

RESEARCH ARTICLE | AUGUST 01 2023

A generalized Kirkwood implicit solvent for the polarizable AMOEBA protein model

Rae A. Corrigan ; Andrew C. Thiel ; Jack R. Lynn; Thomas L. Casavant; Pengyu Ren ;
Jay W. Ponder ; Michael J. Schnieders  



J. Chem. Phys. 159, 054102 (2023)

<https://doi.org/10.1063/5.0158914>



CrossMark



The Journal of Chemical Physics

Special Topic: Adhesion and Friction

Submit Today!



A generalized Kirkwood implicit solvent for the polarizable AMOEBA protein model

Cite as: J. Chem. Phys. 159, 054102 (2023); doi: 10.1063/5.0158914

Submitted: 18 May 2023 • Accepted: 17 July 2023 •

Published Online: 1 August 2023



Rae A. Corrigan,¹ Andrew C. Thiel,¹ Jack R. Lynn,¹ Thomas L. Casavant,² Pengyu Ren,³ Jay W. Ponder,⁴ and Michael J. Schnieders^{1,5,a)}

AFFILIATIONS

¹Roy J. Carver Department of Biomedical Engineering, The University of Iowa, Iowa City, Iowa 52242, USA

²Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, Iowa 52242, USA

³Department of Biomedical Engineering, The University of Texas in Austin, Austin, Texas 78712, USA

⁴Department of Chemistry, Washington University in St. Louis, St. Louis, Missouri 63130, USA

⁵Department of Biochemistry and Molecular Biology, The University of Iowa, Iowa City, Iowa 52242, USA

^{a)}Author to whom correspondence should be addressed: michael-schnieders@uiowa.edu

ABSTRACT

Computational simulation of biomolecules can provide important insights into protein design, protein-ligand binding interactions, and *ab initio* biomolecular folding, among other applications. Accurate treatment of the solvent environment is essential in such applications, but the use of explicit solvents can add considerable cost. Implicit treatment of solvent effects using a dielectric continuum model is an attractive alternative to explicit solvation since it is able to describe solvation effects without the inclusion of solvent degrees of freedom. Previously, we described the development and parameterization of implicit solvent models for small molecules. Here, we extend the parameterization of the generalized Kirkwood (GK) implicit solvent model for use with biomolecules described by the AMOEBA force field via the addition of corrections to the calculation of effective radii that account for interstitial spaces that arise within biomolecules. These include element-specific pairwise descreening scale factors, a short-range *neck* contribution to describe the solvent-excluded space between pairs of nearby atoms, and finally tanh-based rescaling of the overall descreening integral. We then apply the AMOEBA/GK implicit solvent to a set of ten proteins and achieve an average coordinate root mean square deviation for the experimental structures of 2.0 Å across 500 ns simulations. Overall, the continued development of implicit solvent models will help facilitate the simulation of biomolecules on mechanistically relevant timescales.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0158914>

INTRODUCTION

Biomolecular simulation is a powerful tool that can be used to understand important biological processes such as biomolecular folding^{1,2} and binding interactions^{3,4} while also playing a role in the biomolecular design⁵ and interpretation of experiments.⁶ Often, simulations are limited by the use of accurate but costly explicit descriptions of solvents, making it difficult to achieve biologically relevant timescales. Implicit solvent models that represent solvent effects using a dielectric continuum provide a complementary alternative that eliminates the explicit representation of solvent molecules.⁷ The total implicit solvent potential of the mean force can be divided into polar (electrostatic) and non-polar terms. The polar term can be calculated numerically using Poisson–Boltzmann (PB)

solvers such as the adaptive Poisson–Boltzmann solver (APBS),⁸ ddX,⁹ and PyGBe.^{10,11} Alternatively, the popular generalized Born (GB)^{12–14} model for fixed partial charges or the generalized Kirkwood (GK)¹⁵ model for polarizable multipoles offer efficient analytic approximations.

A foundational component of biomolecular simulations is the selection of a force field. Various GB implicit solvent models for proteins and nucleic acids have been described for fixed charge force fields. An early GB implicit solvent model developed by Hawkins, Cramer, and Truhlar¹⁶ (HCT) presented a pairwise descreening method to calculate effective radii (Fig. 1) analytically based on a van der Waals solute volume. Alternatively, the GBSW (GB simple switching)^{17,18} model, implemented in Chemistry at Harvard Molecular Mechanics (CHARMM),^{19,20} samples atomic density

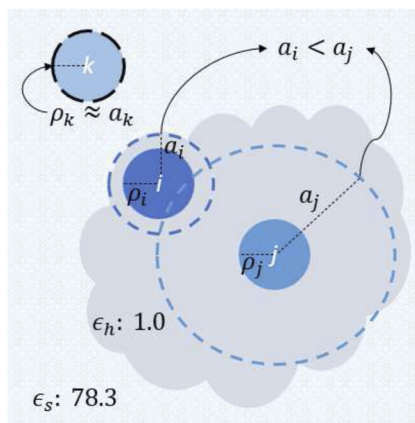


FIG. 1. Pictorial representation of effective radii for an arbitrary globular molecule in implicit water. The effective radii, a_i and a_j , for atoms i and j , respectively, are larger than the intrinsic radii for both atoms (ρ_i and ρ_j). Since atom j is more deeply buried within the molecule than atom i , a_j is larger than a_i . The effective radius for ion k is approximately equivalent to its intrinsic radius ρ_k but could be slightly larger due to descreening by the nearby molecule.

around individual atoms to determine contributions to effective radii and employs a switching function to smooth the dielectric boundary. Ideally, effective radii should be computed using an integral over molecular volume (i.e., Lee–Richards²¹). This motivates the GBMV (GB molecular volume)²² and GBMV2²³ models, also implemented in CHARMM, that leverage a close approximation of molecular volume to calculate effective radii. Further work in AMBER²⁴ led to a model from Onufriev, Bashford, and Case²⁵ (OBC) that added a molecular volume correction in the form of tanh rescaling of effective radii. This correction to effective radii calculated based on pairwise descreening of van der Waals radii helps to account for high dielectric interstitial spaces. An additional molecular volume correction was introduced by Mongan *et al.*²⁶ for the Coulomb field approximation (CFA) and later by Aguilar *et al.*²⁷ for the Grycuk²⁸ solvent field approximation (SFA), which both use an approximate “neck” contribution to describe the solvent-excluded space between pairs of nearby atoms. Both tanh and CFA neck corrections were implemented in AMBER within the GB-neck2 models for proteins²⁹ and nucleic acids,³⁰ and were shown to increase the accuracy of effective radii. A broader description of GB models is presented in a recent review by Onufriev and Case.¹⁴

As an alternative to fixed charge force fields, polarizable force fields have been developed that, in principle, should provide a more transferable description of biomolecular electrostatics. Early work by Maple *et al.* combined a polarizable force field with numerical PB continuum solvation to study protein–ligand interactions.³¹ Similarly, the AMOEBA force field for proteins³² and nucleic acids³³ has been combined with PB,^{8,11} ddCOSMO,³⁴ and GK^{15,35} continuum electrostatic models. More recently, the Drude oscillator’s polarizable force field^{36–50} has been combined with PB electrostatics and used to study pK_a shifts.^{51,52}

Previously, we described a polarizable implicit solvent model for small molecules based on the AMOEBA force field and GK

electrostatics.³⁵ As this model was designed for small molecules, effective radii were computed using an integral over solute van der Waals volume rather than an integral over molecular volume. Here, we describe three modifications to extend the model to biomolecules. These include using element-specific HCT overlap scale factors for pairwise descreening, the addition of a pairwise neck correction to account for solvent-excluded volumes between nearby atoms, and finally a tanh correction to account for three-body (or higher) interstitial spaces. The non-polar model described in the previous work is also extended to biomolecules, which includes a cavitation term based on GaussVol⁵³ and a Weeks–Chandler–Anderson (WCA)⁵⁴ dispersion term. Protein simulations are presented to demonstrate the stability of the model, while future work to support the simulation of nucleic acids and biomolecular complexes is discussed. This AMOEBA/GK model is currently implemented in Force Field X (FFX),⁵⁵ FFX-OpenMM,⁵⁶ and Tinker.^{57,58}

THEORY

The aqueous solvation free energy difference of a molecule (ΔG_{solv}) is the change in free energy between a molecule in a vacuum and water. To formulate an implicit solvent, ΔG_{solv} can be decomposed into three separate path dependent free energy differences⁷ to give

$$\Delta G_{\text{solv}} = \Delta G_{\text{cav}} + \Delta G_{\text{disp}} + \Delta G_{\text{elec}}, \quad (1)$$

where ΔG_{cav} is the unfavorable formation of the molecule-shaped cavity in water and ΔG_{disp} is the favorable addition of solute–water dispersion interactions in the previously formed cavity. Collectively, these first two terms combine to make up the non-polar portion of solvation free energy differences ($\Delta G_{\text{non-polar}} = \Delta G_{\text{cav}} + \Delta G_{\text{disp}}$). Overall, our non-polar term builds on the many advancements and insights contained in the AGBNP family of implicit solvents.^{59–62} The final term ΔG_{elec} accounts for the interaction of solute charge density (e.g., fixed partial atomic charges or polarizable atomic multipoles) with the continuum solvent. The implementation and parameterization of the non-polar term for the current AMOEBA implicit solvent model, as well as the implementation and parameterization of the polar term for small molecules, have been described previously.³⁵ Here, updates to the AMOEBA GK implicit solvent model to facilitate its use with biomolecules are described.

Reference values for ΔG_{elec} in the specific case of the polarizable AMOEBA force field can be determined either by solving the Poisson–Boltzmann equation (PBE) numerically using the APBS multigrid finite-difference solver^{8,63,64} or via a boundary integral approach implemented in PyGBE. While numerical solutions to the PBE can be systematically improved (e.g., by using progressively finer grids or surface meshes), they are generally too expensive to be used for molecular dynamics simulations. For this reason, several approximations have been proposed, including the well-known generalized Born approximation. GB employs a summation over pairwise and self-interactions for fixed atomic partial charge force fields to yield the electrostatic solvation free difference as

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

where ϵ_s is the permittivity of the solvent (78.3 for water), ϵ_h is the permittivity of the homogeneous reference state (1.0 for vacuum), q_i and q_j are the partial charges of atoms i and j , respectively, and a commonly used form of the generalizing function f_{ij} is given by

$$f_{ij} = \sqrt{r_{ij}^2 + a_i a_j} \exp\left(-\frac{r_{ij}^2}{c a_i a_j}\right), \quad (3)$$

where $r_{ij} = \|\mathbf{r}_j - \mathbf{r}_i\|$ is the atomic separation distance in Angstroms, a_i and a_j are the effective Born radii, and c controls the transition from the Born regime ($f_{ij} = 1/a_i$) to the screened Coulomb's law regime ($f_{ij} = 1/r_{ij}$). For most GB implementations, $c = 4$, but here we treat c as a tunable parameter and fix its value to 2.455 as determined previously.¹⁵ GK extends the GB approximation to arbitrary degree multipole moments, which facilitates the use of polarizable atomic multipole solute electrostatics. The GK monopole energy $\Delta G_{GK}^{(q,q)}$ is equivalent to the GB charge-charge term given in Eq. (2). As a second example, the GK dipole energy $\Delta G_{GK}^{(\mu,\mu)}$ is presented beginning from the electrostatic potential at $\mathbf{r}_j$ due to all permanent atomic dipole moments is given by

$$\Phi_{GK}^{(\mu)}(\mathbf{r}_j) = \left[\frac{1}{\epsilon_h} \frac{2(\epsilon_h - \epsilon_s)}{2\epsilon_s + \epsilon_h} \right] \sum_i \frac{r_{\alpha} \mu_{i,\alpha}}{f_{ij}^3}, \quad (4)$$

where μ_i is the permanent atomic dipole moment vector, the subscript α denotes the use of the Einstein summation convention, and $r_{\alpha} = r_{j,\alpha} - r_{i,\alpha}$. The interaction between all permanent dipole moments, including their self-energies, is then given by

$$\begin{aligned} \Delta G_{GK}^{(\mu,\mu)} &= \frac{1}{2} \sum_j \mu_{j,\beta} \nabla_{\beta} \Phi_{GK}^{(\mu)}(\mathbf{r}_j) \\ &= \frac{1}{2} \left[\frac{1}{\epsilon_h} \frac{2(\epsilon_h - \epsilon_s)}{2\epsilon_s + \epsilon_h} \right] \sum_{i,j} \mu_{i,\alpha} \mu_{j,\beta} \left[\frac{3r_{\alpha} r_{\beta} g_{ij}}{f_{ij}^5} + \frac{\delta_{\alpha\beta}}{f_{ij}^3} \right], \end{aligned} \quad (5)$$

where $g_{ij} = \exp(-r_{ij}^2/c a_i a_j)/c - 1$ is the chain rule term, and $\delta_{\alpha\beta}$ is the Kronecker delta. This expression can be simplified to the total permanent atomic dipole self-energy by setting $r_{ij} = 0$ and summing over atomic sites to give

$$\Delta G_{GK}^{(\mu,\text{self})} = \frac{1}{2} \left[\frac{1}{\epsilon_h} \frac{2(\epsilon_h - \epsilon_s)}{2\epsilon_s + \epsilon_h} \right] \sum_i \frac{\mu_{x,i}^2 + \mu_{y,i}^2 + \mu_{z,i}^2}{a_i^3}. \quad (6)$$

Higher-order GK interaction tensors have been described previously¹⁵ and can be generated to a desired order using a tensor recursion⁶⁵ for two Cartesian multipoles in the global frame or after rotation into their quasi-internal (QI) frame.⁶⁶ For the interaction between two multipoles truncated at quadrupole order, the use of the QI frame is ~30% faster for computing the pairwise GK energy, force, and torque (despite the cost of rotating both multipoles from the global frame into the QI frame and the cost of rotating both forces and torques back into the global frame). Reference GK tensor recursion code is available in the "multipole" package of Force Field X (<https://ffx.biochem.uiowa.edu>)⁵⁵ and can be leveraged to create an $O(N \log N)$ treecode-based GK implementation in the future.^{67,68}

For the GB or GK approximation to concord with numerical solutions to the Poisson equation (PE), it has been demonstrated that effective radii should approach being *perfect*.⁶⁹ The reference *perfect* effective Born radius for an atom with a fixed partial charge (q) is defined based on its self-energy, ΔG_{self} , as determined using a numerical solution of the PE with all other atoms in the molecule uncharged,

$$a_i = \frac{1}{2} \left(\frac{1}{\epsilon_s} - \frac{1}{\epsilon_h} \right) \frac{q_i^2}{\Delta G_{\text{self}}}. \quad (7)$$

Although *perfect* effective Born radii enforce that the electrostatic potential at atomic centers match those from the numerical PE solutions, neither the electric field nor its gradient are guaranteed to be correct (i.e., permanent dipole and/or quadrupole self-energy contributions computed using *perfect* effective Born radii deviate from their reference numerical PE values). As an alternative, the contribution of higher order atomic multipole moments to the self-energy can be included to calculate a *perfect* effective Kirkwood radius using the following equation,⁷⁰ which was originally derived by Kirkwood using spherical harmonics and later converted by Applequist to a traceless Cartesian tensor form:⁷¹

$$\begin{aligned} \Delta G_{\text{self}} &= \frac{1}{2} \left[\epsilon_0 \frac{q_i^2}{a_i} + \epsilon_1 \frac{\mu_{x,i}^2 + \mu_{y,i}^2 + \mu_{z,i}^2}{a_i^3} \right. \\ &\quad \left. + \epsilon_2 \frac{2}{3} \frac{\Theta_{xx,i}^2 + \Theta_{yy,i}^2 + \Theta_{zz,i}^2 + 2(\Theta_{xy,i}^2 + \Theta_{xz,i}^2 + \Theta_{yz,i}^2)}{a_i^5} \right], \end{aligned} \quad (8)$$

where Θ is the traceless permanent quadrupole, and the permittivity function ϵ_n for a multipole moment of order n is given by^{70,72}

$$\epsilon_n = \frac{1}{\epsilon_h} \frac{(n+1)(\epsilon_h - \epsilon_s)}{(n+1)\epsilon_s + n\epsilon_h}. \quad (9)$$

The *perfect* effective Kirkwood radius a_i can then be determined using Eq. (8) and a simple numerical search. Note that the right-hand side of Eq. (8) neglects polarization energy (i.e., the interaction of an induced dipole at site i with the reaction field of its permanent dipole), and the computed ΔG_{self} is based on the input of an AMOEBA permanent atomic multipole. In practice, *perfect* effective Born radii and *perfect* effective Kirkwood radii agree to be within ~2% on average (see Table I), and both represent the degree of burial of an atom within a molecule. Mean *perfect* effective Born radii and *perfect* effective Kirkwood radii for several biomolecules are shown in Table I, along with the regression slope, correlation, and MUE of *perfect* Kirkwood radii relative to *perfect* Born radii. The full regression of radii for ubiquitin (PDB ID: 1UBQ) is shown in the supplementary material, Fig. S1, as an example.

Element-specific overlap scale factors

For the calculation of effective radii, the GK implicit solvent model combines the analytic HCT pairwise descreening approximation¹⁶ with the SFA proposed by Gryckuk.²⁸ As a part of this approximation, a unitless scale factor was previously set at 0.72 to account for the atomic overlaps that would otherwise lead to overestimated effective radii.³⁵ While a single scale factor worked well, it was straightforward to achieve a modest improvement in accuracy

TABLE I. Mean *perfect* Born radii and *perfect* Kirkwood radii for each tested molecule. On average, *perfect* Kirkwood radii are slightly smaller than *perfect* Born radii based on AMOEBA permanent multipoles. The slope, correlation, and MUE are presented for each molecule by comparing *perfect* Kirkwood radii to *perfect* Born radii. An example regression for 1UBQ is given in the supplementary material, Fig. S1.

Molecule	Average <i>perfect</i> Born radius for all atoms (Å)	Average <i>perfect</i> Kirkwood radius for all atoms (Å)	Slope (Born vs Kirkwood)	R^2	MUE (Å)
1MIS	2.93	2.86	0.982	0.992	0.209
2JXQ	2.91	2.86	0.988	0.992	0.203
1F5G	2.93	2.88	0.986	0.992	0.206
2L8F	3.03	2.97	0.990	0.988	0.244
1ZIH	3.12	3.07	0.990	0.990	0.237
1SZY	2.94	2.88	0.987	0.992	0.202
2KOC	2.87	2.82	0.989	0.991	0.205
1D20	2.88	2.86	0.997	0.988	0.244
2HKB	2.91	2.88	0.991	0.990	0.232
1BPI	3.48	3.41	0.987	0.989	0.274
1L2Y	2.96	2.89	0.984	0.989	0.240
1UBQ	4.14	4.08	0.995	0.992	0.302
1UCS	4.00	3.92	0.988	0.990	0.315
1VII	3.12	3.05	0.984	0.989	0.248
1WM3	3.94	3.87	0.990	0.992	0.286
2OED	3.53	3.45	0.980	0.990	0.267
2PPN	4.16	4.09	0.991	0.992	0.302
7SKW	4.57	4.51	0.994	0.991	0.338
Average	3.36	3.30	0.988	0.990	0.253

using element specific overlap scale factors. The current implementation of the GK pairwise descreening term with element specific scale factors is given by

$$I^{vdw}(r_{ij}) = \frac{\pi}{12} \left(\frac{3(r_{ij}^2 - (S_{HCT,j} * \rho_j)^2) + 6u^2 - 8ur_{ij}}{u^4 r_{ij}} - \frac{3(r_{ij}^2 - (S_{HCT,j} * \rho_j)^2) + 6l^2 - 8lr_{ij}}{l^4 r_{ij}} \right), \quad (10)$$

where ρ_j is the radius of atom j used for descreening, $S_{HCT,j}$ is the element-specific scaling factor for atom j , and u and l are the upper and lower integration bounds, determined based on the overlap of atoms i and j . A detailed description of how to determine the upper and lower integration bounds has been described previously.^{16,35}

Pairwise neck interstitial space correction

Descreening for the AMOEBA small molecule implicit solvent was based on using a van der Waals definition of solute volume. This approximation is not appropriate for large biomolecules where interstitial spaces become increasingly important, which motivates the use of the more physically realistic Lee-Richards molecular volume. In particular, the use of a van der Waals volume leads to an underestimation of the effective radii for biomolecules by failing to account for descreening due to interstitial spaces that are too small to accommodate water molecules. Although too small to

accommodate an explicit water molecule, these interstitial spaces are nonetheless “filled” by continuum water, leading to artificially favorable electrostatic hydration. A more accurate descreening integral must account for interstitial spaces and properly exclude continuum water. The concept of “neck” regions between pairs of nearby atoms, as shown in Fig. 2, was first proposed by Mongan and co-workers for the Coulomb field approximation (CFA) $|r|^{-4}$ integral²⁶ and later refined by Aguilar and co-workers for the solvent field approximation (SFA) $|r|^{-6}$ integral.²⁷ The functional form of the latter is given by

$$I^{\text{neck}}(r_{ij}) = \frac{4\pi}{3} S_{\text{neck},ij} * A_{ij} (r_{ij} - B_{ij})^4 (\rho_i + \rho_j + 2\rho_w - r_{ij})^4, \quad (11)$$

where ρ_i and ρ_j are the descreening radii of atoms i and j , respectively; ρ_w is the radius of water (1.4 Å); and r_{ij} is the atomic separation distance.

The A_{ij} and B_{ij} constants were originally determined using benchmark values from a numerically exact method of calculating effective Born radii called NSR6, which has been described previously.²⁷ In this work, values of A_{ij} and B_{ij} were calculated for an expanded set of descreening radii using *perfect* effective radii from APBS calculations. The procedure for determining A_{ij} and B_{ij} was otherwise analogous; for pairs of atoms, APBS was used to determine the value of the neck integral at various separation distances. The separation distance at which the value of the neck integral is at its maximum ($neck^{max}$) was recorded as r_{ij}^{max} for that pair of radii. The

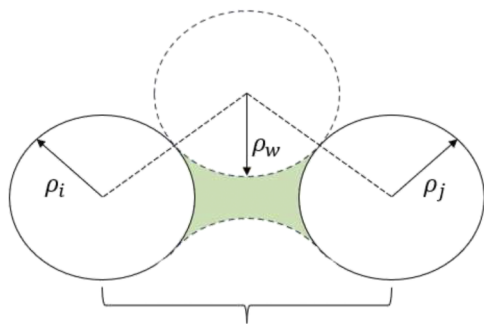


FIG. 2. Shown is the neck region (shaded) between two atoms. A neck is only formed when the atoms are close enough to exclude water, which is represented here by a sphere with radius ρ_w .

value of B_{ij} was then calculated as $B_{ij} = 2r_{ij}^{\max} - (\rho_i + \rho_j + 2\rho_w)$ and A_{ij} was calculated such that $I_{ij}^{\text{neck}}(r_{ij}^{\max}) = \text{neck}^{\max}$. A slight change to the determination of A_{ij} values in this work was to explicitly include the $4\pi/3$ constant in the neck integral equation instead of including it in the A_{ij} values—this was done to facilitate consistency in the components of the descreening integral. A full tabulation of these updated A_{ij} and B_{ij} constants is available in the supplementary material, Tables S1 and S2.

For a single pair of atoms, the neck between them perfectly describes the correction from the van der Waals volume to the Lee–Richards molecule volume. For more than two atoms, however, neck regions can overlap and lead to an overestimation of the interstitial volume. For this reason, a scale factor (S_{neck}) is introduced to prevent over-counting neck overlaps, which is analogous to the HCT scale factors that account for overlap during pairwise descreening. Neck contributions to the molecular volume are only calculated for pairs of atoms that are close enough to exclude water between them such that their separation distance (r_{ij}) satisfies the following criterion:

$$r_{ij} \leq \rho_i + \rho_j + 2\rho_w. \quad (12)$$

To calculate forces (e.g., for optimization or molecular dynamics) the neck correction must be differentiable with respect to the separation distance. This derivative is given by

$$\begin{aligned} \frac{\partial I_{ij}^{\text{neck}}(r_{ij}^{\max})}{\partial r_{ij}} &= \frac{16\pi}{3} * (S_{\text{neck},ij} * A_{ij}(r_{ij} - B_{ij})^3 \\ &\quad \times (\rho_i + \rho_j + 2\rho_w - r_{ij})^4 \\ &\quad - S_{\text{neck},ij} * A_{ij}(r_{ij} - B_{ij})^4 (\rho_i + \rho_j + 2\rho_w - r_{ij})^3). \end{aligned} \quad (13)$$

In previous implementations of the neck correction, a single S_{neck} scaling factor was used in all cases. In this work, we propose a modification to the neck scaling factor based on the number of heavy atoms bound to a particular atom of interest. If no heavy atoms are bound to atom i , then $S_{\text{neck},i} = 1.0$. For all other cases, the scaling factor for atom i is calculated based on the relationship,

$$S_{\text{neck},i} = S_{\text{neck}} * \frac{5.0 - n_{\text{heavy}}}{4.0}, \quad (14)$$

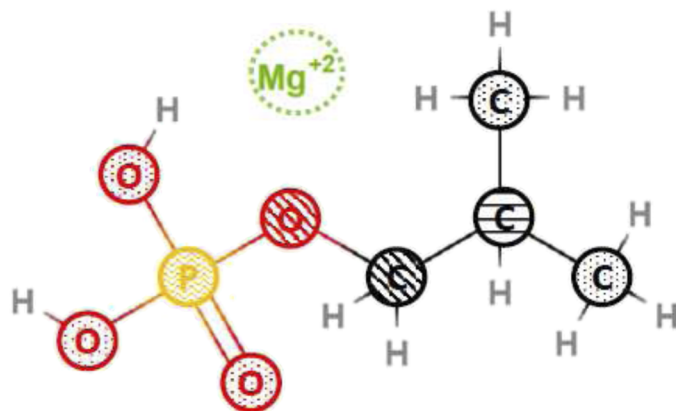
where n_{heavy} is the number of heavy atoms bound to the atom of interest, and S_{neck} is the maximum fit scale factor. A pictorial representation of this S_{neck} scheme is shown in Fig. 3.

This modification to the treatment of interstitial space necks preserves accuracy for free ions (Fig. 4) and results in atoms with fewer bound heavy atoms forming more significant necks than atoms with more bound heavy atoms. Finally, the following combining rule is used to weight the chemical environment of both atoms that form the neck,

$$S_{\text{neck},ij} = (S_{\text{neck},i} + S_{\text{neck},j})/2. \quad (15)$$

Hyperbolic tangent interstitial space correction

The pairwise neck correction, described above, is helpful in accounting for short-range [Eq. (12)] underestimation of molecular volume but does not account for three-body (or higher) effects. A



0 Bound Heavy Atoms	$S_{\text{neck},i} = 1.00$
1 Bound Heavy Atom	$S_{\text{neck},i} = S_{\text{neck}}$
2 Bound Heavy Atoms	$S_{\text{neck},i} = \frac{3 * S_{\text{neck}}}{4}$
3 Bound Heavy Atoms	$S_{\text{neck},i} = \frac{S_{\text{neck}}}{2}$
4 Bound Heavy Atoms	$S_{\text{neck},i} = \frac{S_{\text{neck}}}{4}$

FIG. 3. Shown are the bonding aware neck scaling scheme and associated equations used to calculate an individual atomic $S_{\text{neck},i}$ scale factor. The scaling factor is reduced as the number of heavy atoms bound to the atom of interest increases. If the atom of interest has no bound heavy atoms, the scale factor is set to 1.0.

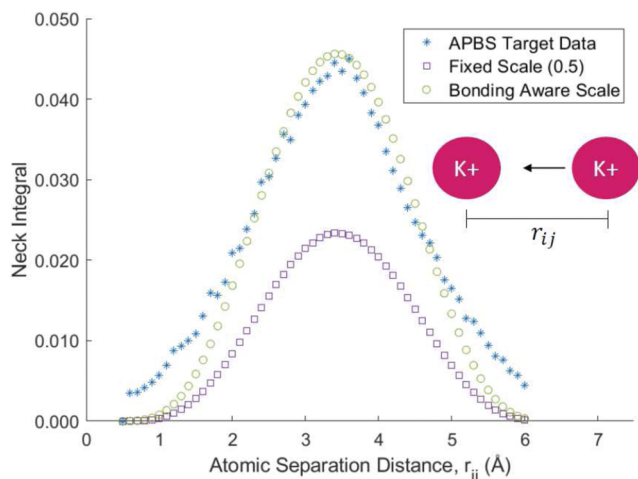


FIG. 4. Demonstration of the bonding aware neck scaling scheme for two ions in the AMOEBA GK implicit solvent. Using a single, fixed scaling factor leads to underestimation of the neck integral value at all separation distances for pairs of ions, while chemically aware scaling concurs with reference PB calculations.

function based on the hyperbolic tangent ($\tanh$) is used to smoothly scale up effective radii, increasing the radii of deeply buried atoms more than the radii of atoms closer to the surface. This type of $\tanh$ rescaling function has been used previously as part of both CFA $|r|^{-4}$ interstitial space corrections^{26,69} and, more recently, SFA $|r|^{-6}$ interstitial space corrections.²⁷ The $\tanh$ correction used in the current model follows the latter, where the maximum effective radius is capped at 30 Å but does not consider the electrostatic size of the solute.²⁷ The $\tanh$ rescaling function and associated component functions are presented below,

$$c_i = \frac{4\pi}{3} \left(\frac{1}{\rho_i^3} - \frac{1}{30^3} \right), \quad (16)$$

$$\Psi_i = \sum_{i \neq j} I^{\text{vdw}}(r_{ij}) + \sum_{i \neq j} I^{\text{neck}}(r_{ij}), \quad (17)$$

$$\frac{1}{a_i} = \frac{3}{4\pi} \left(\frac{4\pi}{3} \rho_i^{-3} - c_i \tanh \left(\beta_0 \Psi_i \rho_i^3 - \beta_1 (\Psi_i \rho_i^3)^2 + \beta_2 (\Psi_i \rho_i^3)^3 \right) \right)^{1/3}. \quad (18)$$

Here a_i is the effective radius of atom i ; β_0, β_1 , and β_2 are tunable parameters; $I^{\text{vdw}}(r_{ij})$ is the $|r|^{-6}$ integral over all van der Waals spheres in the solute [Eq. (10)]; and $I^{\text{neck}}(r_{ij})$ are short-range pairwise neck contributions [Eq. (11)]. The $\tanh$ correction to the effective radius and its derivative is given by

$$\text{scale}(\Psi_i) = c_i \tanh \left(\beta_0 \Psi_i \rho_i^3 - \beta_1 (\Psi_i \rho_i^3)^2 + \beta_2 (\Psi_i \rho_i^3)^3 \right) \quad (19)$$

and

$$\frac{\partial \text{scale}(\Psi_i)}{\partial \Psi_i} = c_i (\beta_0 \rho_i^3 - 2\beta_1 \Psi_i \rho_i^6 + 3\beta_2 \Psi_i^2 \rho_i^9) \times \left(1 - \tanh \left(\beta_0 \Psi_i \rho_i^3 - \beta_1 (\Psi_i \rho_i^3)^2 + \beta_2 (\Psi_i \rho_i^3)^3 \right) \right). \quad (20)$$

PARAMETERIZATION

Element-specific scale factors

Element-specific scaling factors were determined using a limited memory BFGS optimizer and five different target molecules—two proteins, two RNAs, and one DNA—that were chosen from the set of biomolecules used to validate the small molecule implicit solvent models.³⁵ The optimizer target function is given by

$$E(\mathbf{P}) = W_{MUE} \sum_{i=1}^n (\Delta G_{i,\text{self}}^{\text{PB}} - \Delta G_{i,\text{self}}^{\text{GK}})^2 + W_{MSE} \left(\sum_{i=1}^n \Delta G_{i,\text{self}}^{\text{PB}} - \sum_{i=1}^n \Delta G_{i,\text{self}}^{\text{GK}} \right)^2 + W_{\text{Regularization}} \sum_{i=1}^{N_{\text{elements}}} \left(S_{\text{HCT}}^{\text{element}} - 0.72 \right)^2, \quad (21)$$

where $W_{MUE} = 1.0$, $W_{MSE} = 10.0$, and $W_{\text{Regularization}} = 1.0 \times 10^4$. Here, $\Delta G_{i,\text{self}}$ is the self-energy for the atom i calculated using either PB or GK for n atoms and $S_{\text{HCT}}^{\text{element}}$ is the element-specific scale factor for each element (C, N, O, P, and S), where $N_{\text{elements}} = 5$. The HCT scale factors were optimized for each molecule individually and then averaged. Benchmark permanent AMOEBA electrostatic solvation energy values were calculated using APBS based on a van der Waals definition of the solute volume. The decision to use permanent self-energy values was motivated by the expense of using APBS to compute self-consistent reaction fields and by the relatively smaller contribution of self-polarization. Van der Waals radii were used for consistency between the PB and GK electrostatics models and to prevent HCT scale factors from implicitly accounting for interstitial spaces. The benchmark self-energies, $\Delta G_{i,\text{self}}^{\text{PB}}$, used in the HCT scale factor optimizer were determined using monopoles without considering higher order multipole moments. These self-energies are consistent with *perfect* effective Born radii and promote transferability of the final $S_{\text{HCT}}^{\text{element}}$ to other force fields, including those based on fixed partial charges.

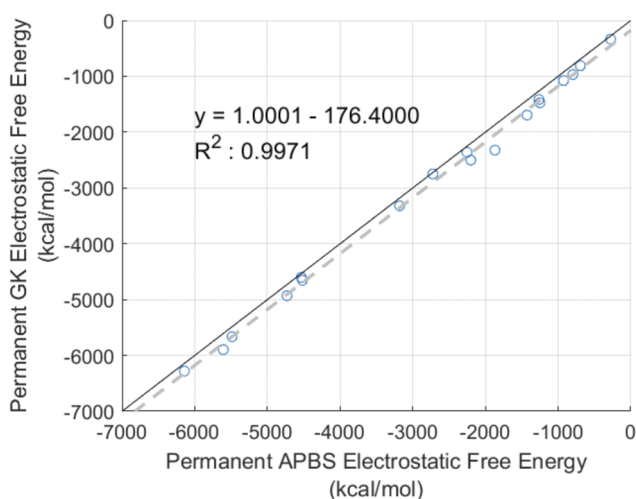
Starting from the initial small molecule scale factor (0.72), element-specific scaling factors were fit for C, N, O, P, and S. Due to the high degree of overlap with their bound heavy atom, the choice was made to exclude hydrogen atoms from contributing to descreening for the AMOEBA GK implicit solvent. For this reason, no scale factor was fit for hydrogen atoms (i.e., the HCT scale factor for hydrogen atoms is 0). The final element-specific scaling factors ($S_{\text{HCT}}^{\text{element}}$) are shown in Table II.

As the $S_{\text{HCT}}^{\text{element}}$ scale factors were fit using a van der Waals description of solute volume; only the *neck* and *tanh* corrections account for molecular volume based descreening of biomolecules.²¹ The final scale factors agree with chemical intuition regarding heavy atom overlaps. Carbon atoms can form four bonds, often with other heavy atoms, necessitating a smaller scale factor than the base value

TABLE II. Element-specific S_{HCT} scale factors optimized for individual protein (1BPI and 1UCS), RNA (1MIS and 1ZIH), and DNA (1D20) molecules. The final scale factors are averages.

Element and Bondi ⁷³ radius (Å)	1BPI	1UCS	1MIS	1ZIH	1D20	Final
C (1.70)	0.7151	0.7294	0.6694	0.6975	0.6634	0.6950
N (1.55)	0.8348	0.7614	0.7659	0.7365	0.7377	0.7673
O (1.50)	0.7635	0.7785	0.8098	0.8048	0.8261	0.7965
P (1.80)	...	...	0.6173	0.6300	0.5878	0.6117
S (1.80)	0.7214	0.7194	...	...	...	0.7204

of 0.72. Phosphorous atoms can form four bonds as well (e.g., to oxygen atoms in the tested nucleic acids) and are larger than carbon atoms, which explains why phosphorous has the smallest HCT scale. Conversely, nitrogen and oxygen atoms are smaller than carbon and phosphorous while also generally forming fewer bonds with other heavy atoms. This is consistent with fewer overlaps and explains the relative increase in those two scale factors during fitting. In the tested protein systems, sulfur atoms form either one or two bonds with heavy atoms. Due to sulfur atoms being larger than nitrogen or oxygen but forming fewer bonds than carbon or phosphorous, an intermediate amount of overlap is expected, which is consistent with the sulfur HCT scale remaining close to the base value. Final S_{HCT} scale factors were used to compute the electrostatic portion of solvation free energy differences for all 18 molecules used to validate the small molecule implicit solvent models.³⁵ When compared with APBS results, GK energy differences produced a slope of 1.0001 and an R-squared of 0.9971 (Fig. 5). This supports the conclusion that the relatively simple HCT pairwise descreening approach is reliable for integration over van der Waals solute volumes.

**FIG. 5.** Comparison of permanent electrostatic energies for biomolecules calculated in APBS and GK using a van der Waals definition of the solute volume. GK values used element specific HCT scaling factors. The dashed gray line is the best fit regression line with slope = 1.0001 and $R^2 = 0.9971$. The solid black is $x = y$ to guide the eye.

Neck and tanh interstitial space corrections

To facilitate the use of the GK implicit solvent model with mixed protein/nucleic acid simulations, a single set of $\tanh \beta$ parameters was fit for all target biomolecules. During initial MD testing with the full biomolecule test set, proteins and nucleic acids showed distinct sensitivities to the magnitude of the neck scaling factor. The original goal of fitting a single neck scaling factor for use with both proteins and nucleic acids led to a balancing of errors whereby folded protein electrostatic energies were too negative and folded nucleic acid electrostatic energies were too positive (compared to PB reference values). For this reason, neck scale factors for proteins and nucleic acids will be fitted to separate optimal values.

The $\{\beta_0, \beta_1, \beta_2\}$ parameters were initially optimized simultaneously using a genetic algorithm. Each run of the genetic algorithm included 1000 generations of 500 individuals, with the top 20% of individuals being carried over directly to the next generation and a mutation rate of 0.3. Permanent self-energies ($\Delta G_i^{Self, Perfect}$), permanent electrostatic energies ($\Delta G_i^{Elec, Perfect}$), and *perfect* effective Kirkwood radii ($\bar{R}_{perfect}$) were calculated using APBS and used as target data for optimization according to the following objective function:

$$\begin{aligned}
 E(\mathbf{P}) = & W_{MUE} \left(\sum_{i=1}^n \left| \Delta G_i^{Elec, GK} - \Delta G_i^{Elec, Perfect} \right|^2 \right. \\
 & \left. + \sum_{i=1}^n \left| \Delta G_i^{Self, GK} - \Delta G_i^{Self, Perfect} \right|^2 \right) \\
 & + W_{MSE} \left(\sum_{i=1}^n \Delta G_i^{Elec, GK} - \sum_{i=1}^n \Delta G_i^{Elec, Perfect} \right)^2 \\
 & + W_{Rad} \left(\sum_{i=1}^n (\bar{R}_{perfect} - \bar{R}_{GK})^2 \right), \quad (22)
 \end{aligned}$$

where $W_{MUE} = 0.001$, $W_{MSE} = 1.0$, and $W_{Rad} = 1.0$. The parameters for each new (non-mutant) individual were selected randomly from uniform distributions across the following ranges: $\beta_0 \in \{0.5000, 1.5000\}$, $\beta_1 \in \{0.1000, 0.4000\}$, and $\beta_2 \in \{0.0004, 0.2000\}$. The permanent energies and *perfect* effective Kirkwood radii used as benchmarks were from a set of nine proteins (1BPI, 1L2Y, 1UBQ, 1UCS, 1VII, 1WM3, 2OED, 2PPN, 7SKW) and nine nucleic acids (1MIS, 2JXQ, 1F5G, 2L8F, 1SZY, 1ZIH, 2KOC, 1D20, 2HKB). The fitting of $\tanh$ parameters is a multiple-minima problem,^{25,27,30} and for this reason, a local optimization approach (e.g., using an L-BFGS optimizer) will not explore the parameter space effectively. Instead,

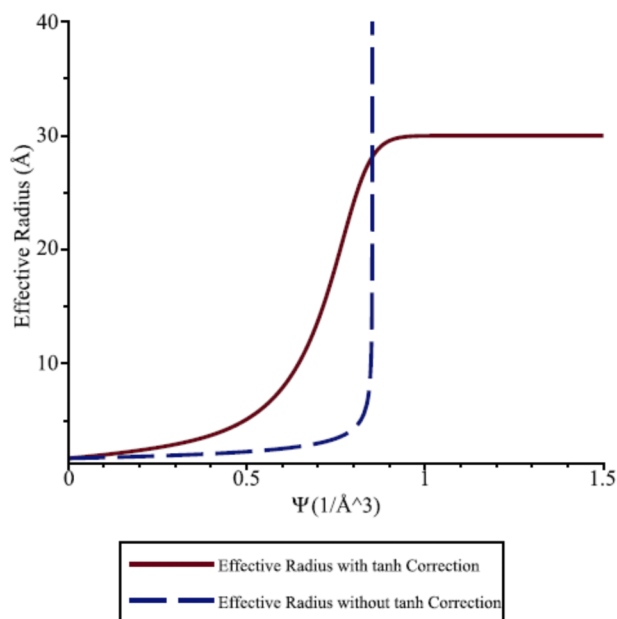


FIG. 6. Effective radii for a molecule with a 1.7 Å base radius (carbon atom) without rescaling (dashed line) and with the final tanh rescaling function based on $\beta_0 = 0.9563$, $\beta_1 = 0.2578$, and $\beta_2 = 0.0810$ (solid line) across a range of input volume integrals (Ψ).

the genetic algorithm was used to control parameter ranges based on results from prior work^{25,29,30} and trial and error.

The candidate parameter sets produced by the optimization runs were used to calculate GK permanent electrostatic energies for the biomolecule target set. Additionally, the GK effective radii were plotted against the tanh input (Ψ) to check the shape of the tanh function for given β parameter sets. Any tanh function that did not have a positive first derivative across the full functional range was eliminated. Output parameters from the genetic algorithm were slightly adjusted manually to improve the electrostatic energy regression for all tested biomolecules, resulting in the following parameter set: $\{\beta_0 = 0.9563, \beta_1 = 0.2578, \beta_2 = 0.0810\}$, which is plotted in Fig. 6.

With the tanh β parameters fixed, the final S_{neck} scaling factor for proteins was then determined using progressively finer S_{neck} scans. The bonding awareness scheme described above was used for all scans. The final protein scale factor ($S_{neck,pr} = 0.1350$) helped improve permanent self-energies relative to only using a tanh correction (Fig. 7). The final optimized parameters are given in Table III.

Effective radii calculated in GK using the protein parameter set and thioredoxin (2TRX), which was not used for fitting, are plotted against perfect effective Kirkwood radii (calculated using APBS) in Fig. 8.

In Fig. 9, total electrostatic hydration free energy differences calculated using APBS with a molecular surface are plotted against total electrostatic hydration free energy differences calculated using GK with interstitial space corrections for all target molecules.

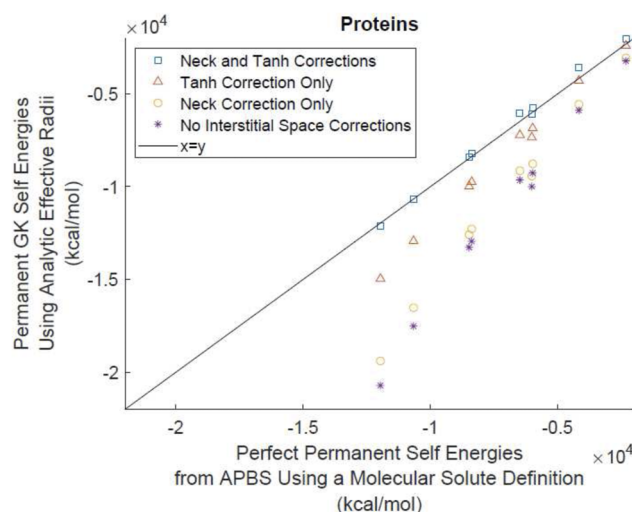


FIG. 7. Comparison of permanent self-energies with perfect effective Kirkwood radii and GK fit radii for proteins. Energies calculated with only the tanh correction have a slope of 1.192 and a R^2 of 0.997; energies calculated with the full correction (neck and tanh) have an improved slope of 0.987 and a R^2 of 0.998. The solid black line is $x = y$ to guide the eye.

TABLE III. The implicit solvent parameters are used to correct from a van der Waals volume integral to a molecular volume integral with interstitial spaces. The neck scale factor is applied in a bonding aware manner, as described in Fig. 3.

Parameter	Value
β_0	0.9563
β_1	0.2578
β_2	0.0810
$S_{neck, protein}$	0.1350

Tuning based on molecular dynamics trajectories

The original base radii for the GK implicit solvent model were previously fitted using small molecule solvation free energy differences.³⁵ During initial molecular dynamics tests on biomolecules, in some cases, overcounting in the pairwise descreening integrals was observed. This was traced to the SFA $|r|^{-6}$ descreening integral evaluating to its largest values for small separation distances (i.e., the HCT overlap scale factors are appropriate on average but can be too large for some overlaps at very short ranges). To alleviate this, a small descreening offset of 0.3 Å was added to push the beginning of the descreening integral away from the atomic center. It was also observed that repeated backbone atoms tended to favor intramolecular interactions (such as hydrogen bonding) over interactions with the GK continuum. This is in part due to the lack of hydrogen bonding within the fitting set of small molecule solvation free energy differences. Slight alterations to selected atomic base radii for proteins helped alleviate the incorrect preference of certain groups to form intramolecular hydrogen bonds with backbone groups in place of interacting with implicit water. Radii for protein carbonyl carbon and oxygen atoms, asparagine and glutamine amide nitrogen atoms,

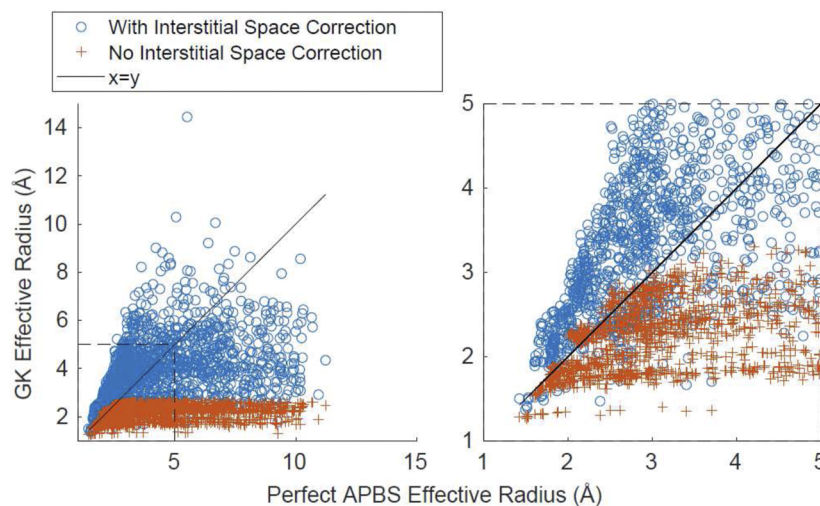


FIG. 8. Comparison of 2TRX effective radii calculated in GK without interstitial space corrections (orange pluses) and with interstitial space corrections (blue circles) to perfect effective Kirkwood radii calculated in APBS. Fit GK base radii are used for both series. The first plot (left) shows all effective radii for 2TRX (thioredoxin), while the second plot (right) shows the range of effective radii from 1 to 5 Å. The solid black line is $x = y$ to guide the eye.

and lysine and arginine HN atoms were modified slightly—a tabulation of the updated GK base radii is available in the supplementary material, Table S3.

A known limitation of the current implicit solvent model involves the use of GaussVol to determine the surface area used in cavitation free energy calculations. GaussVol is comparable to the

more general algorithms by Connolly^{74,75} when calculating a van der Waals surface area, but GaussVol cannot be used to compute molecular surface area at this time. Due to the Connolly algorithms not being available on graphics processing units (GPUs) (e.g., neither in Tinker nor OpenMM), the use of the molecular surface area is not currently feasible. For these reasons, an inflated van der Waals surface area computed using GaussVol via scaling up atomic radii by 15% was selected as the best available option (i.e., to approximately remove interstitial spaces). This modest increase in atomic radii preserved simulation efficiency, while even small additional increases (e.g., to 20% or 25%) reduced simulation speed by almost a factor of 2 (due to the nonlinear increase in GaussVol atomic overlaps as a function of atomic radii). Future work to extend the GaussVol approach to efficiently handle molecular surface areas will benefit both fixed charge and polarizable implicit solvents.

RESULTS

The parameterization of the AMOEBA/GK biomolecular implicit solvent model was designed to enable simulations of both proteins and nucleic acids. The results for proteins are presented here, while those for nucleic acids will be described in a later contribution. MD simulations were performed for the set of 10 proteins previously studied during the development of the AMOEBA protein force field.³² The 7SKW structure of lysozyme was used in this work in place of the 6LYT structure used previously due to the improved resolution of 7SKW across the same lysozyme sequence. Final implicit solvent parameters (Table III) and updated GK base radii (the supplementary material, Table S3) were used for all simulations. Each molecule was simulated continuously for 500 ns. Explicit neutralizing chloride ions (nine total ions) were added to the lysozyme simulation (7SKW) and restrained using flat-bottom potentials. These restraints help maintain the neutralizing

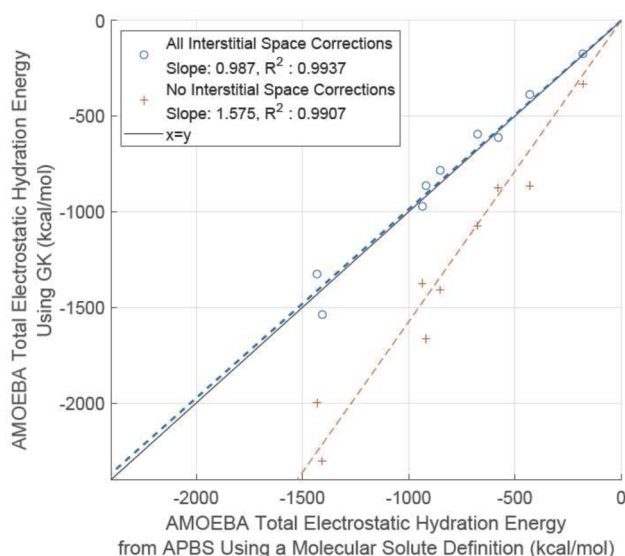


FIG. 9. Electrostatic hydration energy of target proteins was calculated using APBS and GK, including contributions from permanent multipoles and induced dipoles determined via a self-consistent reaction field. APBS used a molecular surface based on fit GK base radii. GK also used fit radii, with effective Born radii computed with (Slope: 0.987, R^2 : 0.9937) and without (Slope: 1.575, R^2 : 0.9907) interstitial space corrections.

TABLE IV. Average backbone (BB) heavy atom RMSD values across 500 ns MD trajectories for protein molecules in Angstroms.

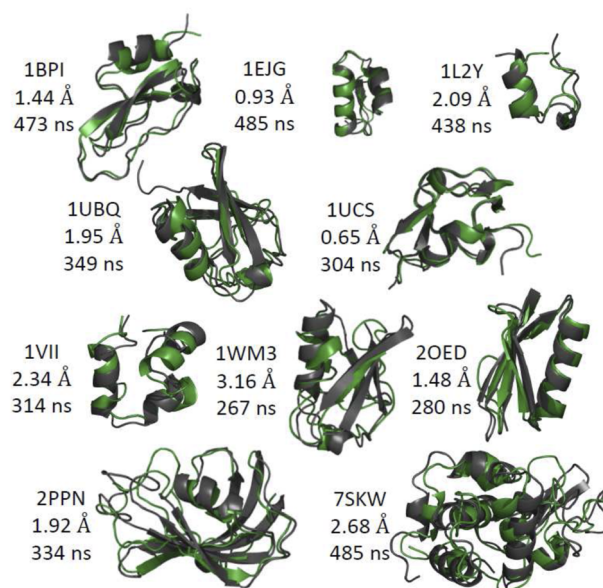
PDB ID	No. Residues	Formal charge	Average BBRMSD with all residues (Å)		Average non-terminal BBRMSD (Å)	
			30 ns	500 ns	30 ns	500 ns
1BPI	58	6.0	1.73	1.69	1.37	1.57
1EJG	46	0.0	0.78	1.13	0.76	1.09
1L2Y	20	1.0	1.99	2.86	1.24	2.10
1UBQ	76	1.0	2.98	3.97	1.48	1.94
1UCS	64	0.0	1.22	1.76	0.83	1.12
1VII	76	2.0	2.18	2.76	2.01	2.56
1WM3	88	1.0	1.51	3.23	1.48	3.08
2OED	56	-2.0	1.80	2.03	1.78	2.02
2PPN	107	4.0	2.89	3.47	1.65 ^a	2.05 ^a
7SKW	129	9.0	2.53	2.86	2.44	2.73
Average			1.96	2.58	1.50	2.03

^aIn addition, excludes flexible loop (residues 82–96).

ion cloud around the solute and prevent diffusion toward entropically favored states. The restraints enforced a maximum separation distance of 45.0 Å from the center of mass of the lysozyme protein, while no minimum distance penalty was used. Explicit ions were energy-minimized to an rms gradient of 1.0 kcal/(mol Å) and then equilibrated for 1 ns at 100 K, 1 ns at 200 K, and 1 ns at 300 K with the position of the biomolecule fixed. Each protein was energy-minimized to an rms gradient of 1.0 kcal/(mol Å), then equilibrated for 1 ns at 100 K, 1 ns at 200 K, and 1 ns at 300 K. During equilibration, protein C-alpha atoms were fixed to promote relaxation of side chains before allowing the full biomolecule to move. The production runs completed 500 ns of sampling at 298.15 K with self-consistent field (SCF) convergence to 1.0×10^{-6} rms D. A 2 fs Langevin multiple time step integrator was used for all simulations, along with mass repartitioning from heavy atoms to their bound hydrogen atoms. All production simulations were run on either NVIDIA GeForce RTX 2080TI or NVIDIA A10 Tensor Core GPUs. The wall clock time to generate 500 ns for the smallest tested system (1L2Y, trp-cage) was 428.9 h (~17.8 days), and that for the largest tested system (7SKW, lysozyme) was 1076.3 h (~44.8 days).

Output MD trajectories were compared to the experimental structure for each protein system. Where experimental structures consisted of NMR ensembles, the first structure in the ensemble was selected for comparison. Coordinate root mean square deviations (RMSDs) for proteins are reported in Table IV. All RMSDs reported are for peptide backbone heavy atoms and were calculated in FFX using the Superpose utility script. Average backbone heavy atom RMSDs for proteins were 2.58 Å for all residues and 2.03 Å for non-terminal residues (Table IV). Average RMSDs after 30 ns of simulation are also available in Table IV in order to directly compare them to previous protein simulations in explicit AMOEBA water.³² The average backbone RMSD for explicit AMOEBA water simulations after 30 ns as reported in the original paper was 1.33 Å,³² while the average backbone RMSD for implicit AMOEBA water simulations was 1.96 Å (Table IV). Part of this difference is likely due to the reduced viscosity of continuum solvents, which results in faster kinetics. Trajectory snapshots were clustered based on non-terminal

backbone heavy atom RMSD into ten clusters, and representative structures from the largest clusters are presented superposed with the base experimental structures in Fig. 10. Representative structures were the minimum RMSD structures from the second half of the trajectory (250 ns and beyond) in the largest cluster. RMSD trajectories across the full 500 ns simulation time are presented in Fig. 11.

**FIG. 10.** Superposition of the deposited x-ray crystallography or NMR structure (gray) with the lowest-RMSD structure from the largest cluster (green). The time step of the snapshot in the 500 ns trajectory and its RMSD to the experimental structure are displayed. All representative snapshots were taken from the second half of the trajectories.

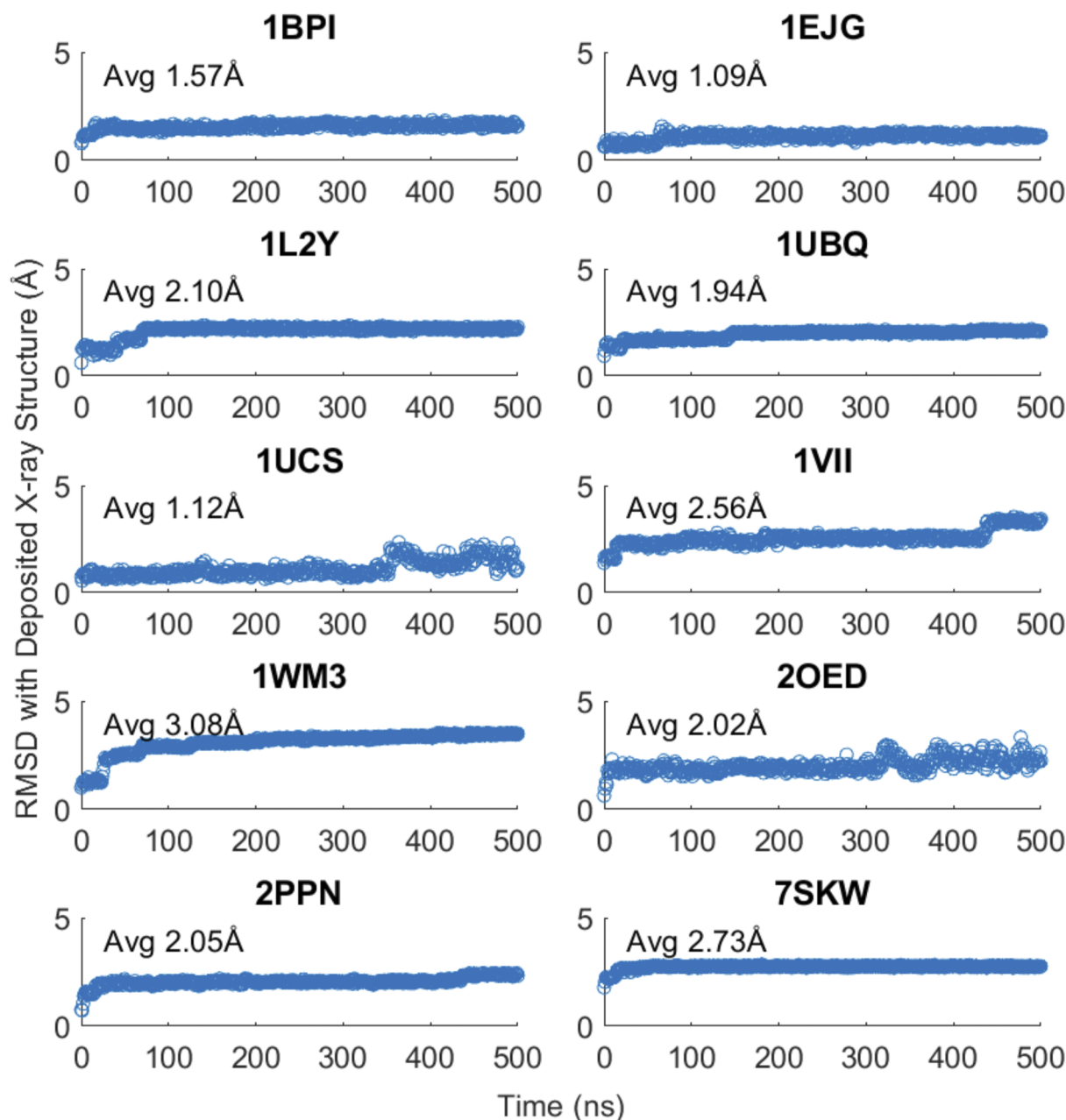


FIG. 11. Protein non-terminal backbone RMSDs to the experimental x-ray crystallography or NMR structure across 500 ns production trajectories (Å).

Testing of the GBNeck2 implicit solvent model²⁹ included simulations of the trp-cage protein, which was also simulated here (1L2Y). A histogram of RMSD probability across two trajectories at 300 K was presented, analogous to the 1L2Y histogram (second row, left) in Fig. 12. GBNeck2 trajectories, reported in the original paper in the supplementary material, (Fig. S9), are for 160 ns trajectories at 300 K with enhanced sampling from replica exchange

molecular dynamics (REMD). The use of REMD simulations facilitates comparison to our 500 ns trajectories that do not include enhanced sampling. For both models, trp-cage RMSDs at or below 2.0 Å are the most probable, although the distributions exhibit different features. RMSD histograms for all proteins tested in this work are shown in Fig. 12. The percentage of snapshots across the 500 ns trajectory that fall into each of the RMSD bins used to

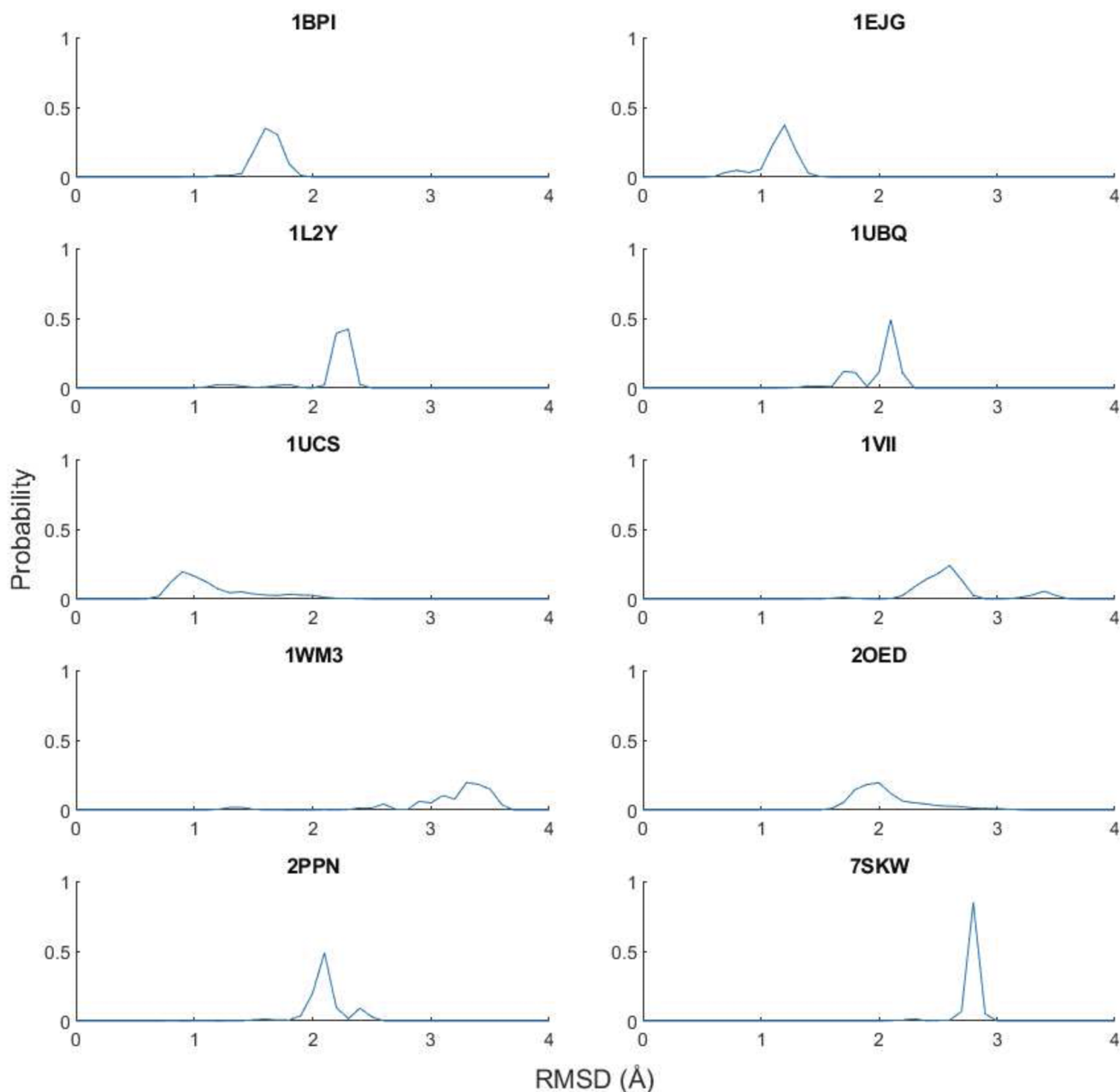


FIG. 12. Histograms of RMSD values across 500 ns of simulation for AMOEBA/GK implicit solvent are shown for each protein.

create the histograms is tabulated in the supplementary material, Table S4.

Average dipole moment magnitudes across production MD trajectories were calculated for all proteins in both vacuum and GK implicit solvents. Dipole moment magnitudes in GK implicit solvents were calculated with and without interstitial space corrections.

Average magnitudes in GK implicit solvent were ~30% to 35% larger than those in vacuum, with the addition of interstitial space corrections slightly reducing average dipole moment magnitudes (Fig. 13). The change between vacuum and condensed phase dipole moment magnitudes can only be captured using polarizable force fields, such as the AMOEBA and Drude models.

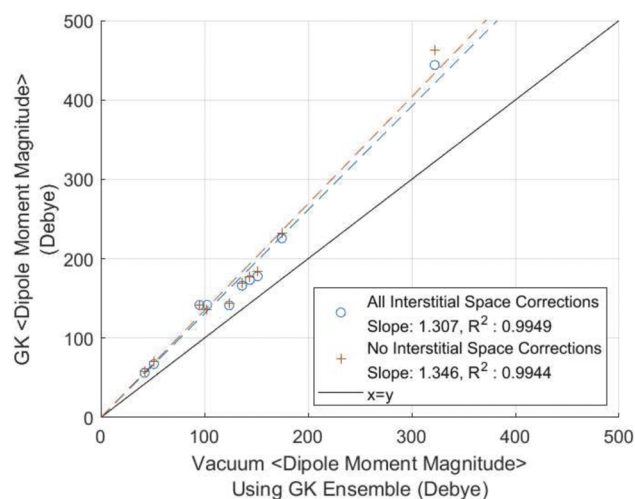


FIG. 13. Comparison of average dipole moment magnitudes across production MD trajectories in a vacuum and GK implicit solvent. Implicit solvent dipole moment magnitudes are reported with (blue circles) and without (orange pluses) interstitial space corrections. The dashed lines are the best fit regression lines for GK dipole moment magnitudes, while the solid black line is $x = y$ to guide the eye.

CONCLUSIONS

In this work, updates to the AMOEBA GK implicit solvent to support protein simulations have been described. We detailed the addition of nine parameters to the original implicit solvent model to improve its performance for interstitial spaces that arise commonly in biomolecules. Element-specific HCT scale factors S_x ($x = [C, N, O, P, S]$) are used for pair-wise descreening,¹⁶ which are strictly less than 1.0 since their role is to avoid overcounting of overlaps. The neck scaling factor, S_{neck} , corrects for overlaps between atomic neck regions (Fig. 2) and is also strictly less than 1.0. Finally, a tanh function with three parameters β_x ($x = 0, 1, 2$) increases the descreening integral to account for interstitial spaces due to many-body interactions, especially for deeply buried atoms.

The model was tested by running 5 μ s of MD simulation for a varied set of proteins. The protein trajectories were generally stable across 500 ns, with an average (non-terminal) backbone RMSD of 2.03 Å. The current fitting of GK base radii is biased toward experimentally available small molecule solvation free energy differences, which do not feature repetitive elements such as those in a protein or nucleic acid backbone. For this reason, additional tuning of base radii to account for repeated chemical groups was performed. Tensor recursion formulations for GK using both Cartesian and QI frames have been made available to ease implementation in software packages such as FFX,⁵⁵ OpenMM,⁵⁶ and Tinker.^{57,58} This recursive scheme will help to facilitate $n \cdot \log(n)$ implicit solvent implementations based multipolar methods,⁷⁶ including fixed charge GB models.⁷⁷

Future work may benefit from force matching^{78,79} data from explicit solvent biomolecular simulations to augment traditional fitting based on small molecule solvation free energy differences. One force matching parameterization method was described for 16 GRO-MOS atom types by Kleinjung,⁸⁰ and similar procedures could be

used with AMOEBA implicit solvent model fitting. The parameterization of interstitial space correction terms was designed to ensure that the implicit solvent model can be used to simulate proteins and nucleic acids simultaneously. The fitting of nucleic acid specific parameters, including the nucleic acid neck scale factor, and tuning of previously fit electrostatic radii for nucleic acid atom types to account for repetitive backbone chemistries will be addressed in future work.

The implicit solvent model for proteins is currently being used in the development of new protein optimization and design methods within FFX, including a family of side-chain optimization methods. These algorithms use a many-body energy expansion to determine optimal side chain conformations and titration states (e.g., for LYS, HIS, ASP, and GLU residues) from a set of low-energy conformations known as rotamers. Typically, the polypeptide backbone remains fixed while the side chains are moved through their rotamers. The use of a continuum solvent is essential to eliminate steric clashes with explicit solvent molecules as the energy of each rotamer (or pair of rotamers) is computed. This approach can be used in conjunction with new techniques for experimental structure determination (e.g., CryoEM, time resolved x-ray crystallography) that leverage the AMOEBA force field during refinement.^{6,81–85} However, the experimental resolution is rarely high enough to assign titratable amino acid protons. Additionally, manual placement of side chains during model building is time-consuming and can result in energetically nonoptimal structures. Using global sidechain optimization methods built on rotamer libraries,⁸⁶ it is possible to optimize both side-chain conformations and their titration states during model building and refinement.

This GK implicit solvent for proteins can also facilitate the development of constant pH molecular dynamics (CpHMD) algorithms for the AMOEBA force field. Implicit solvents have already been shown to work well with CpHMD methods using fixed charge force fields.^{52,87–89} For example, GB-CpHMD⁹⁰ in the AMBER simulation package has been used to predict pK_a shifts using the Amber ff14sb force field. The advantage of using an implicit solvent for CpHMD simulations is two-fold. First, the number of atoms being simulated is reduced. A second advantage is that solvents relax instantaneously to changes in ionization state, which avoids the relatively slow kinetics associated with water reorientation. This is especially apparent for enhanced sampling methods such as pH replica exchange, where different ionization states are immediately accommodated by continuum water during exchanges (i.e., promoting efficient pH replica exchange rates).

An implicit solvent was recently used in conjunction with a deep learning approach to calculate the absolute binding free energy difference of a host-guest system via the DeepBAR method presented by Ding and Zhang.⁹¹ With the implicit solvent model described here, the DeepBAR approach could now be applied to the series of host-guest systems modeled successfully by the AMOEBA polarizable force field in the context of the SAMPL challenges.^{92,93} Additional applications that stand to benefit from implicit solvation, such as the simulation of protein/nucleic acid complexes³⁰ or intrinsically disordered proteins,^{94,95} could be explored in future work. While the current model is parameterized for use with the AMOEBA force field, it will also be adapted for force fields with similar electrostatics models (e.g., AMOEBA+⁹⁶ and HIPPO⁹⁷) as they are developed. This will ensure transferability and the continued use

of more advanced force fields. Overall, the stability of the current model for a broad array of proteins with the addition of only a few new parameters shows promise for the expanded use of GK implicit solvent for biomolecular simulations.

SUPPLEMENTARY MATERIAL

The supplementary material contains (1) a Word document with two tables of neck integral constants, a plot showing perfect effective Born radii vs perfect effective Kirkwood radii for 1UBQ, GK diameters for selected AMOEBA protein atoms, and the numerical data used to generate the RMSD histograms; and (2) a tar archive with Force Field X input files that can be used to generate the protein simulation trajectories.

ACKNOWLEDGMENTS

Author R.A.C. was supported by the NSF (National Science Foundation) Graduate Research Fellowship under Grant No. 000390183. Authors J.W.P. and P.R. were supported by NIH Grant Nos. R01GM114237 and R01GM106137. Author MJS was supported by NIH Grant No. R01DC012049 and NSF Grant No. CHE-1751688.

AUTHOR DECLARATIONS

Conflict of Interest

J. W. Ponder and P. Ren are cofounders of Qubit Pharmaceuticals.

Author Contributions

Rae A. Corrigan: Conceptualization (equal); Data curation (lead); Formal analysis (equal); Investigation (lead); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (lead); Writing – review & editing (equal). **Andrew C. Thiel:** Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Writing – review & editing (equal). **Jack R. Lynn:** Data curation (equal); Formal analysis (equal); Methodology (equal); Software (equal); Writing – review & editing (equal). **Thomas L. Casavant:** Investigation (equal); Writing – review & editing (equal). **Pengyu Ren:** Conceptualization (equal); Formal analysis (equal); Methodology (equal); Software (equal); Validation (equal); Writing – review & editing (equal). **Jay W. Ponder:** Conceptualization (equal); Formal analysis (equal); Methodology (equal); Software (equal); Validation (equal); Writing – review & editing (equal). **Michael J. Schnieders:** Conceptualization (lead); Data curation (equal); Formal analysis (equal); Funding acquisition (lead); Investigation (equal); Methodology (equal); Project administration (lead); Resources (lead); Software (equal); Supervision (lead); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Raw data were generated at the Force Field X (<https://github.com/SchniedersLab/forcefieldx>, <https://ffx.biochem.uiowa.edu/>) large scale facility. Derived data supporting the findings

of this study are available from the corresponding author upon reasonable request.

REFERENCES

- C. D. Snow, E. J. Sorin, Y. M. Rhee, and V. S. Pande, “How well can simulation predict protein folding kinetics and thermodynamics?,” *Annu. Rev. Biophys. Biomol. Struct.* **34**, 43–69 (2005).
- K. A. Dill and J. L. MacCallum, “The protein-folding problem, 50 Years on,” *Science* **338**, 1042–1046 (2012).
- E. Gallicchio, M. Lapelosa, and R. M. Levy, “Binding energy distribution analysis method (BEDAM) for estimation of Protein–Ligand binding affinities,” *J. Chem. Theory Comput.* **6**, 2961–2977 (2010).
- E. Gallicchio and R. M. Levy, in *Advances in Protein Chemistry and Structural Biology*, edited by C. Christov (Academic Press, 2011), Vol. 85, pp. 27–80.
- E. Michael and T. Simonson, “How much can physics do for protein design?,” *Curr. Opin. Struct. Biol.* **72**, 46–54 (2022).
- M. J. Schnieders, T. D. Fenn, and V. S. Pande, “Polarizable atomic multipole X-ray refinement: Particle mesh Ewald electrostatics for macromolecular crystals,” *J. Chem. Theory Comput.* **7**, 1141–1156 (2011).
- B. Roux and T. Simonson, “Implicit solvent models,” *Biophys. Chem.* **78**, 1–20 (1999).
- M. J. Schnieders, N. A. Baker, P. Y. Ren, and J. W. Ponder, “Polarizable atomic multipole solutes in a Poisson-Boltzmann continuum,” *J. Chem. Phys.* **126**, 124114 (2007).
- M. Nottoli, A. Mikhalev, B. Stamm, and F. Lipparini, “Coarse-graining ddCOSMO through an interface between Tinker and the ddX library,” *J. Phys. Chem. B* **126**, 8827–8837 (2022).
- C. D. Cooper, J. P. Bardhan, and L. A. Barba, “A biomolecular electrostatics solver using Python, GPUs and boundary elements that can handle solvent-filled cavities and Stern layers,” *Comput. Phys. Commun.* **185**, 720–729 (2014).
- C. D. Cooper, “A boundary-integral approach for the Poisson-Boltzmann equation with polarizable force fields,” *J. Comput. Chem.* **40**, 1680–1692 (2019).
- W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, “Semianalytical treatment of solvation for molecular mechanics and dynamics,” *J. Am. Chem. Soc.* **112**, 6127–6129 (1990).
- D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still, “The GB/SA continuum model for solvation: A fast analytical method for the calculation of approximate Born radii,” *J. Phys. Chem. A* **101**, 3005–3014 (1997).
- A. V. Onufriev and D. A. Case, in *Annual Review of Biophysics*, edited by K. A. Dill (Annual Reviews, 2019), Vol. 48, pp. 275–296.
- M. J. Schnieders and J. W. Ponder, “Polarizable atomic multipole solutes in a generalized Kirkwood continuum,” *J. Chem. Theory Comput.* **3**, 2083–2097 (2007).
- G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, “Pairwise solute descreening of solute charges from a dielectric medium,” *Chem. Phys. Lett.* **246**, 122–129 (1995).
- W. P. Im, M. S. Lee, and C. L. Brooks, “Generalized Born model with a simple smoothing function,” *J. Comput. Chem.* **24**, 1691–1702 (2003).
- E. J. Arthur and C. L. Brooks, “Parallelization and improvements of the generalized Born model with a simple sWitching function for modern graphics processors,” *J. Comput. Chem.* **37**, 927–939 (2016).
- A. D. MacKerell *et al.*, “All-atom empirical potential for molecular modeling and dynamics studies of proteins,” *J. Phys. Chem. B* **102**, 3586–3616 (1998).
- R. B. Best *et al.*, “Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles,” *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
- F. M. Richards, “Areas, volumes, packing and protein structure,” *Annu. Rev. Biophys. Bioeng.* **6**, 151–176 (1977).
- M. S. Lee, F. R. Salsbury, and C. L. Brooks, “Novel generalized Born methods,” *J. Chem. Phys.* **116**, 10606–10614 (2002).
- M. S. Lee, M. Feig, F. R. Salsbury, and C. L. Brooks, “New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations,” *J. Comput. Chem.* **24**, 1348–1356 (2003).

- ²⁴V. Hornak *et al.*, "Comparison of multiple amber force fields and development of improved protein backbone parameters," *Proteins* **65**, 712–725 (2006).
- ²⁵A. Onufriev, D. Bashford, and D. Case, "Exploring protein native states and large-scale conformational changes with a modified generalized Born model," *Proteins* **55**, 383–394 (2004).
- ²⁶J. Mongan, C. Simmerling, J. A. McCammon, D. A. Case, and A. Onufriev, "Generalized Born model with a simple, robust molecular volume correction," *J. Chem. Theory Comput.* **3**, 156–169 (2007).
- ²⁷B. Aguilar, R. Shadrach, and A. V. Onufriev, "Reducing the secondary structure bias in the generalized Born model via R6 effective radii," *J. Chem. Theory Comput.* **6**, 3613–3630 (2010).
- ²⁸T. Gryczuk, "Deficiency of the Coulomb-field approximation in the Generalized Born model: An improved formula for Born radii evaluation," *J. Chem. Phys.* **119**, 4817–4826 (2003).
- ²⁹H. Nguyen, D. R. Roe, and C. Simmerling, "Improved generalized Born solvent model parameters for protein simulations," *J. Chem. Theory Comput.* **9**, 2020–2034 (2013).
- ³⁰H. Nguyen, A. Perez, S. Bermeo, and C. Simmerling, "Refinement of generalized Born implicit solvation parameters for nucleic acids and their complexes with proteins," *J. Chem. Theory Comput.* **11**, 3714–3728 (2015).
- ³¹J. R. Maple *et al.*, "A polarizable force field and continuum solvation methodology for modeling of Protein–Ligand interactions," *J. Chem. Theory Comput.* **1**, 694–715 (2005).
- ³²Y. Shi *et al.*, "Polarizable atomic multipole-based AMOEBA force field for proteins," *J. Chem. Theory Comput.* **9**, 4046–4063 (2013).
- ³³C. Zhang *et al.*, "AMOEBA polarizable atomic multipole force field for nucleic acids," *J. Chem. Theory Comput.* **14**, 2084–2108 (2018).
- ³⁴F. Lipparini *et al.*, "Polarizable molecular dynamics in a polarizable continuum solvent," *J. Chem. Theory Comput.* **11**, 623–634 (2015).
- ³⁵R. A. Corrigan *et al.*, "Implicit solvents for the polarizable atomic multipole AMOEBA force field," *J. Chem. Theory Comput.* **17**, 2323–2341 (2021).
- ³⁶S. Miertus, E. Scrocco, and J. Tomasi, "Electrostatic interaction of a solute with a continuum. A direct utilization of AB initio molecular potentials for the prevision of solvent effects," *Chem. Phys.* **55**, 117–129 (1981).
- ³⁷C. J. Cramer and D. G. Truhlar, "An SCF solvation model for the hydrophobic effect and absolute free energies of aqueous solvation," *Science* **256**, 213–217 (1992).
- ³⁸A. Klamt and G. Schuurmann, "COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," *J. Chem. Soc., Perkin Trans. 2*, 799–805 (1993).
- ³⁹A. Klamt, "Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena," *J. Phys. Chem.* **99**, 2224–2235 (1995).
- ⁴⁰E. Cancès, B. Mennucci, and J. Tomasi, "A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics," *J. Chem. Phys.* **107**, 3032–3041 (1997).
- ⁴¹C. J. Cramer and D. G. Truhlar, "Implicit solvation models: Equilibria, structure, spectra, and dynamics," *Chem. Rev.* **99**, 2161–2200 (1999).
- ⁴²J. W. Ponder and D. A. Case, *Advances in Protein Chemistry* (Academic Press, 2003), Vol. 66, pp. 27–85.
- ⁴³J. Tomasi, "Thirty years of continuum solvation chemistry: A review, and prospects for the near future," *Theor. Chem. Acc.* **112**, 184–203 (2004).
- ⁴⁴J. Tomasi, B. Mennucci, and R. Cammi, "Quantum mechanical continuum solvation models," *Chem. Rev.* **105**, 2999–3093 (2005).
- ⁴⁵C. P. Kelly, C. J. Cramer, and D. G. Truhlar, "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.* **1**, 1133–1152 (2005).
- ⁴⁶A. V. Marenich, C. J. Cramer, and D. G. Truhlar, "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* **113**, 6378–6396 (2009).
- ⁴⁷J. A. Lemkul, J. Huang, B. Roux, and A. D. MacKerell, "An empirical polarizable force field based on the classical drude oscillator model: Development history and recent applications," *Chem. Rev.* **116**, 4983–5013 (2016).
- ⁴⁸J. A. Lemkul and A. D. MacKerell, "Polarizable force field for RNA based on the classical Drude oscillator," *J. Comput. Chem.* **39**, 2624–2646 (2018).
- ⁴⁹P. P. Poier and F. Jensen, "Including implicit solvation in the bond capacity polarization model," *J. Chem. Phys.* **151**, 000006 (2019).
- ⁵⁰P. P. Poier and F. Jensen, "Polarizable charges in a generalized Born reaction potential," *J. Chem. Phys.* **153**, 000010 (2020).
- ⁵¹A. Aleksandrov, F.-Y. Lin, B. Roux, and A. D. MacKerell, Jr., "Combining the polarizable Drude force field with a continuum electrostatic Poisson–Boltzmann implicit solvation model," *J. Comput. Chem.* **39**, 1707–1719 (2018).
- ⁵²A. Aleksandrov, B. Roux, and A. D. MacKerell, "pKa calculations with the polarizable Drude force field and Poisson–Boltzmann solvation model," *J. Chem. Theory Comput.* **16**, 4655–4668 (2020).
- ⁵³B. Zhang, D. Kilburg, P. Eastman, V. S. Pande, and E. Gallicchio, "Efficient Gaussian density formulation of volume and surface areas of macromolecules on graphical processing units," *J. Comput. Chem.* **38**, 740–752 (2017).
- ⁵⁴J. D. Weeks, D. Chandler, and H. C. Andersen, "Role of repulsive forces in determining the equilibrium structure of simple liquids," *J. Chem. Phys.* **54**, 5237–5247 (1971).
- ⁵⁵M. J. Schnieders, Force Field X, Version 1.0, <https://ffx.biochem.uiowa.edu>, 2021.
- ⁵⁶P. Eastman *et al.*, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics," *PLoS Comput. Biol.* **13**, e1005659 (2017).
- ⁵⁷J. A. Rackers *et al.*, "Tinker 8: Software tools for molecular design," *J. Chem. Theory Comput.* **14**, 5273–5289 (2018).
- ⁵⁸L. Lagardère *et al.*, "Tinker-HP: A massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields," *Chem. Sci.* **9**, 956–972 (2018).
- ⁵⁹E. Gallicchio, M. M. Kubo, and R. M. Levy, "Enthalpy–Entropy and cavity decomposition of alkane hydration free Energies: Numerical results and implications for Theories of hydrophobic solvation," *J. Phys. Chem. B* **104**, 6271–6285 (2000).
- ⁶⁰E. Gallicchio, L. Y. Zhang, and R. M. Levy, "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.* **23**, 517–529 (2002).
- ⁶¹E. Gallicchio and R. M. Levy, "AGBNP: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling," *J. Comput. Chem.* **25**, 479–499 (2004).
- ⁶²E. Gallicchio, K. Paris, and R. M. Levy, "The AGBNP2 implicit solvation model," *J. Chem. Theory Comput.* **5**, 2544–2564 (2009).
- ⁶³N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, "Electrostatics of nanosystems: Application to microtubules and the ribosome," *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–10041 (2001).
- ⁶⁴E. Jurrus *et al.*, "Improvements to the APBS biomolecular solvation software suite," *Protein Sci.* **27**, 112–128 (2018).
- ⁶⁵M. Challacombe, E. Schwegler, and J. Almlöf, "Recurrence relations for calculation of the Cartesian multipole tensor," *Chem. Phys. Lett.* **241**, 67–72 (1995).
- ⁶⁶A. C. Simmonett, F. C. Pickard, H. F. Schaefer, and B. R. Brooks, "An efficient algorithm for multipole energies and derivatives based on spherical harmonics and extensions to particle mesh Ewald," *J. Chem. Phys.* **140**, 184101 (2014).
- ⁶⁷Z. Xu, X. Cheng, and H. Yang, "Treecode-based generalized Born method," *J. Chem. Phys.* **134**, 064107 (2011).
- ⁶⁸D. Cha *et al.*, in *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics* (Association for Computing Machinery, Atlanta, GA, 2015), Vol. 222–231.
- ⁶⁹A. Onufriev, D. A. Case, and D. Bashford, "Effective Born radii in the generalized Born approximation: The importance of being perfect," *J. Comput. Chem.* **23**, 1297–1304 (2002).
- ⁷⁰J. G. Kirkwood, "Theory of solutions of molecules containing widely separated charges with special application to zwitterions," *J. Chem. Phys.* **2**, 351–361 (1934).
- ⁷¹J. Applequist, "Traceless Cartesian tensor forms for spherical harmonic functions: New theorems and applications to electrostatics of dielectric media," *J. Phys. A: Math. Gen.* **22**, 4303–4330 (1989).

- ⁷²C. J. F. Böttcher, *Dielectrics in Static Fields*, 1st ed. (Elsevier Publishing Company, 1952), Vol. 1.
- ⁷³A. Bondi, "van der Waals volumes and radii," *J. Phys. Chem.* **68**, 441–451 (1964).
- ⁷⁴M. Connolly, "Analytical molecular surface calculation," *J. Appl. Crystallogr.* **16**, 548–558 (1983).
- ⁷⁵M. L. Connolly, "Computation of molecular volume," *J. Am. Chem. Soc.* **107**, 1118–1124 (1985).
- ⁷⁶J. Barnes and P. Hut, "A hierarchical $O(N \log N)$ force-calculation algorithm," *Nature* **324**, 446–449 (1986).
- ⁷⁷C. Bajaj and W. Zhao, "Fast molecular solvation energetics and forces computation," *SIAM J. Sci. Comput.* **31**, 4524–4552 (2010).
- ⁷⁸S. Bottaro, K. Lindorff-Larsen, and R. B. Best, "Variational optimization of an all-atom implicit solvent force field to match explicit solvent simulation data," *J. Chem. Theory Comput.* **9**, 5641–5652 (2013).
- ⁷⁹J. Kleinjung and F. Fraternali, "Design and application of implicit solvent models in biomolecular simulations," *Curr. Opin. Struct. Biol.* **25**, 126–134 (2014).
- ⁸⁰J. Kleinjung, W. R. P. Scott, J. R. Allison, W. F. van Gunsteren, and F. Fraternali, "Implicit solvation parameters derived from explicit water forces in large-scale molecular dynamics simulations," *J. Chem. Theory Comput.* **8**, 2391–2403 (2012).
- ⁸¹M. J. Schnieders, T. D. Fenn, V. S. Pande, and A. T. Brunger, "Polarizable atomic multipole X-ray refinement: Application to peptide crystals," *Acta Crystallogr., Sect. D* **65**, 952–965 (2009).
- ⁸²T. D. Fenn, M. J. Schnieders, A. T. Brunger, and V. S. Pande, "Polarizable atomic multipole X-ray refinement: Hydration geometry and application to macromolecules," *Biophys. J.* **98**, 2984–2992 (2010).
- ⁸³T. D. Fenn, M. J. Schnieders, and A. T. Brunger, "A smooth and differentiable bulk-solvent model for macromolecular diffraction," *Acta Crystallogr., Sect. D* **66**, 1024–1031 (2010).
- ⁸⁴T. D. Fenn *et al.*, "Reintroducing electrostatics into macromolecular crystallographic refinement: Application to neutron crystallography and DNA hydration," *Structure* **19**, 523–533 (2011).
- ⁸⁵T. D. Fenn and M. J. Schnieders, "Polarizable atomic multipole X-ray refinement: Weighting schemes for macromolecular diffraction," *Acta Crystallogr., Sect. D* **67**, 957–965 (2011).
- ⁸⁶S. D. LuCore *et al.*, "Dead-end elimination with a polarizable force field repacks PCNA structures," *Biophys. J.* **109**, 816–826 (2015).
- ⁸⁷A. M. Baptista, P. J. Martel, and S. B. Petersen, "Simulation of protein conformational freedom as a function of pH: Constant-pH molecular dynamics using implicit titration," *Proteins* **27**, 523–544 (1997).
- ⁸⁸M. S. Lee, F. R. Salsbury, and C. L. Brooks, "Constant-pH molecular dynamics using continuous titration coordinates," *Proteins* **56**, 738–752 (2004).
- ⁸⁹J. Khandogin and C. L. Brooks, "Constant pH molecular dynamics with proton tautomerism," *Biophys. J.* **89**, 141–157 (2005).
- ⁹⁰R. C. Harris and B. M. Pettitt, "Effects of geometry and chemistry on hydrophobic solvation," *Proc. Natl. Acad. Sci. U. S. A.* **111**, 14681–14686 (2014).
- ⁹¹X. Q. Ding and B. Zhang, "DeepBAR: A fast and exact method for binding free energy computation," *J. Phys. Chem. Lett.* **12**, 2509–2515 (2021).
- ⁹²M. K. J. Chung, R. J. Miller, B. Novak, Z. Wang, and J. W. Ponder, "Accurate host-guest binding free energies using the AMOEBA polarizable force field," *J. Chem. Inf. Model.* **63**, 2769–2782 (2023).
- ⁹³M. Amezcua, J. Setiadi, Y. Ge, and D. L. Mobley, "An overview of the SAMPL8 host-guest binding challenge," *J. Comput. Aided Mol. Des.* **36**, 707–734 (2022).
- ⁹⁴T. H. Click, D. Ganguly, and J. H. Chen, "Intrinsically disordered proteins in a physics-based world," *Int. J. Mol. Sci.* **11**, 5292–5309 (2010).
- ⁹⁵J. E. Shea, R. B. Best, and J. Mittal, "Physics-based computational and theoretical approaches to intrinsically disordered proteins," *Curr. Opin. Struct. Biol.* **67**, 219–225 (2021).
- ⁹⁶C. W. Liu, J. P. Piquemal, and P. Y. Ren, "AMOEBA+ classical potential for modeling molecular interactions," *J. Chem. Theory Comput.* **15**, 4122–4139 (2019).
- ⁹⁷J. A. Rackers, R. R. Silva, Z. Wang, and J. W. Ponder, "Polarizable water potential derived from a model electron density," *J. Chem. Theory Comput.* **17**, 7056–7084 (2021).