced mean unsigned errors of 0.70, 0.63, and 0.51 kcal/mol compared to experiment. All three implicit solvent models produced hydration free energy difference RMSEs within 0.2 kcal/mol of AMOEBA explicit solvent 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 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 significantly 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 will often fall within the surface area scaling regime. A future goal is to account for local molecular curvature to promote transferability of 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 non-spherical solutes, which adds physical detail to

the previously described Born radii-based dispersion model⁶⁸. 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 agreement of GK self-energies with 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. Precedent for this split is given in the original description of the HCT pairwise descreening approximation, where the scale factor magnitude generally decreases with increasing atomic size³². Additionally, agreement between GK cross term energies and calculated *perfect* PB multipolar cross term energies might be improved by 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 longer range for monopoles than for dipoles. Therefore, the transition between the Born ion regime (or Kirkwood multipole regime) and the Coulomb regime, which is tuned by the constant in the generalizing function (Equation 32), 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 boundary¹¹⁷⁻¹²¹, 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 0.7, MSE 0.0) is comparable to the recent Drude/PB implicit solvent (RMSE 0.8, MSE 0.0)⁴⁸. The fit radii reproduce experimental solvation free energy differences better than original vdW 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^{56, 122}.

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 13). 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 dynamics simulations of biomolecules in implicit solvent compared to explicit solvent to assess 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, they each support polarizable atomic multipole electrostatics and will be adapted to emerging models such as AMOEBA+ and HIPPO¹²³⁻¹²⁵.

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^{126, 127}. The approximation of a van der Waals description is modest for small solutes but becomes problematic as molecular size and complexity increases (*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 continuum water access to spaces not accessible to explicit water. Favorable hydration effects of continuum water in interstitial spaces promotes swelling of biomolecules and opposes hydrophobic compaction forces. For this reason, future work will incorporate methods

that have been proposed to account for interstitial spaces^{108, 127-129} into the Amoeba family of implicit solvents.

Acknowledgements

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. MJS was supported by NIH R01DK110023, NIH R01DC012049, and NSF CHE-1751688. JWP and PR were supported by R01GM106137 and R01GM114237. JPP and LL were supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program No 810367. RAC and ACT were supported by the NSF Graduate Research Fellowship Program NSF DGE-1945994 and GQ 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 (7059), 640-647.
2. Dill, K. A.; MacCallum, J. L., The Protein-Folding Problem, 50 Years On. *Science* **2012**, 338 (6110), 1042-1046.
3. Zhang, Q. C.; Petrey, D.; Deng, L.; Qiang, L.; Shi, Y.; Thu, C. A.; Bisikirska, 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 (7421), 556-60.
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 (4), 427-91.
5. Huang, P.-S.; Boyken, S. E.; Baker, D., The coming of age of de novo protein design. *Nature* **2016**, 537, 320.
6. 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 (7), 8.
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 (30), 2494-2507.
8. Cheatham, T. E.; Case, D. A., Twenty-five years of nucleic acid simulations. *Biopolymers* **2013**, 99 (12), 969-977.
9. 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 (9), 1578-1590.
10. Musil, M.; Konegger, H.; Hong, J.; Bednar, D.; Damborsky, J., Computational Design of Stable and Soluble Biocatalysts. *Acs Catalysis* **2019**, 9 (2), 1033-1054.
11. Roux, B.; Simonson, T., Implicit solvent models. *Biophys. Chem.* **1999**, 78 (1-2), 1-20.
12. Onufriev, A. V.; Izadi, S., Water models for biomolecular simulations. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2018**, 8 (2), 40.
13. Tan, C.; Tan, Y. H.; Luo, R., Implicit nonpolar solvent models. *J. Phys. Chem. B* **2007**, 111 (42), 12263-12274.

14. Decherchi, S.; Masetti, M.; Vyalov, I.; Rocchia, W., Implicit solvent methods for free energy estimation. *Eur. J. Med. Chem.* **2015**, *91*, 27-42.
15. 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* (31), 9523-9530.
16. 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* (22), 8331-8336.
17. 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* (8), 9.
18. Harris, R. C.; Pettitt, B. M., Effects of geometry and chemistry on hydrophobic solvation. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (41), 14681-14686.
19. Michael, E.; Polydorides, S.; Simonson, T.; Archontis, G., Simple models for nonpolar solvation: parameterization and testing. *J. Comput. Chem.* **2017**, *38* (29), 2509-2519.
20. Gilson, M. K.; Honig, B. H., Calculation of electrostatic potentials in an enzyme active-site. *Nature* **1987**, *330* (6143), 84-86.
21. Gilson, M. K.; Honig, B., Calculation of the total electrostatic energy of a macromolecular system - solvation energies, binding-energies, and conformational-analysis. *Proteins-Structure Function and Genetics* **1988**, *4* (1), 7-18.
22. 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* (4), 1454-1455.
23. Honig, B.; Nicholls, A., Classical electrostatics in biology and chemistry. *Science* **1995**, *268* (5214), 1144-1149.
24. 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* (18), 10037-41.
25. Simonson, T., Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.* **2001**, *11* (2), 243-252.
26. 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* (12), 889-897.
27. Tan, Y. H.; Luo, R., Continuum treatment of electronic polarization effect. *J. Chem. Phys.* **2007**, *126* (9), 6.

28. 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* (25), 7675-7688.
29. 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* (1), 112-128.
30. Rashin, A. A.; Honig, B., Reevaluation of the Born model of ion hydration. *J. Phys. Chem.* **1985**, *89* (26), 5588-5593.
31. Roux, B.; Yu, H. A.; Karplus, M., Molecular-basis for the Born model of ion solvation. *J. Phys. Chem.* **1990**, *94* (11), 4683-4688.
32. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246* (1-2), 122-129.
33. 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* (16), 3005-14.
34. 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* (26), 5239-5248.
35. 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* (4), 479-499.
36. Gallicchio, E.; Paris, K.; Levy, R. M., The AGBNP2 implicit solvation model. *J. Chem. Theory Comput.* **2009**, *5* (9), 2544-2564.
37. 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* (4), 1788-1794.
38. Onufriev, A. V.; Case, D. A., Generalized Born implicit solvent models for biomolecules. In *Annual Review of Biophysics, Vol 48*, Dill, K. A., Ed. Annual Reviews: Palo Alto, 2019; Vol. 48, pp 275-296.
39. Ren, P.; Wu, C.; Ponder, J. W., Polarizable atomic multipole-based molecular mechanics for organic molecules. *J. Chem. Theory Comput.* **2011**, *7* (10), 3143-3161.
40. 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* (4), 2084-2108.

41. 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* (12), 124114.
42. 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* (2), 623-634.
43. Schnieders, M. J.; Ponder, J. W., Polarizable atomic multipole solutes in a generalized Kirkwood continuum. *J. Chem. Theory Comput.* **2007**, *3* (6), 2083-2097.
44. 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* (4), 694-715.
45. Cortis, C. M.; Friesner, R. A., Numerical solution of the Poisson-Boltzmann equation using tetrahedral finite-element meshes. *J. Comput. Chem.* **1997**, *18* (13), 1591-1608.
46. 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* (9), 4983-5013.
47. Lemkul, J. A.; MacKerell Jr., A. D., Polarizable force field for RNA based on the classical drude oscillator. *J. Comput. Chem.* **2018**, *39* (32), 2624-2646.
48. 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* (22), 1707-1719.
49. 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* (7), 4655-4668.
50. 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.; Kuczera, 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* (18), 3586-616.
51. 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* (9), 3257-3273.
52. Poier, P. P.; Jensen, F., Including implicit solvation in the bond capacity polarization model. *J. Chem. Phys.* **2019**, *151* (11), 6.

53. Poier, P. P.; Jensen, F., Polarizable charges in a generalized Born reaction potential. *The Journal of Chemical Physics* **2020**, *153* (2), 024111.
54. Cramer, C. J.; Truhlar, D. G., Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.* **1999**, *99* (8), 2161-2200.
55. Tomasi, J., Thirty years of continuum solvation chemistry: a review, and prospects for the near future. *Theor. Chem. Acc.* **2004**, *112* (4), 184-203.
56. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105* (8), 2999-3093.
57. Miertys, 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* (1), 117-129.
58. 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* (8), 3032-3041.
59. Cramer, C. J.; Truhlar, D. G., An SCF solvation model for the hydrophobic effect and absolute free energies of aqueous solvation. *Science* **1992**, *256* (5054), 213-217.
60. 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* (6), 1133-1152.
61. 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* (18), 6378-6396.
62. Klamt, A.; Schuurmann, G., COSMO - a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society: Perkin Transactions 2* **1993**, (5), 799-805.
63. Klamt, A., Conductor-like screening model for real solvents - a new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* **1995**, *99* (7), 2224-2235.
64. 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* (10), 5273-5289.
65. Stone, A. J., Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005**, *1* (6), 1128-1132.
66. 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* (7), 711-720.

67. 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* (5), 1559-1569.
68. 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* (5), 517-529.
69. 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* (1), 128-139.
70. Bartmess, J. E., "Negative ion energetics data". In *NIST Chemistry WebBook*, Technology, N. I. o. S. a., Ed. Gaithersburg, MD, p 20899.
71. Hunter, E. P.; Lias, S. G., "Proton affinity evaluation". In *NIST Chemistry WebBook*, National Institute of Standards and Technology: Gaithersburg, MD, p 20899.
72. Stewart, R., *The proton: applications to organic chemistry*. Academic Press, Inc.: New York, 1985; Vol. 46.
73. 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.
74. Fawcett, W. R., Thermodynamic parameters for the solvation of monatomic ions in water. *The Journal of Physical Chemistry B* **1999**, *103* (50), 11181-11185.
75. Ashbaugh, H. S., Convergence of molecular and macroscopic continuum descriptions of ion hydration. *The Journal of Physical Chemistry B* **2000**, *104* (31), 7235-7238.
76. Wu, J. C.; Chattree, G.; Ren, P. Y., Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012**, *131* (3), 11.
77. Wu, J. C.; Chattree, G.; Ren, P., Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012**, *131* (3), 1-11.
78. 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* (15), 8324-8327.
79. Huang, D. M.; Geissler, P. L.; Chandler, D., Scaling of hydrophobic solvation free energies. *J. Phys. Chem. B* **2001**, *105* (28), 6704-6709.
80. Huang, D. M.; Chandler, D., The hydrophobic effect and the influence of solute-solvent attractions. *J. Phys. Chem. B* **2002**, *106* (8), 2047-2053.

81. Chen, J. H.; Brooks, C. L., Critical importance of length-scale dependence in implicit modeling of hydrophobic interactions. *J. Am. Chem. Soc.* **2007**, *129* (9), 2444-+.
82. Connolly, M. L., Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16* (5), 548-558.
83. Connolly, M. L., Computation of molecular volume. *J. Am. Chem. Soc.* **1985**, *107* (5), 1118-1124.
84. Connolly, M. L., The molecular-surface package. *J. Mol. Graph.* **1993**, *11* (2), 139-143.
85. 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* (3), 602-612.
86. Richards, F. M., Areas, volumes, packing and protein structure. *Annual review of biophysics and bioengineering* **1977**, *6* (1), 151-176.
87. Schnieders, M. J. The Theory and Effect of Solvation on Biomolecules. Dissertation, Washington University in St. Louis, St. Louis, MO, 2007.
88. Ren, P.; Ponder, J. W., Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003**, *107* (24), 5933-5947.
89. Steam, I. A. f. t. P. o. W. In *IAPWS release on surface tension of heavy water substance*, Physical Chemistry of Aqueous Systems: Proceedings of the 12th International Conference on the Properties of Water and Steam, Orlando, FL, Orlando, FL, 1994; pp a107-a138.
90. 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* (20), 7827-7843.
91. Weeks, J. D.; Chandler, D.; Andersen, H. C., Role of repulsive forces in determining the equilibrium structure of simple liquids. *The Journal of Chemical Physics* **1971**, *54* (12), 5237-5247.
92. Chandler, D.; Weeks, J. D.; Andersen, H. C., Van der Waals picture of liquids, solids, and phase transformations. *Science* **1983**, *220* (4599), 787-794.
93. 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* (26), 6271-6285.
94. 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.* **1996**, *100* (51), 19824-19839.

95. 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* (4), 1638-1651.
96. 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* (1), 180-190.
97. Simmonett, A. C.; IV, F. C. P.; Ponder, J. W.; Brooks, B. R., An empirical extrapolation scheme for efficient treatment of induced dipoles. *The Journal of Chemical Physics* **2016**, *145* (16), 164101.
98. Lipparini, F.; Lagardere, 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* (6), 953-958.
99. Lipparini, F.; Scalmani, G.; Lagardere, 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* (18), 13.
100. Stamm, B.; Lagardere, 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* (1), 15.
101. Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardere, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W., Tinker 8: software tools for molecular design. *J. Chem. Theory Comput.* **2018**.
102. Holst, M.; Saied, F., Multigrid solution of the Poisson-Boltzmann equation. *J. Comput. Chem.* **1993**, *14* (1), 105-113.
103. Holst, M. J.; Saied, F., Numerical solution of the nonlinear Poisson-Boltzmann equation: developing more robust and efficient methods. *J. Comput. Chem.* **1995**, *16* (3), 337-364.
104. Lagardere, 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* (4), 956-972.
105. Bashford, D.; Case, D. A., Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129-152.

106. 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* (9), 4817-4826.
107. Mongan, J.; Svrcek-Seiler, W. A.; Onufriev, A., Analysis of integral expressions for effective Born radii. *J. Chem. Phys.* **2007**, *127* (18), 10.
108. 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* (12), 3613-3630.
109. Kirkwood, J. G., Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J. Chem. Phys.* **1934**, *2* (7), 351-361.
110. 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* (14), 1297-1304.
111. Schnieders, M. J.; Ponder, J. W., Implicit solvents for the AMOEBA force field based on Poisson-Boltzmann and Generalized Kirkwood electrostatics (in preparation).
112. 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* (8), 2549-2564.
113. 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* (3), 712-725.
114. 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* (28), 6474-6487.
115. 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* (34), 6669-6678.
116. 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* (4), L36-L38.
117. Nina, M.; Im, W.; Roux, B., Optimized atomic radii for protein continuum electrostatics solvation forces. *Biophys. Chem.* **1999**, *78* (1-2), 89-96.
118. 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* (3), 484-493.

119. 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* (1), 170-183.
120. Green, D. F., Optimized parameters for continuum solvation calculations with carbohydrates. *J. Phys. Chem. B* **2008**, *112* (16), 5238-5249.
121. 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* (29), 2132-2139.
122. 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* (30), 9122-9129.
123. Liu, C. W.; Piquemal, J. P.; Ren, P. Y., AMOEBA plus classical potential for modeling molecular interactions. *J. Chem. Theory Comput.* **2019**, *15* (7), 4122-4139.
124. Rackers, J. A.; Ponder, J. W., Classical Pauli repulsion: an anisotropic, atomic multipole model. *J. Chem. Phys.* **2019**, *150* (8), 22.
125. Liu, C.; Piquemal, J.-P.; Ren, P., Implementation of geometry-dependent charge flux into the polarizable AMOEBA+ potential. *The Journal of Physical Chemistry Letters* **2020**, *11* (2), 419-426.
126. 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* (31), 14769-14772.
127. 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* (1), 156-169.
128. Onufriev, A.; Bashford, D.; Case, D., Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55* (2), 383-394.
129. Nguyen, H.; Roe, D. R.; Simmerling, C., Improved generalized Born solvent model parameters for protein simulations. *J. Chem. Theory Comput.* **2013**, *9* (4), 2020-2034.

Appendix

Supplementary Figure S1. Description of the phase potential of a cavity under three different simulation conditions. Neither molecular mechanics force fields nor continuum solvents should implicitly include the effect of the interface potential $\Phi_{\text{Interface}}$ when fitting against experimental hydration free energy differences for charged molecules. However, continuum solvents can implicitly include the cavity potential Φ_{Cavity} during fitting of solute atomic radii.

- An empty cavity has no electrostatic potential under continuum electrostatics (*i.e.* for the PBE with the potential at the boundary of the domain set to zero).
- An empty cavity experiences a potential (Φ_{Cavity}) for explicit water simulations under periodic boundary conditions (*i.e.* PME) due to preferential orientation of water around the empty cavity.
- In addition to the cavity potential, an empty cavity under experimental conditions (or for an explicit water droplet simulation embedded in a vacuum) experiences a potential due to preferential orientation of water at the vacuum-liquid interface ($\Phi_{\text{Interface}}$).

Supplementary Table S1. The implicit solvent dispersion term was fit to best match dispersion free energy calculated for a set of small molecules in explicit solvent using an integral offset ($d = 1.056\text{\AA}$) which pushes the start of the dispersion integral further into the solvent. R^2 for explicit vs implicit dispersion values: 0.987

Molecule	ΔW_{disp} Explicit	ΔW_{disp} Implicit	SE	UE
acetic acid	-8.40	-7.89	0.51	0.51
formic acid	-6.10	-5.25	0.85	0.85
ethanol	-7.90	-7.48	0.42	0.42
isopropanol	-9.50	-9.46	0.04	0.04
methanol	-5.60	-4.89	0.71	0.71
propanol	-9.80	-9.71	0.09	0.09
acetaldehyde	-7.40	-6.64	0.76	0.76
formaldehyde	-5.00	-4.33	0.67	0.67
butane	-10.70	-10.92	-0.22	0.22
methane	-4.40	-3.64	0.76	0.76
acetamide	-9.00	-8.53	0.47	0.47
dimethylacetamide	-12.50	-12.91	-0.41	0.41
dimethylformamide	-10.90	-10.96	-0.06	0.06
N-methylacetamide	-11.20	-10.92	0.28	0.28
N-methylformamide	-9.10	-8.71	0.39	0.39
propionamide	-10.70	-10.75	-0.05	0.05
ammonia	-3.40	-2.93	0.47	0.47
dimethylamine	-8.30	-8.10	0.20	0.20
ethylamine	-8.20	-8.04	0.16	0.16
methylamine	-6.10	-5.66	0.44	0.44
propylamine	-10.00	-10.22	-0.22	0.22
trimethylamine	-9.80	-10.26	-0.46	0.46
benzene	-11.70	-11.80	-0.10	0.10
p-cresol	-14.10	-14.76	-0.66	0.66
ethylbenzene	-14.70	-15.80	-1.10	1.10
phenol	-12.50	-12.80	-0.30	0.30
toluene	-13.30	-13.86	-0.56	0.56
ethylimidazole	-13.60	-14.07	-0.47	0.47
imidazole	-10.30	-9.95	0.35	0.35
ethylindole	-17.90	-20.42	-2.52	2.52
N-methylpyrrolidine	-11.90	-13.46	-1.56	1.56
pyrrolidine	-10.60	-11.54	-0.94	0.94
dimethylsulfide	-10.10	-9.68	0.42	0.42
ethanethiol	-9.70	-9.39	0.31	0.31
methyl ethyl sulfide	-12.00	-11.93	0.07	0.07
methanethiol	-7.70	-7.06	0.64	0.64
water	-2.50	-2.07	0.43	0.43
			0.00	0.52

Supplementary Table S2. Initial (van der Waals) and optimized solute diameters for GK, ddCOSMO, and PB. All diameter values given in Angstroms (Å). APBS, ddCOSMO, and GK optimized diameters deviate from original van der Waals values by an average of 9.3%, 9.9%, and 14.0%, respectively.

Element	SMARTS Group	vdW	APBS	ddCOSMO	GK
H	[#1]([#8])	2.665	3.247	2.822	1.851
H	[#1](c(n))	3.000	2.906	2.602	2.973
H	[#1]([CH3](N(C(=O))))	2.930	2.289	2.003	2.797
H	[#1][#6;D3]	2.583			
	[#1]([O;D2;H2])	2.655	2.574	2.758	3.247
	[#1]([OH1](C=O))	2.655			
H	[#1]([N;D3]([C]([#1])))	2.700			
	[#1]([N+;D3])	2.700	3.649	3.404	2.910
	[#1][C;D4][OH]	2.870			
H	[#1]([N;D4])	2.480			
	[#1](C=O)	2.920	3.035	2.686	2.950
	[#1]([CH2]([N]))	2.960			
H	[#1][#16;D2]	2.770			
	[#1]([CH4])	2.900	2.996	2.928	3.362
	[#1][CH3][CH3]	2.960			
	[#1]([CH3]([CH2](S))))	2.980			
H	[#1](c)	2.980	3.002	2.635	3.250
H	[#1][O;D2][#6]	2.655			
	[#1]([C][S;D2])	2.870	2.653	3.392	2.799
	[#1]([CH3]([O]([CH2,CH3,c])))	2.890			
H	[#1]([N](C=O))	2.590			
	[#1](N(~C(c)))	2.590			
	[#1]([N;D3])	2.700	2.909	2.836	3.048
	[#1]([CH1;D4])	2.980			
	[#1]([CH3](C=O))	2.980			
H	[#1]([CH2]([CH3]))	2.980	3.021	3.885	3.465
	[#1]([CH2])	2.980			
H	[#1][C;D4][N,c;D3]	2.880	3.056	2.996	3.488
H	[#1]([CH2]([CH2]([N])))	2.960	3.991	4.144	3.806
H	[#1][CH3]	2.960	3.143	3.374	3.268
C	[cH1][n+]	3.650			
	c(n)	3.780	3.859	3.766	3.542
	c([cH0])	3.800			
C	[CH0;D4]	3.600			
	[CH1]([CH3])([CH3])([OH1])	3.650	3.580	3.623	4.761
	[CH3]([OH1])	3.760			
C	[#6;D3](=[#8])	3.820	4.267	3.776	3.777
	[CH3]([CH2](S))	3.820			
C	[CH4]	3.780	3.411	3.545	3.813
C	c	3.800			
	c(c([cH0]))	3.800			
	c(c(c([cH0])))	3.800	3.829	3.893	4.708
	ccccncc	3.800			
	cccnccc	3.800			

	cnccccc	3.800			
	cnccccc	3.800			
C	[CH3]	3.820	3.506	3.309	4.225
C	[#6;D3]([*])([*])(=[*])	3.800	2.930	4.506	4.305
C	[cH0]	3.800	3.691	3.766	3.788
C	[CH1]	3.650			
	C(S)	3.780			
	[#6;D3](=[#8-])	3.820	3.397	3.255	4.425
	[C](=O)([N,c])	3.820			
C	[#6]	3.820	3.303	2.854	4.367
N	n	3.710	3.602	4.034	3.239
N	[N]([C]([#1]))	3.710	2.884	2.990	3.245
N	[NH3]	3.710	3.524	3.297	3.501
N	[N+]	3.710			
	[#7+](=[#6](-[#7](-[H])-[H])-[#7](-[H])-[H])(-[H])-[H])	3.710	2.226	4.700	2.854
N	N(~C(c))	3.710	3.710	3.498	3.904
N	[#7;D3]	3.710			
	[N](C=O)	3.710	3.424	3.491	4.678
O	[#8;D1]=[#6]	3.300			
	[O](=C)	3.300			
	O(=C(c))	3.300	2.984	3.356	2.986
	[#8;D1]=[#15]	3.360			
	[#8;D2](~[#15])	3.450			
O	[#8;D1][*]	3.400	2.801	3.128	3.029
O	[O;D2][#6]	3.405	3.168	3.134	3.089
O	[O;D2;H2]	3.405	3.124	2.906	2.696
O	[OD1]~C~[OD1]	3.450	3.039	3.156	3.233
O	[OH1](C=O)	3.405	3.399	2.819	2.631
O	[O;D1](~S)	3.510	3.711	3.787	4.914
F	[#9]	3.220	4.114	4.166	3.489
P	[#15]	4.450	4.870	4.353	4.105
S	[SH2]	4.005	4.066	3.800	4.047
S	[#16]	4.005			
	[SH]2	4.005	4.435	4.194	4.459

Supplementary Table S3. Solvation free energy components - all listed in kcal/mol. Optimized radii were used in the computation of all electrostatic energies. Non-polar terms were optimized using GK and used with all three electrostatics models. A constant solvent pressure of 0.0334 kcal/mol/Å³, a constant dispersion offset of 1.056 Å, and a unitless scale factor of 0.75 were used in the calculation of all nonpolar energies.

Molecule	ΔW_{cav}	ΔW_{disp}	ΔW_{apolar}	ΔU_{elec}		
				APBS	ddCOSMO	GK
methane	6.49	-3.64	2.86	-0.43	-0.46	-0.94
ethane	9.06	-6.42	2.65	-0.34	-0.55	-0.23
propane	11.27	-8.72	2.55	-0.31	-0.27	-0.19
isobutane	13.22	-10.56	2.66	-0.60	-0.59	-0.82
butane	13.44	-10.92	2.52	-0.43	-0.32	-0.84
isopentane	15.08	-12.40	2.68	-0.48	-0.40	-0.30
neopentane	14.88	-12.29	2.59	-0.56	-0.55	-0.42
pentane	15.59	-13.08	2.51	-0.41	-0.34	-0.18
cyclopentane	13.84	-12.17	1.67	-0.46	-0.27	-0.22
methylcyclopentane	15.79	-13.85	1.94	-0.74	-0.42	-0.90
2,3-dimethylbutane	16.77	-13.92	2.85	-0.62	-0.63	-0.37
neohexane	16.69	-14.02	2.67	-0.57	-0.53	-0.50
hexane	17.76	-15.24	2.52	-0.63	-0.40	-0.60
cyclohexane	15.56	-13.85	1.71	-0.37	-0.82	0.05
methylcyclohexane	17.46	-15.47	1.99	-0.48	-0.63	-0.06
heptane	19.92	-17.40	2.52	-0.50	-0.38	-0.23
octane	22.08	-19.55	2.53	-0.55	-0.42	-0.27
decane	26.40	-23.86	2.54	-0.64	-0.46	-0.36
methanol	7.43	-4.89	2.54	-6.09	-6.06	-6.39
ethanol	9.95	-7.48	2.47	-6.56	-6.36	-5.94
isopropanol	11.95	-9.46	2.49	-6.26	-6.16	-6.87
propanol	12.05	-9.71	2.34	-6.66	-6.87	-6.57
butanol	14.22	-11.89	2.33	-6.51	-6.73	-6.52
butan-2-ol	14.07	-11.58	2.49	-5.88	-5.30	-5.18
pentanol	16.38	-14.06	2.32	-6.64	-6.89	-6.65
pentan-2-ol	16.22	-13.72	2.49	-5.80	-5.37	-6.30
hexanol	18.52	-16.21	2.31	-6.67	-6.89	-6.71
3-methylbutan-1-ol	15.87	-13.37	2.50	-6.80	-7.01	-6.96
cyclopentanol	14.41	-12.80	1.62	-5.52	-4.37	-6.02
ethyleneglycol	10.51	-8.44	2.07	-10.33	-10.20	-11.23
glucose	19.80	-18.29	1.51	-29.21	-29.52	-27.83
phenol	13.91	-12.80	1.10	-7.33	-6.58	-7.21
2-fluorophenol	14.23	-13.48	0.75	-5.91	-5.66	-6.94
4-fluorophenol	14.29	-13.50	0.80	-7.05	-6.99	-6.49
ammonia	5.54	-2.93	2.61	-5.83	-6.92	-6.91
methylamine	8.05	-5.66	2.39	-6.40	-6.93	-6.60
ethylamine	10.45	-8.04	2.41	-6.55	-7.37	-6.29
propylamine	12.60	-10.22	2.38	-6.28	-7.09	-6.25

dimethylamine	10.36	-8.10	2.26	-4.98	-6.29	-6.83
diethylamine	15.12	-12.58	2.54	-3.16	-3.89	-7.35
diisopropylamine	18.53	-15.75	2.78	-4.29	-4.22	-6.46
dibutylamine	24.05	-21.21	2.84	-4.01	-5.27	-7.84
trimethylamine	12.36	-10.26	2.10	-5.44	-5.19	-3.64
triethylamine	18.72	-15.92	2.80	-7.37	-6.89	-4.69
morpholine	15.07	-12.48	2.59	-9.63	-12.94	-12.05
benzenamine	14.54	-13.33	1.20	-5.55	-6.77	-7.44
acetamide	10.33	-8.53	1.80	-12.57	-12.01	-11.70
N-methylacetamide	12.76	-10.92	1.83	-11.33	-10.92	-11.02
dimethylacetamide	14.55	-12.91	1.64	-9.91	-10.49	-11.65
N-methylformamide	10.40	-8.71	1.69	-11.06	-10.91	-10.19
dimethylformamide	12.52	-10.96	1.56	-9.69	-10.34	-10.32
propionamide	12.46	-10.75	1.70	-11.58	-10.37	-11.14
benzamide	16.28	-15.86	0.42	-10.78	-11.30	-11.68
N,N,4-trimethylbenzamide	22.43	-21.79	0.63	-9.46	-10.04	-10.19
pyrrole	11.38	-10.24	1.15	-6.16	-5.93	-6.30
pyrrolidine	13.50	-11.54	1.97	-4.27	-6.57	-6.90
pyridine	12.68	-11.59	1.09	-5.46	-5.63	-5.09
N-methylpyrrolidine	15.58	-13.46	2.12	-4.55	-6.50	-6.11
N-methyl-2-pyridone	15.46	-14.58	0.88	-10.30	-10.09	-12.11
3-methylindole	18.23	-18.21	0.01	-6.42	-6.02	-5.60
imidazole	10.84	-9.95	0.88	-10.31	-10.68	-10.57
5-fluorouracil	13.72	-13.89	-0.17	-17.36	-17.35	-15.85
benzene	13.18	-11.80	1.38	-2.68	-3.09	-2.51
toluene	15.17	-13.86	1.32	-2.72	-2.91	-1.90
p-cresol	15.90	-14.76	1.14	-7.36	-6.51	-6.66
ethylbenzene	17.32	-15.80	1.53	-2.74	-2.94	-1.87
fluorobenzene	13.54	-12.58	0.96	-2.57	-3.46	-2.54
methoxymethane	9.85	-7.57	2.28	-4.10	-4.48	-4.25
methoxyethane	12.30	-9.92	2.38	-4.23	-4.46	-4.60
ethoxyethane	14.72	-12.21	2.51	-4.08	-4.18	-4.22
1,2-dimethoxyethane	15.54	-13.35	2.19	-7.10	-6.61	-8.17
1-methoxypropane	14.48	-12.14	2.34	-4.42	-4.60	-4.98
tetrahydrofuran	12.72	-11.33	1.39	-4.41	-5.32	-4.53
2-methyltetrahydrofuran	14.83	-13.09	1.74	-4.91	-5.80	-5.05
tetrahydropyran	14.56	-13.10	1.46	-4.22	-4.15	-3.37
1,4-dioxane	13.55	-12.30	1.25	-8.56	-6.93	-7.75
methanethiol	8.69	-7.06	1.63	-3.04	-2.82	-2.66
ethanethiol	11.06	-9.39	1.67	-2.94	-2.88	-2.59
propanethiol	13.25	-11.55	1.71	-4.02	-3.51	-3.60
butanethiol	15.39	-13.74	1.65	-3.06	-2.94	-2.55
acetic acid	9.85	-7.89	1.96	-8.52	-8.71	-8.48

propionic acid	12.03	-10.14	1.89	-8.36	-8.21	-8.14
butyric acid	14.20	-12.33	1.87	-8.57	-8.34	-8.63
dimethylsulfide	10.87	-9.68	1.19	-3.30	-2.75	-2.96
hydrogen sulfide	6.28	-4.14	2.14	-3.20	-2.91	-2.90
methyl ethyl sulfide	13.25	-11.93	1.33	-2.97	-2.52	-2.95
acetaldehyde	9.02	-6.64	2.38	-6.56	-6.22	-5.95
formaldehyde	6.61	-4.33	2.28	-5.12	-5.47	-4.79
dimethylsulfoxide	12.20	-10.43	1.77	-9.38	-11.16	-12.65
triethylphosphate	24.37	-23.15	1.22	-11.52	-10.24	-8.74
water	4.80	-2.07	2.73	-8.10	-9.07	-9.01
acetate (-)	9.82	-7.78	2.04	-90.48	-88.72	-89.79
propanoate (-)	11.97	-10.03	1.94	-87.61	-89.04	-87.98
methyl hydrogen phosphate (-)	13.05	-12.98	0.07	-86.49	-86.36	-86.20
dimethylphosphate (-)	15.33	-15.03	0.30	-85.42	-85.66	-85.77
methylammonium (+)	8.15	-5.67	2.48	-68.79	-66.41	-69.03
dimethylammonium (+)	10.64	-8.25	2.39	-61.75	-60.61	-60.93
trimethylammonium (+)	12.76	-10.51	2.26	-55.27	-54.34	-52.92
butylammonium (+)	14.77	-12.36	2.40	-64.40	-65.22	-64.01
guanidinium (+)	10.42	-8.56	1.86	-53.35	-53.37	-53.95
methylguanidinium (+)	12.67	-10.79	1.89	-52.25	-51.15	-49.19
imidazolium (+)	10.28	-9.76	0.52	-57.13	-56.30	-56.81
4-methylimidazolium (+)	12.83	-11.95	0.89	-54.31	-53.83	-53.39

Supplementary Table S4. Experimental and calculated total solvation free energies for all test molecules used to parameterize the AMOEBA implicit solvent model. Experimental energies were sourced from the FreeSolv database (version 0.51) unless otherwise noted. All energies reported in kcal/mol.

Molecule	ΔG_{exp}	$\Delta W_{\text{hydration}}$		
		APBS	ddCOSMO	GK
methane	2.00	2.43	2.39	1.92
ethane	1.83	2.31	2.10	2.42
propane	2.00	2.24	2.28	2.36
isobutane	2.30	2.06	2.07	1.83
butane	2.10	2.09	2.21	1.68
isopentane	2.38	2.20	2.28	2.38
neopentane	2.51	2.03	2.04	2.17
pentane	2.30	2.10	2.17	2.33
cyclopentane	1.20	1.21	1.40	1.45
methylcyclopentane	1.59	1.20	1.52	1.03
2,3-dimethylbutane	2.34	2.23	2.22	2.48
neohexane	2.51	2.10	2.14	2.17
hexane	2.48	1.89	2.12	1.92
cyclohexane	1.23	1.35	0.89	1.76
methylcyclohexane	1.70	1.52	1.36	1.93
heptane	2.67	2.02	2.15	2.29
octane	2.88	1.97	2.11	2.26
decane	3.16	1.90	2.08	2.18
methanol	-5.10	-3.55	-3.51	-3.84
ethanol	-5.00	-4.09	-3.89	-3.48
isopropanol	-4.74	-3.77	-3.66	-4.38
propanol	-4.85	-4.33	-4.53	-4.23
butanol	-4.72	-4.18	-4.40	-4.20
butan-2-ol	-4.62	-3.38	-2.80	-2.69
pentanol	-4.57	-4.32	-4.57	-4.32
pentan-2-ol	-4.39	-3.31	-2.87	-3.81
hexanol	-4.40	-4.36	-4.58	-4.40
3-methylbutan-1-ol	-4.42	-4.29	-4.50	-4.46
cyclopentanol	-5.49	-3.90	-2.75	-4.41
ethyleneglycol	-9.30	-8.26	-8.14	-9.16
glucose	-25.47	-27.70	-28.01	-26.32
phenol	-6.60	-6.23	-5.48	-6.11
2-fluorophenol	-5.29	-5.16	-4.91	-6.19
4-fluorophenol	-6.19	-6.26	-6.19	-5.69
ammonia	-4.29	-3.21	-4.31	-4.29
methylamine	-4.55	-4.00	-4.53	-4.21
ethylamine	-4.50	-4.14	-4.97	-3.88
propylamine	-4.39	-3.90	-4.71	-3.87

dimethylamine	-4.29	-2.72	-4.03	-4.57
diethylamine	-4.07	-0.62	-1.35	-4.81
diisopropylamine	-3.22	-1.52	-1.45	-3.68
dibutylamine	-3.24	-1.17	-2.42	-5.00
trimethylamine	-3.20	-3.34	-3.09	-1.54
triethylamine	-3.22	-4.56	-4.09	-1.89
morpholine	-7.17	-7.04	-10.35	-9.46
benzenamine	-5.49	-4.35	-5.57	-6.24
acetamide	-9.71	-10.78	-10.22	-9.91
N-methylacetamide	-10.00	-9.50	-9.09	-9.19
dimethylacetamide	-8.50 ^a	-8.28	-8.85	-10.01
N-methylformamide	-10.00 ^a	-9.36	-9.21	-8.49
dimethylformamide	-7.81	-8.13	-8.78	-8.76
propionamide	-9.40	-9.88	-8.66	-9.43
benzamide	-11.00	-10.36	-10.88	-11.26
N,N,4-trimethylbenzamide	-9.76	-8.83	-9.41	-9.56
pyrrole	-4.78	-5.02	-4.78	-5.16
pyrrolidine	-5.48	-2.31	-4.61	-4.93
pyridine	-4.69	-4.37	-4.54	-4.00
N-methylpyrrolidine	-3.98 ^a	-2.44	-4.39	-3.99
N-methyl-2-pyridone	-10.00 ^a	-9.42	-9.21	-11.23
3-methylindole	-5.88	-6.40	-6.00	-5.59
imidazole	-9.63	-9.43	-9.79	-9.68
5-fluorouracil	-16.92	-17.53	-17.52	-16.02
benzene	-0.90	-1.30	-1.70	-1.13
toluene	-0.90	-1.41	-1.60	-0.58
p-cresol	-6.13	-6.22	-5.37	-5.52
ethylbenzene	-0.79	-1.21	-1.41	-0.35
fluorobenzene	-0.80	-1.61	-2.50	-1.58
methoxymethane	-1.91	-1.82	-2.20	-1.97
methoxyethane	-2.10	-1.85	-2.08	-2.22
ethoxyethane	-1.59	-1.56	-1.66	-1.71
1,2-dimethoxyethane	-4.84	-4.91	-4.42	-5.98
1-methoxypropane	-1.66	-2.08	-2.26	-2.64
tetrahydrofuran	-3.47	-3.03	-3.94	-3.15
2-methyltetrahydrofuran	-3.30	-3.17	-4.06	-3.31
tetrahydropyran	-3.12	-2.77	-2.70	-1.92
1,4-dioxane	-5.06	-7.31	-5.68	-6.50
methanethiol	-1.20	-1.41	-1.18	-1.02
ethanethiol	-1.14	-1.27	-1.20	-0.92
propanethiol	-1.10	-2.31	-1.80	-1.89
butanethiol	-0.99 ^b	-1.41	-1.29	-0.90
acetic acid	-6.69	-6.56	-6.75	-6.53

propionic acid	-6.46	-6.47	-6.32	-6.25
butyric acid	-6.35	-6.70	-6.46	-6.76
dimethylsulfide	-1.61 ^b	-2.11	-1.56	-1.77
hydrogen sulfide	-0.70	-1.06	-0.77	-0.77
methyl ethyl sulfide	-1.50 ^b	-1.64	-1.19	-1.62
acetaldehyde	-3.50	-4.18	-3.84	-3.57
formaldehyde	-2.75	-2.84	-3.19	-2.50
dimethylsulfoxide	-9.28	-7.61	-9.39	-10.88
triethylphosphate	-7.50	-10.29	-9.02	-7.52
water	-6.30	-5.36	-6.34	-6.28
acetate (-)	-87.76	-88.44	-86.68	-87.75
propanoate (-)	-85.94	-85.67	-87.10	-86.04
methyl hydrogen phosphate (-)	-85.63 ^c	-85.91	-85.79	-85.62
dimethylphosphate (-)	-85.28 ^c	-85.03	-85.26	-85.38
methylammonium (+)	-66.60	-66.31	-63.93	-66.55
dimethylammonium (+)	-58.87	-59.36	-58.22	-58.55
trimethylammonium (+)	-51.28	-53.02	-52.09	-50.66
butylammonium (+)	-61.15	-61.99	-62.81	-61.61
guanidinium (+)	-52.31 ^c	-51.50	-51.51	-52.09
methylguanidinium (+)	-47.18 ^c	-50.36	-49.27	-47.30
imidazolium (+)	-56.02	-56.61	-55.78	-56.29
4-methylimidazolium (+)	-52.57	-53.43	-52.95	-52.51

^aGallicchio, 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 (5), 517-529.

^bRizzo, Robert C., et al. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. *Journal of Chemical Theory and Computation* **2006**, 2 (1), 128-39.

^c Determined using BAR with AMOEBA explicit solvent, see Supplementary Table S6

Supplementary Table S6. Signed and unsigned errors, relative to experiment, for all test molecules. Errors given in kcal/mol. Experimental and calculated total solvation free energies used to compute errors shown in Supplementary Table S4.

Molecule	SE			UE		
	APBS	ddCOSMO	GK	APBS	ddCOSMO	GK
methane	0.43	0.39	-0.08	0.43	0.39	0.08
ethane	0.48	0.27	0.59	0.48	0.27	0.59
propane	0.24	0.28	0.36	0.24	0.28	0.36
isobutane	-0.24	-0.23	-0.47	0.24	0.23	0.47
butane	-0.01	0.11	-0.42	0.01	0.11	0.42
isopentane	-0.18	-0.10	0.00	0.18	0.10	0.00
neopentane	-0.48	-0.47	-0.34	0.48	0.47	0.34
pentane	-0.20	-0.13	0.03	0.20	0.13	0.03
cyclopentane	0.01	0.20	0.25	0.01	0.20	0.25
methylcyclopentane	-0.39	-0.07	-0.56	0.39	0.07	0.56
2,3-dimethylbutane	-0.11	-0.12	0.14	0.11	0.12	0.14
neohexane	-0.41	-0.37	-0.34	0.41	0.37	0.34
hexane	-0.59	-0.36	-0.56	0.59	0.36	0.56
cyclohexane	0.12	-0.34	0.53	0.12	0.34	0.53
methylcyclohexane	-0.18	-0.34	0.23	0.18	0.34	0.23
heptane	-0.65	-0.52	-0.38	0.65	0.52	0.38
octane	-0.91	-0.77	-0.62	0.91	0.77	0.62
decane	-1.26	-1.08	-0.98	1.26	1.08	0.98
methanol	1.55	1.59	1.26	1.55	1.59	1.26
ethanol	0.91	1.11	1.52	0.91	1.11	1.52
isopropanol	0.97	1.08	0.36	0.97	1.08	0.36
propanol	0.52	0.32	0.62	0.52	0.32	0.62
butanol	0.54	0.32	0.52	0.54	0.32	0.52
butan-2-ol	1.24	1.82	1.93	1.24	1.82	1.93
pentanol	0.25	0.00	0.25	0.25	0.00	0.25
pentan-2-ol	1.08	1.52	0.58	1.08	1.52	0.58
hexanol	0.04	-0.18	0.00	0.04	0.18	0.00
3-methylbutan-1-ol	0.13	-0.08	-0.04	0.13	0.08	0.04
cyclopentanol	1.59	2.74	1.08	1.59	2.74	1.08
ethyleneglycol	1.04	1.16	0.14	1.04	1.16	0.14
glucose	-2.23	-2.54	-0.85	2.23	2.54	0.85
phenol	0.37	1.12	0.49	0.37	1.12	0.49
2-fluorophenol	0.13	0.38	-0.90	0.13	0.38	0.90
4-fluorophenol	-0.07	0.00	0.50	0.07	0.00	0.50
ammonia	1.08	-0.02	0.00	1.08	0.02	0.00
methylamine	0.55	0.02	0.34	0.55	0.02	0.34
ethylamine	0.36	-0.47	0.62	0.36	0.47	0.62
propylamine	0.49	-0.32	0.52	0.49	0.32	0.52
dimethylamine	1.57	0.26	-0.28	1.57	0.26	0.28

diethylamine	3.45	2.72	-0.74	3.45	2.72	0.74
diisopropylamine	1.70	1.77	-0.46	1.70	1.77	0.46
dibutylamine	2.07	0.82	-1.76	2.07	0.82	1.76
trimethylamine	-0.14	0.11	1.66	0.14	0.11	1.66
triethylamine	-1.34	-0.87	1.33	1.34	0.87	1.33
morpholine	0.13	-3.18	-2.29	0.13	3.18	2.29
benzenamine	1.14	-0.08	-0.75	1.14	0.08	0.75
acetamide	-1.07	-0.51	-0.20	1.07	0.51	0.20
N-methylacetamide	0.50	0.91	0.81	0.50	0.91	0.81
dimethylacetamide	0.22	-0.35	-1.51	0.22	0.35	1.51
N-methylformamide	0.64	0.79	1.51	0.64	0.79	1.51
dimethylformamide	-0.32	-0.97	-0.95	0.32	0.97	0.95
propionamide	-0.48	0.74	-0.03	0.48	0.74	0.03
benzamide	0.64	0.12	-0.26	0.64	0.12	0.26
N,N,4-trimethylbenzamide	0.93	0.35	0.20	0.93	0.35	0.20
pyrrole	-0.24	0.00	-0.38	0.24	0.00	0.38
pyrrolidine	3.17	0.87	0.55	3.17	0.87	0.55
pyridine	0.32	0.15	0.69	0.32	0.15	0.69
N-methylpyrrolidine	1.54	-0.41	-0.01	1.54	0.41	0.01
N-methyl-2-pyridone	0.58	0.79	-1.23	0.58	0.79	1.23
3-methylindole	-0.52	-0.12	0.29	0.52	0.12	0.29
imidazole	0.20	-0.16	-0.05	0.20	0.16	0.05
5-fluorouracil	-0.61	-0.60	0.90	0.61	0.60	0.90
benzene	-0.40	-0.80	-0.23	0.40	0.80	0.23
toluene	-0.51	-0.70	0.32	0.51	0.70	0.32
p-cresol	-0.09	0.76	0.61	0.09	0.76	0.61
ethylbenzene	-0.42	-0.62	0.44	0.42	0.62	0.44
fluorobenzene	-0.81	-1.70	-0.78	0.81	1.70	0.78
methoxymethane	0.09	-0.29	-0.06	0.09	0.29	0.06
methoxyethane	0.25	0.02	-0.12	0.25	0.02	0.12
ethoxyethane	0.03	-0.07	-0.12	0.03	0.07	0.12
1,2-dimethoxyethane	-0.07	0.42	-1.14	0.07	0.42	1.14
1-methoxypropane	-0.42	-0.60	-0.98	0.42	0.60	0.98
tetrahydrofuran	0.44	-0.47	0.32	0.44	0.47	0.32
2-methyltetrahydrofuran	0.13	-0.76	-0.01	0.13	0.76	0.01
tetrahydropyran	0.35	0.42	1.20	0.35	0.42	1.20
1,4-dioxane	-2.25	-0.62	-1.44	2.25	0.62	1.44
methanethiol	-0.21	0.02	0.18	0.21	0.02	0.18
ethanethiol	-0.13	-0.06	0.22	0.13	0.06	0.22
propanethiol	-1.21	-0.70	-0.79	1.21	0.70	0.79
butanethiol	-0.42	-0.30	0.09	0.42	0.30	0.09
acetic acid	0.13	-0.06	0.16	0.13	0.06	0.16
propionic acid	-0.01	0.14	0.21	0.01	0.14	0.21

butyric acid	-0.35	-0.11	-0.41	0.35	0.11	0.41
dimethylsulfide	-0.50	0.05	-0.16	0.50	0.05	0.16
hydrogen sulfide	-0.36	-0.07	-0.07	0.36	0.07	0.07
methyl ethyl sulfide	-0.14	0.31	-0.12	0.14	0.31	0.12
acetaldehyde	-0.68	-0.34	-0.07	0.68	0.34	0.07
formaldehyde	-0.09	-0.44	0.25	0.09	0.44	0.25
dimethylsulfoxide	1.67	-0.11	-1.60	1.67	0.11	1.60
triethylphosphate	-2.79	-1.52	-0.02	2.79	1.52	0.02
water	0.94	-0.04	0.02	0.94	0.04	0.02
acetate (-)	-0.68	1.08	0.01	0.68	1.08	0.01
propanoate (-)	0.27	-1.16	-0.10	0.27	1.16	0.10
methyl hydrogen phosphate (-)	-0.28	-0.16	0.01	0.28	0.16	0.01
dimethylphosphate (-)	0.25	0.02	-0.10	0.25	0.02	0.10
methylammonium (+)	0.29	2.67	0.05	0.29	2.67	0.05
dimethylammonium (+)	-0.49	0.65	0.32	0.49	0.65	0.32
trimethylammonium (+)	-1.74	-0.81	0.62	1.74	0.81	0.62
butylammonium (+)	-0.84	-1.66	-0.46	0.84	1.66	0.46
guanidinium (+)	0.81	0.80	0.22	0.81	0.80	0.22
methylguanidinium (+)	-3.18	-2.09	-0.12	3.18	2.09	0.12
imidazolium (+)	-0.59	0.24	-0.27	0.59	0.24	0.27
4-methylimidazolium (+)	-0.86	-0.38	0.06	0.86	0.38	0.06
All	0.05	0.00	0.00	0.70	0.63	0.51
Neutrals	0.14	0.01	-0.01	0.67	0.58	0.55

Supplementary Table S5. BAR free energies and standard deviations for phosphate and guanidinium compounds. All BAR values calculated across seven equally spaced lambda windows. All values reported in kcal/mol.

vdW λ Windows	Methyl-phosphate		Dimethyl-phosphate		Guanidinium		Methyl-guanidinium	
	Free Energy	Std. Dev.	Free Energy	Std. Dev.	Free Energy	Std. Dev.	Free Energy	Std. Dev.
0.000 - 0.143	0.05	0.00	0.06	0.00	0.04	0.00	0.05	0.00
0.143 - 0.286	1.08	0.00	1.31	0.01	0.75	0.01	0.99	0.01
0.286 - 0.429	4.71	0.06	5.46	0.07	3.32	0.06	4.22	0.07
0.429 - 0.571	2.22	0.07	2.26	0.07	2.00	0.08	2.33	0.10
0.571 - 0.714	-0.26	0.03	-0.34	0.03	-0.11	0.03	-0.14	0.04
0.714 - 0.857	-1.97	0.02	-2.31	0.02	-1.35	0.02	-1.67	0.02
0.857 - 1.000	-3.69	0.01	-4.23	0.01	-2.56	0.02	-3.24	0.02
vdW Total	2.13	0.09	2.21	0.10	2.09	0.11	2.54	0.13
Electrostatic λ Windows								
0.000 - 0.143	-2.57	0.02	-2.46	0.03	-0.53	0.02	-0.39	0.03
0.143 - 0.286	-6.65	0.03	-6.27	0.03	-2.86	0.02	-2.35	0.03
0.286 - 0.429	-11.08	0.04	-10.42	0.04	-5.24	0.03	-4.39	0.03
0.429 - 0.571	-15.94	0.05	-15.14	0.04	-7.75	0.03	-6.48	0.03
0.571 - 0.714	-22.09	0.06	-21.07	0.06	-10.52	0.03	-8.76	0.03
0.714 - 0.857	-29.14	0.07	-27.45	0.06	-13.48	0.03	-11.18	0.04
0.857 - 1.000	-37.14	0.06	-34.44	0.06	-16.67	0.04	-13.88	0.04
Electrostatic Total	-124.62	0.19	-117.24	0.19	-57.04	0.07	-47.43	0.08
Vacuum λ Windows								
0.000 - 0.143	-0.73	0.00	-0.60	0.00	-0.05	0.00	0.05	0.00
0.143 - 0.286	-2.21	0.01	-1.79	0.01	-0.15	0.00	0.14	0.00
0.286 - 0.429	-3.69	0.01	-2.99	0.01	-0.26	0.00	0.24	0.00
0.429 - 0.571	-5.15	0.01	-4.18	0.01	-0.36	0.01	0.33	0.00
0.571 - 0.714	-6.69	0.01	-5.43	0.01	-0.48	0.01	0.42	0.00
0.714 - 0.857	-8.21	0.01	-6.69	0.01	-0.61	0.01	0.51	0.00
0.857 - 1.000	-9.78	0.01	-7.98	0.01	-0.73	0.01	0.60	0.00
Vacuum Total	-36.46	0.03	-29.65	0.03	-2.64	0.01	2.28	0.00
Hydration Total	-86.02	0.21	-85.38	0.22	-52.31	0.14	-47.18	0.15

Supplementary Table S7. (A) Detailed simulation conditions for validation set molecules with explicit solvent. (B) Simulation box content for validation set molecules in explicit water.

(A) Equilibration and production simulation protocols.

Purpose	Simulation Time	Ensemble	Temperature
	(ns)		(K)
Equilibration	1	NVT	100
	1	NVT	200
	1	NVT	300
	1	NPT	298
Production	1	NVT	298

(B) Simulation box sizes and content

Type	PDB	Experimental Structure Method	Explicit Solvent Water/Na+/Cl- Number	Explicit Solvent Cubic Box Side Length (Å)
RNA Helix	1MIS	NMR	11955/14/0	50.591
	2JXQ	NMR	16968/18/0	56.617
	1F5G	NMR	19629/18/0	59.169
	2L8F	NMR	24054/20/0	63.333
RNA Hairpin	1ZIH	NMR	7629/11/0	43.769
	1SZY	NMR	22941/20/0	62.323
	2KOC	NMR	12933/13/0	51.622
DNA Helix	1D20	NMR	14022/18/0	53.443
	2HKB	NMR	22806/22/0	62.464
Protein	1BPI	X-ray (1.09 Å)	18906/0/6	58.716
	1L2Y	NMR	7164/0/1	42.473
	1UBQ	X-ray (1.80 Å)	23052/0/1	62.933
	1UCS	X-ray (0.62 Å)	13182/0/0	52.623
	1VII	NMR	11505/0/2	50.037
	1WM3	X-ray (1.20 Å)	15390/0/1	55.628
	2OED	NMR	14451/2/0	54.063
	2PPN	X-ray (0.92 Å)	26052/0/4	65.701
	6LYT	X-ray (1.90 Å)	29667/0/9	68.637