JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2019, 15, 4122−4139

pubs.acs.org/JCTC

AMOEBA+ Classical Potential for Modeling Molecular Interactions

Chengwen Liu,[†] Jean-Philip Piquemal,[†,‡,§] and Pengyu Ren*[,†]

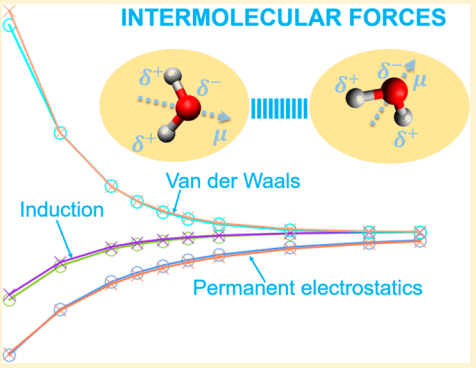
[†]Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712, United States

[‡]Laboratoire de Chimie Théorique, Sorbonne Université, UMR7616 CNRS, Paris 75252, France

[§]Institut Universitaire de France, Paris Cedex 05, 75005, France

Supporting Information

ABSTRACT: Classical potentials based on isotropic and additive atomic charges have been widely used to model molecules in computers for the past few decades. The crude approximations in the underlying physics are hindering both their accuracy and transferability across chemical and physical environments. Here we present a new classical potential, AMOEBA+, to capture essential intermolecular forces, including permanent electrostatics, repulsion, dispersion, many-body polarization, short-range charge penetration, and charge transfer, by extending the polarizable multipole-based AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) model. For a set of common organic molecules, we show that AMOEBA+ with general parameters can reproduce both quantum mechanical interactions and energy decompositions according to Symmetry-Adapted Perturbation Theory (SAPT). Additionally, a new water model based on the AMOEBA+ framework captures various liquid-phase properties in molecular dynamics simulations while remaining consistent with SAPT energy decompositions, utilizing both ab initio data and experimental liquid properties. Our results demonstrate that it is possible to improve the physical basis of classical force fields to advance their accuracy and general applicability.



1. INTRODUCTION

Classical molecular dynamics (MD) simulations are widely used to study the physical properties of chemical and biological systems. One of the essential ingredients in the MD simulations is the accuracy of the underlying classical potentials, or force fields (FFs), which include the functional forms that describe the intra- and intermolecular potential energy surface (PES) and the parameters associated with different chemistry. The traditional FFs, such as AMBER (Assisted Model Building with Energy Refinement),[1−4] CHARMM (Chemistry at HARvard Macromolecular Mechanics),[5−7] GROMOS (Groningen Molecular Simulation),[8,9] and OPLS (Optimized Potential for Liquid Simulations),[10−12] employ fixed atomic charges to model the electrostatic interactions via pairwise additive Coulombic interactions. While they have been widely used in complex systems due to their efficiency in simulations, there has been increasing effort to improve the underlying physics, particularly the many-body polarization that can vary significantly depending on chemical and physical environments.[13−15] For example, it is well-known that a water molecule in isolation has a dipole moment of 1.85 D that increases by about 50% to 2.4−3.0 D in the liquid phase due to polarization at room temperature.[16−18] Mainly three approaches have been adopted to explicitly capture many-body polarization in FFs, including the Drude oscillators,[19−21] fluctuating charges,[6,22] and induced dipole schemes.[13,16,23,24] Polarizable AMBER,[25] OPLS,[26] GROMOS,[27] CHARMM,[28] NEMO (Nonempirical Molecular Orbital),[29,30] and PFF (Polarizable Force Field for proteins)[31] employ one of the three approaches. Another area of improvement focuses on the deficiency of spherical atomic charge representation of permanent electrostatics. Atomic multipole expansion was adopted to capture the anisotropic electrostatic potentials (ESPs) around atoms; examples include AMOEBA,[16] NEMO,[32] EFP (Effective Fragment Potential),[33,34] and SIBFA (Sum of Interactions Between Fragments Ab initio computed).[35−37] More sophisticated approaches are used by Gaussian-based GEM (Gaussian Electrostatic Model)[38,39] and fragment-based electronic structure method X-Pol.[24,40] We have been developing the AMOEBA potential that employs atomic multipole expansion up to quadrupoles to represent the permanent charge distributions and inducible atomic dipoles to account for the polarization response.[16,41] It has been applied to water,[16,42] ions,[43−45] small organic molecules,[23] and complex proteins[46] and nucleic acids,[41] with extensive applications to protein−ligand and ion binding.[47−52]

While computationally expensive, ab initio quantum mechanical (QM) energy decomposition analysis (EDA) such as SAPT[53] provide detailed components of the interaction energy including electrostatics, induction, exchange-repulsion, and dispersion. However, classical FFs have not systematically utilized such information due to the intrinsic limitations of the underlying physical approximations. Our studies over the past few years suggest that it is possible to model after SAPT with

Received: March 13, 2019
Published: May 28, 2019

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

relatively simple classical terms. For example, electrostatic interactions between molecules are well represented by multipole expansion at long distance. At short distances where the electron clouds overlap, the ESPs of atoms deviate from the pure Coulomb form $(1/r)$ due to electron shielding. This is known as the charge penetration (CP) effect.[54] While the CP correction is short-ranged, it can be significant at equilibrium geometry of typical molecular complexes. Empirical damping formulas for treating CP have been proposed by several research groups.[55−60] We have demonstrated that combining CP correction with point multipoles[58,60] leads to electrostatic interactions closely matching those from SAPT decomposition even at short distances while the added computational cost is minimal. This is a major step toward eliminating error cancellation between what are referred to as electrostatic and van der Waals interactions in classical FFs. A prominent example of the CP effect is the benzene−benzene system.[61] The electrostatic interaction of a stacked benzene is attractive according to ab initio EDA.[62,63] However, classical point electrostatic models will yield repulsive interactions when C and H atoms are stacked on top of the same C and H. Thus, to achieve reasonable benzene stacking energy, the vdW repulsion has to be reduced artificially, which however leads to transferability issues with other molecules. Induction is another fundamental force and often separated in concept into charge transfer (CT) and polarization. In contrast to CP, the definition of the CT effect has been far from clear, despite the fact that this concept has been proposed for more than 60 years.[64] Several quantum chemical EDA methods attempted to separate the CT and polarization from the total induction energy. However, the resulting CT energy ranges from −38.37 to −1.49 kJ·mol^{-1} for the hydrogen-bonding water dimer depending on the methods used (see the paper by Mao et al.[65] and references therein). Stone has pointed out that the EDAs based on the natural bond orbital to separate CT and polarization are intrinsically problematic.[66] An empirical scheme has been proposed to separate CT from the SAPT induction energy.[67] To overcome the uncertainty in CT energy, we have employed an effective approach where we first determine the polarization model against the many-body MP2 energy for a wide range of molecular clusters in various intermolecular distances.[68] The CT energy is then obtained by subtracting the model polarization energy from the SAPT induction energy.

In this paper, we integrate our progress over the past few years and present a new intermolecular potential, AMOEBA+. By incorporating explicit CP and CT terms, as well as improved polarization and van der Waals functions, the AMOEBA+ potential is able to accurately describe intermolecular interactions and their components with general parameters. In addition, a water model based on the AMOEBA+ framework is demonstrated. The water model performs well in simulating liquid-phase properties, while the individual intermolecular force components remain consistent with SAPT energy decomposition.

2. METHODOLOGY

2.1. AMOEBA+ Potential Energy Terms. The total potential energy of the AMOEBA+ model can be expressed as the sum of the bonded and nonbonded energy terms

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}} \tag{1}$$

where E_{bonded} and $E_{\text{nonbonded}}$ can be expressed as

$$E_{\text{bonded}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{b-a}} + E_{\text{oop}} + E_{\text{torsion}} \tag{2}$$

$$E_{\text{nonbonded}} = E_{\text{electrostatics}}^{\text{CP-corrected}} + E_{\text{polarization}} + E_{\text{charge-transfer}} + E_{\text{VDW}} \tag{3}$$

The detailed functional forms for bonded terms, including bond, angle, bond−angle coupling, out-of-plane, and torsion of the AMOEBA model have been retained in AMOEBA+.[14,23] As for the AMOEBA+ water model, the bonded terms include bond stretching, angle bending, and the Urey−Bradley, as with the current AMOEBA water models.[16,69] The main improvements of AMOEBA+ over the AMOEBA FF on functional form lie in the nonbonded interactions, as will be described in detail next.

2.1.1. Electrostatics. The current AMOEBA model employs point atomic multipoles truncated at quadrupoles to compute the electrostatic interactions. In AMOEBA, the atomic multipole moments are usually derived with the distributed multipole analysis (DMA) approach[70] and then optimized against a high-level ab initio ESP. Although the atomic charge is not a quantum observable, various population analysis schemes exist to derive atomic multipoles that facilitate our chemical understanding and FF development. Atom-in-molecule (AIM) methods, such as Hirshfeld,[71] Hirshfeld-I,[72,73] and iterative stockholder analysis (ISA),[74−76] underwent much improvement in the past decades. Several attempts to develop molecular mechanical potentials using these approaches also appeared recently.[72,77,78] In this work, both the DMA- and ISA-based multipoles were explored to examine their performance on simulating water properties. The CP effect was incorporated into AMOEBA+ using the scheme by Gordon and coworkers,[55,60] where damping functions were applied to the full multipole−multipole interaction. As an illustration, the CP-corrected charge−charge interaction between sites i and j is expressed as

$$E_{\text{elst}}^{\text{chg-chg}} = \frac{Z_i Z_j}{r} + \frac{Z_i q_j}{r} f_{\text{damp}}(r) + \frac{Z_j q_i}{r} f_{\text{damp}}(r) + \frac{q_i q_j}{r} f_{\text{damp}}^{\text{overlap}}(r) \tag{4}$$

where the Z_i and Z_j are nuclei charges and q_i and q_j are electron charges. Thus, $Z_i + q_i = c_i^{\text{DMA}}$ will be the actual net charge on atom i. When the damping is removed ($f = 1$), the above equation reduces to the familiar Coulombic form $(c_i^{\text{DMA}} c_j^{\text{DMA}}/r)$ in common FFs. In AMOEBA+, damping is systematically applied to the monopole, dipole, and quadrupole. The electrostatic energy of AMOEBA+ is then the sum of the multipole−multipole interaction energy and the CP correction. A compact form of the CP-corrected electrostatic energy between sites i and j can be written as

$$E_{\text{elst}}^{\text{CP-corr}} = \frac{Z_i Z_j}{r} + Z_i T_{ij}^{\text{damp}} M_j + Z_j T_{ji}^{\text{damp}} M_i + M_i^{\text{t}} T_{ij}^{\text{overlap}} M_j \tag{5}$$

Within the particle mesh Ewald (PME), the CP was incorporated as "modifications" in real space, i.e., the original (non-Ewald) pairwise multipole interactions were subtracted while the CP-corrected terms described above were added. The effect of CP damping is short-ranged and dies off at typical real-space Ewald cutoffs. For example, the magnitude of CP

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

correction is on the order of 10^{-6} and 10^{-8} kcal·mol^{-1} when two water molecules are 6 and 7 Å apart, respectively.

2.1.2. Polarization. The AMOEBA potential utilizes the interactive atomic dipole induction scheme to represent the many-body polarization effect. A point dipole is induced at each polarizable site (atom) by the total electric field felt by that site

$$\mu_i^{\text{ind}} = \alpha_i \left(\sum_j T_{ij}^{\text{damped}} M_j + \sum_j T_{ij}^{\text{damped}} \mu_j^{\text{ind}} \right) \tag{6}$$

where T_{ij} is the multipole−multipole interaction matrix for *direct* induction and T_{ij}^{11} includes only the dipole field-related terms for *mutual* induction. The Thole damping scheme[79] is used in the AMOEBA model to ensure the finite nature of the intermolecular polarization and proper anisotropic molecular response. This scheme is rather successful in reproducing molecular polarizability tensors for a broad range of organic molecules using element-based isotropic atomic polarizabilities.[79,80] AMOEBA uses the same damping function for the direct and mutual induction. The damping function corresponding to the T_α matrix is

$$f_{\text{Thole}}(r) = 1 - e^{-au^3(r)} \tag{7}$$

where $u(r) = r_{ij}/(\alpha_i\alpha_j)^{1/6}$ is the polarizability-normalized distance between sites i, j and r_{ij} is the actual distance, and α is the atomic polarizability. The current AMOEBA model uses the same damping factor a for both the direct and mutual parts.[16] In our recent work, we systematically examined the damping functional form and the damping factor by explicitly comparing against the MP2 many-body interaction energy for a range of molecular clusters at various orientations and separations.[68] We found that a better distance dependence behavior of the three-body energy (E_{3B}) can be achieved by modifying the functional form for damping the *permanent* field to

$$f_{\text{MB}}^{\text{direct}}(r) = 1 - e^{-au^{3/2}(r)} \tag{8}$$

This new damping function for the permanent field was adopted in the AMOEBA+ potential. The damping functional form and parameter for mutual induction remain the same as those for AMOEBA. More details of the modified polarization model can be found in a previous publication.[68]

2.1.3. Charge Transfer. The CT refers to the *stabilization energy* between atoms at near-covalent distances. In our model, we are not explicitly considering the transfer of electrons between them. In our model, CT energy is computed as the difference between the SAPT2+[81] induction energy and the polarization energy, as described above: $E_{\text{Induction}}^{\text{SAPT2+}} - E_{\text{Polarization}}^{\text{AMOEBA+}}$. SIBFA has an elaborate many-body charge transfer term using a triple overlap function of the damped distributed point multipoles.[82] Simpler pairwise additive functions are also used to model CT energy, especially in water, including the discrete CT approach[83−85] and electron density overlap approximation method[38] where the monomer density is approximated by exponential functions. We noted that the CT with exponential summation only between hydrogen and oxygen atoms in water[85] may have a deficiency in treating the case where two heavy atoms closely contact. Here we employ a pairwise exponential function to describe this interaction between *any* two atoms belonging to two molecules

$$E_{\text{charge-transfer}} = -\sum_{ij} a_{ij} \exp(-b_{ij} r_{ij}) \tag{9}$$

where a_{ij} is related to the magnitude of the energy and b_{ij} controls the distance dependence behavior. Combining rules were used for two heteroatoms as

$$a_{ij} = \sqrt{a_i a_j} \tag{10}$$

$$b_{ij} = \frac{1}{2}(b_i + b_j) \tag{11}$$

In practice, CT interaction is very short-ranged, and a 6 Å cutoff is used for the CT energy and gradient. Polynomial functions are also added to switch off the CT energy and forces near the cutoff distance. The many-body effect of CT can be important for certain systems such as high-valence ions, which will be treated separately in our future work.

2.1.4. van der Waals. The AMOEBA+ model retains the same vdW functional form used in the current AMOEBA, the Halgren's buffered-14−7 potential[92]

$$E_{\text{vdW}} = \varepsilon_{ij} \left(\frac{1+\delta}{\sigma_{ij}+\gamma} \right)^7 \left(\frac{1+\gamma}{\sigma_{ij}^7+\gamma} - 2 \right) \tag{12}$$

where ε_{ij} is the potential well depth and $\sigma_{ij} = r_{ij}/r_{ij}^0$, with r_{ij} as the separation between i and j. We use fixed values of $\delta = 0.07$ and $\gamma = 0.12$ in AMOEBA+. In the current AMOEBA FF, CUBIC-MEAN (eq 13) and HHG (eq 14) combining rules are used for r_{ij}^0 and ε_{ij}, respectively.

$$r_{ij}^0 = \frac{(r_{ii}^0)^3 + (r_{jj}^0)^3}{(r_{ii}^0)^2 + (r_{jj}^0)^2} \tag{13}$$

$$\varepsilon_{ij} = \frac{4\varepsilon_{ii}\varepsilon_{jj}}{(\sqrt{\varepsilon_{ii}} + \sqrt{\varepsilon_{jj}})^2} \tag{14}$$

In this work, we also examined the performance of the Waldman−Hagler (W−H) rule (eq 15) for ε_{ij}.

$$\varepsilon_{ij} = 2\sqrt{\varepsilon_{ii}\varepsilon_{jj}} \frac{(r_{ii}^0 r_{jj}^0)^3}{(r_{ii}^0)^6 + (r_{jj}^0)^6} \tag{15}$$

2.2. Code Implementation on CPU and GPU Platforms. The AMOEBA+ model was implemented on the CPU platform based on the Tinker 8.2 source code[86] and CUDA platform supported in the Tinker-OpenMM toolkit.[87,88] It mostly involves adding the new CP and CT energy and forces and the new vdW mixing rule to the existing AMOEBA routines and modifying the damping for permanent (direct) polarization. Specifically, CT-related subroutines in Tinker-CPU and CUDA kernel functions in Tinker-OpenMM were implemented in the same manner as the pairwise vdW. An individual cutoff distance (default 6 Å) and a switching function were used in the calculation of the CT energy and forces. CP was incorporated as pairwise corrections to the multipole−multipole interactions in the real space of the Ewald sum. The reciprocal space Ewald summation code was not affected. The AMOEBA+ polarization model, where the damping functions of direct polarization were changed, was implemented as a modification to the direct polarization energy, field, and force. The code related to mutual induction was not affected. On a single CPU, the total computational cost of the AMOEBA+ energy and gradient evaluation, compared to AMOEBA, is

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

increased slightly by ~ 6%, which includes the cost of added CP and CT, as well as the modified (direct) polarization energy and gradient. The detailed computational cost of both the energy and gradient by individual energy components is also provided in Table 1. The canonical Tinker-CPU code serves as the reference code for developing new algorithms and FFs. It is also critical to implement computationally efficient and high-performance simulation packages that support the use of new potentials with more complex potential energy functions. It has been shown the GPU implementation of an AMOEBA-based MD engine[88] offers a roughly 200-fold acceleration compared to a single CPU core. The preliminary implementation of AMOEBA on GPU can achieve ~20 ns/day on RTX2080 for a DHFR system (23 555 atoms) using a 2 fs time step. In another implementation, Tinker-HP[89] takes advantage of the massive parallelization of MPI and provides excellent scalability of performance over a large number of CPU cores for systems of ~1 million atoms. Preliminary implementation of AMOEBA+ GPU code in Tinker-OpenMM has been completed and utilized in this work. Implementation of AMOEBA+ in Tinker-HP, for CPU and a CPU−GPU hybrid platform, is in progress. The Tinker-CPU code is available at the TinkerTools GitHub site as the AMOEBA+ branch.

Table 1. Evaluation of the Computational Cost of the AMOEBA+ Model[a−c]

	AMOEBA	AMOEBA+	difference	difference %	change in AMOEBA+
E_{elst}	9.7	11.2	+1.5	+15.2	added CP correction
E_{pol}	40.9	41.9	+1.1	+2.6	split the direct and mutual induction
E_{ct}		0.9	+0.9		added new CT term
E_{tot}	53.1	56.4	+3.3	+6.3	all of the above

[a]CPU architecture: Intel (R) Xeon (R) CPU E5-2680 0 @ 2.70 GHz. [b]Important settings: The 36 Å^3 simulation box contains 1600 water molecules; cutoffs for vdW and Ewald real space and CT were 12, 7, and 6 Å, respectively; the neighbor-list method was used; induced dipoles were converged to 10×10^{-4}; 500 repetitions were run to evaluate the energy and force. [c]A total wall time (in second) of 500 energy and gradient calculations for each potential energy component was evaluated by the TIMER program in Tinker.

2.3. Parameterization of Organic Molecules Based on SAPT. All SAPT2+ data, Cartesian coordinate structure, and parameter files are also included in the SI. As described in our previous work, the S108×7 dimer set was systematically constructed, and the interaction energy was decomposed using the SAPT2+ method. Briefly, the S108×7 set contains 108 organic dimers, each with seven intermolecular separations, namely, at 0.70, 0.80, 0.90, 0.95, 1.00, 1.05, and 1.10 times the equilibrium distances. All 38 molecules in the S108 set are listed in Figure 1.

The interaction energy components of SAPT2+ can be expressed as

$$E_{SAPT2+} = E_{electrostatics} + E_{induction} + E_{exchange-repulsion} + E_{dispersion} \quad (16)$$

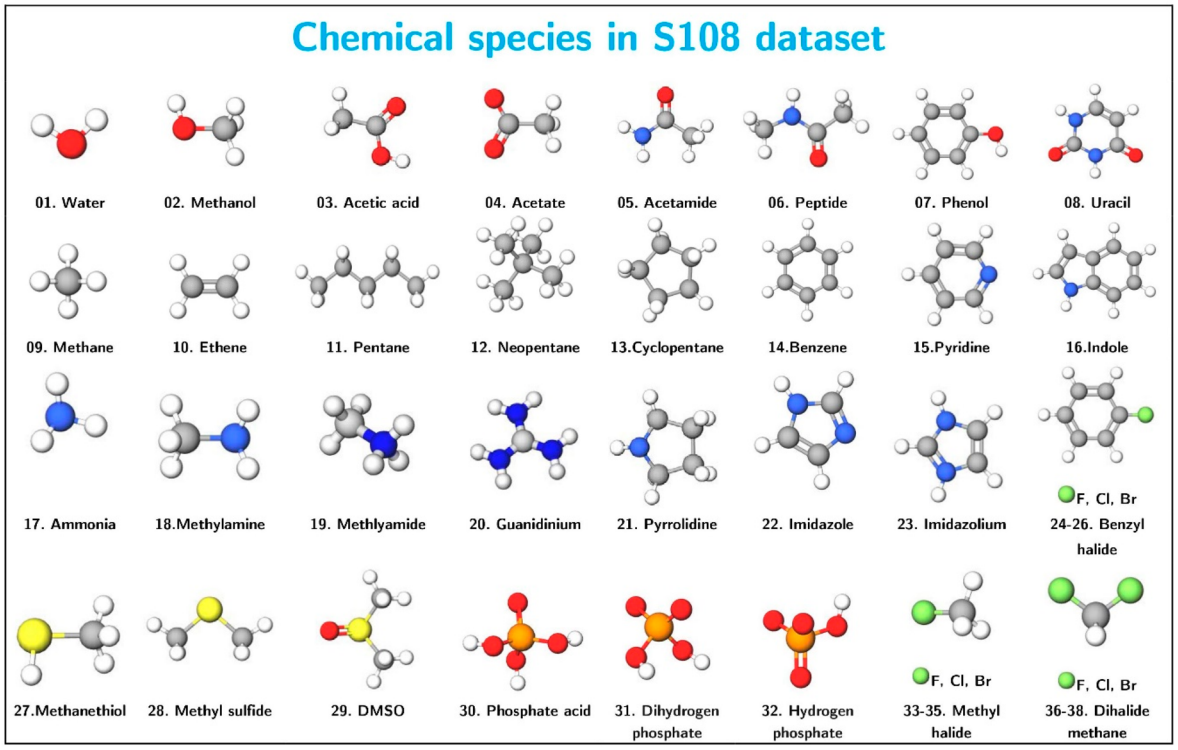


Figure 1. Molecular structures of the 38 molecules in the S108 data set. These molecules cover nine chemical elements: hydrogen (white), oxygen (red), carbon (gray), nitrogen (blue), sulfur (yellow), phosphor (orange), and halogen (green). Both the neutral molecules and charged molecules/fragments (e.g., 04, 19, 20, 23, 31, and 32) are included due to their relevance in biology as amino acid side-chain analogues or fragments of nucleic acids.

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

We first parametrized the AMOEBA+ model (eq 3) from QM calculations of the monomer and energy components by SAPT2+. The detailed one-to-one mappings are

$$E_{\text{electrostatics}}^{\text{AMOEBA+}} \rightarrow E_{\text{electrostatics}}^{\text{SAPT2+}}$$
$$E_{\text{polarization}}^{\text{AMOEBA+}} + E_{\text{charge-transfer}}^{\text{AMOEBA+}} \rightarrow E_{\text{induction}}^{\text{SAPT2+}}$$
$$E_{\text{van der Waals}}^{\text{AMOEBA+}} \rightarrow E_{\text{exchange-repulsion}}^{\text{SAPT2+}} + E_{\text{dispersion}}^{\text{SAPT2+}} \quad (17)$$

The atomic multipoles were derived for the molecules in the S108 data set in a previous work using a systematic procedure,[41] where the DMA approach was used to derive the initial multipoles, which were then optimized to high-level ESP. The CP parameters of the organic molecules were derived for the S101 data set,[42] and here, we directly expanded to the S108 set without further optimization. We modified the direct damping factor from the original 0.75 in our previous publication[68] to 0.70 in this work. It was found that this change leads to better agreement in the many-body energy of water clusters with MP2 values and only has a subtle influence on organic compounds. After subtracting the polarization energy from the total induction energy of SAPT2+, we parametrized the CT model with the remaining energy as the target. In the parametrization of the vdW model, we experimented with both the HHG and W–H combining rules for ε_{ij}. The cost function in the optimization of the vdW model was designed as the root-mean-square error (RMSE) of the model predicted vdW energy and that from SAPT2+. We fixed the shape parameters γ and δ to their original values due to better dispersion behavior beyond the equilibrium distances, which influence the bulk properties much more than small clusters.[90−92] To prevent overfitting and ensure the physical sense of vdW parameters, we slightly constrained the parameters in reasonable ranges by using a regularization term in the cost function.

2.4. Parametrization of the AMOEBA+ Water Model Based on Ab Initio and Experimental Data. As an application of the AMOEBA+ potential, the AMOEBA+ water model has been parametrized for liquid simulations using ForceBalance (FB).[3,93,94] FB uses a series of thermodynamic fluctuation equations to obtain the parametric derivatives of condensed-phase properties from MD simulations. Thus, it allows one to use both experimental properties and ab initio data in parameter optimization. FB has been used to develop new FFs, such as the iAMOEBA[95] and uAMOEBA,[96] and to revise the parameters of existing FFs, such as AMBER for proteins,[3] AMOEBA water,[69] and TIP3P/TIP4P waters.[94] In the optimization, the FF parameters were updated by FB iteratively until satisfied gas-phase and liquid properties were obtained (or the cost function minimized). The initial parameters of the bonded terms were taken from AMOEBA water03.prm.[16] Table 2 lists the targeting data in the parametrization of the AMOEBA+ water model. The objective (cost) function was designed as the sum of weighted mean-square errors from gas and liquid targets. Weights were applied among different fitting targets and physical properties within a target (Table 2). In this work, the optimization was carried out using the trust radius Newton–Raphson algorithm with an adaptive trust radius. The algorithm requires the first and second derivatives of the objective function in the parameter space.

We tested both DMA and ISA-based multipoles in developing water models. POLTYPE tool[97] was used to derive the initial DMA multipoles. We detail the computational procedure for deriving ISA multipoles in the SI. In FB optimization, the ANALYZE program of Tinker was used to compute the energy of water clusters and liquid boxes from MD simulations. The DYNAMIC_OMM program was used to perform liquid-phase *NPT* simulations. The box dimension was ~26 × 26 × 26 Å^3, containing 590 water molecules. The vdW cutoff was set to 10 Å in the FB run with a long-range correction. For simulations to correct the size effect on the self-diffusion constant, 12 Å was used in the larger water boxes. Long-range electrostatic interaction was treated using PME with a 7 Å real-space cutoff, which was also used as the cutoff for calculating CP correction. A RESPA integrator,[98,99] BUSSI thermostat,[100] Monte Carlo barostat, and 0.5 fs integrating time step were used in the *NPT* simulations. We also note that a larger time step (2.0 fs) using RESPA is very stable but will lead to slightly different properties. For example, with 2 fs MD, the water density will be about 1.0 kg·m^{-3} lower at 298 K than that of 0.5 fs. In each FB iteration, 250 ps of equilibration *NPT* simulation was first performed to equilibrate the water box. Then 5 ns of production run was performed to calculate the physical properties and gradients with respect to the parameters, which are listed in Table 5.

Table 2. Target Ab Initio QM Data and Experimental Measurements for Temperatures Ranging from 265 to 369 K (12 Data Points in Total)[a]

system and property		data point	data type	weight[b]	
clusters	hydrogen-bonding dimer	10	SAPT2+ interaction energy	1.0	
	Smith dimer	10	CCSD(T) BE	1.0	
	trimer to hexamer	11	CCSD(T) BE	1.0	
	octamer to 20-mer	18	MP2 BE	0.5	
liquid	density	12	experiment	1.0	3.0[c]
	enthalpy of vaporization	12	experiment		2.0[c]
	thermal expansion coefficient	12	experiment		1.0[c]
	static dielectric constant	12	experiment		1.0[c]
	isobaric heat capacity	12	experiment		0.1[c]
	isothermal compressibility	12	experiment		1.0[c]

[a]All of these data can be found in the "studies/015_amoeba_tinker" directory in the FB program, except for the "hydrogen-bonding dimer" set, of which the interaction energy was decomposed using SAPT2+ for dimers in 10 intermolecular separations. [b]Weights applied in FB optimization among different targets and properties within the target. Here each cluster system is a target, and liquid is another target. [c]Weights applied to the six liquid properties. Weights were also applied to properties at different temperatures (not shown).

3. RESULTS AND DISCUSSION

We previously obtained SAPT2+ intermolecular energy components for the S108×7 data set.[58] The parameters of CP, CT, and vdW obtained in this study are summarized in Tables S1 and S2. In total, AMOEBA+ has 18 atom classes for CP and CT, 28 atom classes for vdW, and 9 element-based atomic polarizability parameters. Thus, a very limited number of parameters (18 for CP, 36 for CT, 9 for polarizability, and 56 for vdW) are used in AMOEBA+. With optimized parameters, the AMOEBA+ model can accurately capture three energy components. We then developed an AMOEBA+ model for

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

Table 3. Statistics Evaluation of the AMOEBA+ Model on Each Energy Component and Total Interactions for the S108×7 Dimer Set Compared to SAPT2+ Data[a]

data points	statistics[b]	E_{elst}	E_{ind}	E_{vdW}[c]		E_{inter}[d]
short distances ($0.70-0.80 \times r_{eq}$)	MUE	1.56	0.96	2.82	(3.06)	3.15
	MSE	0.25	−0.15	0.26	(−0.06)	0.36
	RMSE	2.32	1.54	4.52	(4.87)	4.89
medium distances ($0.90-1.10 \times r_{eq}$)	MUE	0.39	0.33	0.88	(1.04)	0.68
	MSE	0.06	0.27	−0.33	(−0.58)	0.00
	RMSE	0.61	0.48	1.34	(1.70)	1.10
all distances ($0.70-1.10 \times r_{eq}$)	MUE	0.73	0.51	1.43	(1.62)	1.39
	MSE	0.12	0.15	−0.16	(−0.43)	0.11
	RMSE	1.34	0.92	2.67	(2.97)	2.78

[a]All energy values are in kcal·mol^{-1}. [b]MUE: mean-unsigned error; MSE: mean-signed error; RMSE: root-mean-square error. [c]Values from the W−H and HHG combining rules (in parentheses) for ε_{ij} are provided. [d]Sum of three components where the E_{vdW} was calculated with the W−H combining rule for ε_{ij}.

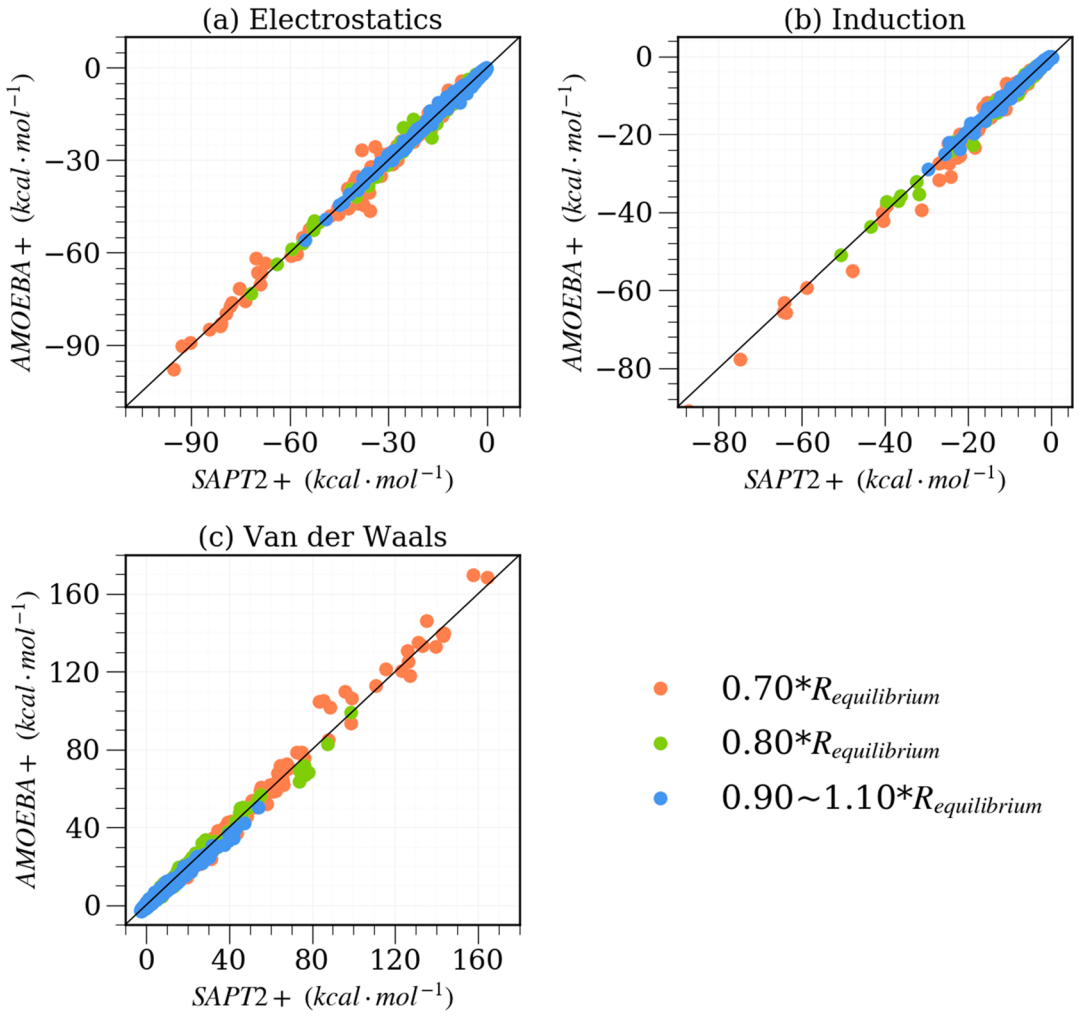


Figure 2. Correlation plot of the intermolecular energy and three decomposed components for the AMOEBA+ model against SAPT2+. (a) Electrostatics energy; (b) induction energy, where sums of E_{CT} and $E_{polarization}$ are served as the AMOEBA+ values, and (c) van der Waals energy, where the AMOEBA+ values are calculated from parameters with the W−H combining rule and SAPT2+ values are sums of $E_{dispersion}$ and $E_{exchange-repulsion}$. Different intermolecular distances are labeled with three different colors (orange, green, and blue). The black solid line shows the perfect correlation.

water, where both the gas-phase cluster energy and various liquid properties are included in parametrization and validation processes.

3.1. Intermolecular Interactions for Organic Molecules. The three nonbonded interactions of AMOEBA+ (electrostatics, CT and vdW) were parametrized separately using the SAPT2+ S108 database as the training set. Polarization parameters were derived separately using the MP2 interaction energy with many-body expansion.[68] As explained above, the sum of AMOEBA+ polarization and CT energy is compared to the SAPT induction energy. The performance of AMOEBA+ for energy components is

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

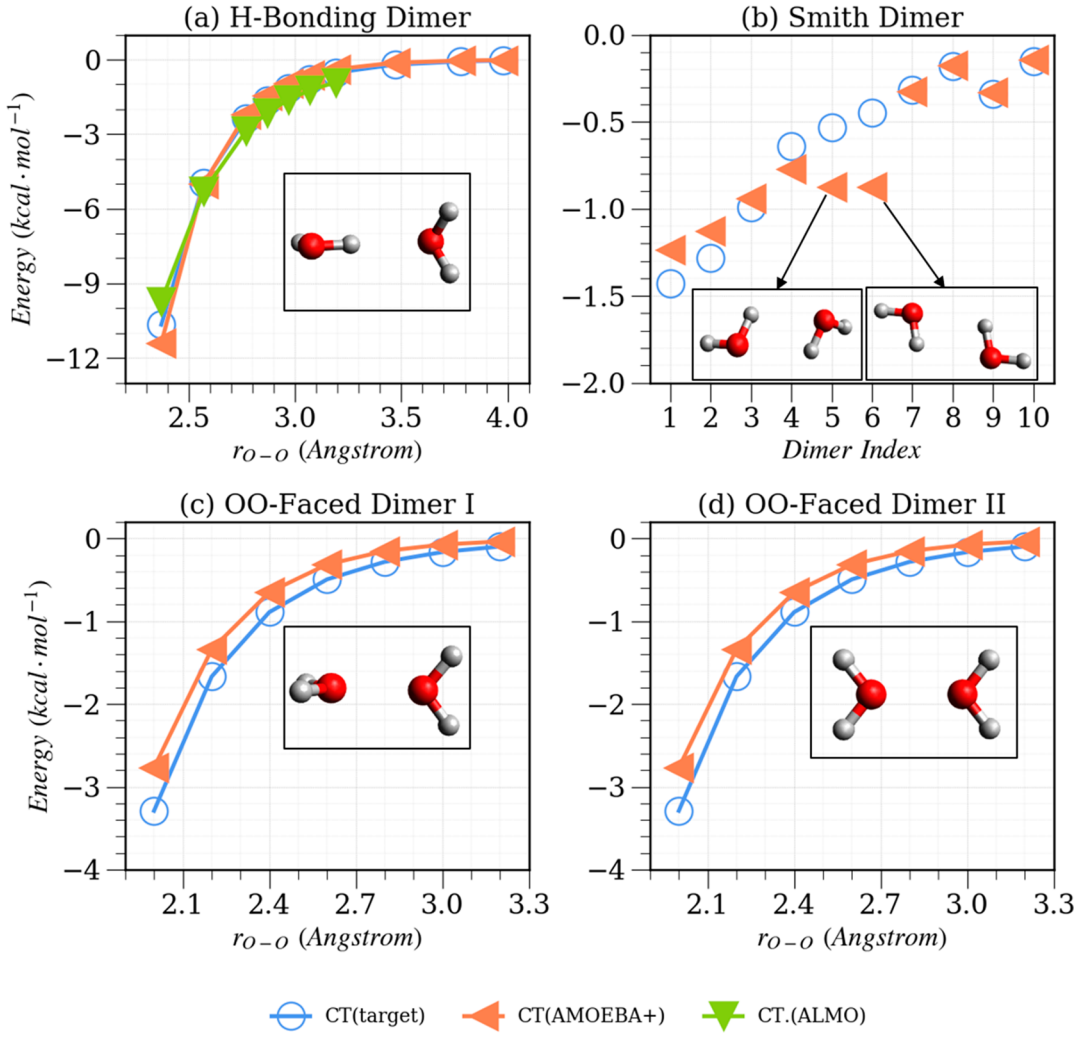


Figure 3. Comparison of the QM and AMOEBA+ CT energy for water dimers of different configurations and intermolecular distances. The target CT energy is defined as $E_{\mathrm{Ind}}^{\mathrm{SAPT2+}} - E_{\mathrm{Pol}}^{\mathrm{AMOEBA+}}$. (a) Hydrogen-bonding water dimer at 10 intermolecular distances, where "CT.(ALMO)" is the energy predicted by the ALMO EDA approach at the HF/6-311++G(2d,2p) level. (b) Smith dimers at various orientations. (c,d) Oxygen−oxygen-faced dimers at seven intermolecular distances.

summarized in Table 3, and the correlations between AMOEBA+ vs SAPT2+ are shown in Figure 2. To validate the transferability, these parameters were further tested on additional SAPT2+ data published by others. In total, there are 707 data points in the training set and 309 in the testing set.

3.1.1. S108×7 Database (Training Set). 3.1.1.1. Electrostatics. The multipole parameters were previously derived based on the DMA fitting procedure.[58] The CP model this work used was previously developed and parametrized on the S101 data set.[60] In this work, we directly applied CP on the S108 data set without reoptimizing the parameters. Using transferable, expanded element-based parameters, the CP-corrected point multipoles are able to reproduce the SAPT2+ electrostatic energy for 108 homo- and heterodimers at 7 intermolecular distances, with an RMSE of 1.34 kcal·mol^{-1} (Table 3). Excellent agreement was found for the medium-distance regions (0.90−1.10 rows in Table 3 and blue dots in Figure 2), with an RMSE of 0.61 kcal·mol^{-1}. It is worth mentioning that for several organic molecules that we studied (not shown here) the CP parameters derived from DMA-based multipoles can also be directly applied on the ISA multipoles without adjusting the CP parameters, which indicates the robustness and transferability of the CP model and parameters.

3.1.1.2. Charge Transfer. Considering the simplicity and isotropic nature of the current CT potential, a possible concern is its capability of capturing the CT energy of different configurations across various chemical environments. We examined the robustness of the model on water and various organic dimers in the S108 data set. Figure 3 shows water dimers of very different configurations at various intermolecular separations. It is encouraging that the absolute errors with respect to the QM target energy, $E_{\mathrm{Ind}}^{\mathrm{SAPT2+}} - E_{\mathrm{Pol}}^{\mathrm{AMOEBA+}}$, are within 0.5 kcal·mol^{-1} in the equilibrium or even shorter distances. Figure 3a shows the excellent agreement between the exponential function used by AMOEBA+ and SAPT2+ for the classic hydrogen-bonding dimer. The AMOEBA+ CT energy is also consistent with that given by the ALMO (absolutely localized molecular orbital)[101,102] EDA method. For the "Smith dimer" configurations (Figure 3b), the two obvious but still small errors appear for the fifth and sixth dimers, which have cyclic structures with the oxygen of one molecule pointing to the hydrogen of the other, suggesting that CT is directional. For the OO-faced dimers (Figure 3c,d), the

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

deviations are *at the most* 0.5 kcal·mol^{-1} at all intermolecular distances. Besides water molecules, we also examined other special cases for several organic dimers in different orientations (Figure S1). Overall, we can conclude that the CT model is robust across these molecules and configurations. We combined the AMOEBA+ CT and polarization energy when comparing with the induction energy of SAPT2+. The agreement is remarkably good even for the short-distance dimers (Table 3 and Figure 2). For all seven intermolecular distances, AMOEBA+ predicted the induction energy of SAPT2+ with an RMSE of 0.92 kcal·mol^{-1}.

3.1.1.3. van der Waals. Here we mainly explore the difference between W−H and HHG combining rules of ε in the buffered-14−7 potential by fitting two sets of vdW parameters to the SAPT2+ repulsion and dispersion energy. Our results indicate that the W−H is slightly better than the HHG combining rule for ε_{ij}, with a reduction of 0.36 and 0.30 kcal·mol^{-1} in RMSE for the medium distances and all seven intermolecular distances, respectively (Table 3). We note that W−H is a more complicated functional form that incorporates the effects of atom radius into ε_{ij}.

Interestingly, when examining the total interaction energy of AMOEBA+, the errors in individual components also partially cancel. AMOEBA+ displays an RMSE of 1.10 kcal·mol^{-1} for the medium-range distances, and the most substantial deviations are in the short distances. The RMSE for all distance points is 2.78, and it will be reduced to 2.55 kcal·mol^{-1} if very repulsive total energy values (>10.0 kcal·mol^{-1}) are excluded.

3.1.2. Validation of AMOEBA+ on an Additional Data Set. We further performed validation of AMOEBA+ using SAPT2+(3) energy decomposition data published by others.[103−107] Molecular dimers in the S59 data set are the same in composition as those in our training set, but only the structures with intermolecular separations not present in S108 are included. The majority of the dimers in the other three sets, S14, Ionic, and X6, include heterodimers not present in S108. Overall, AMOEBA+ performs similarly well on the training and testing sets, with an RMSE of 0.97 kcal·mol^{-1} for the total interaction energy of SAPT2+(3) and <0.7 kcal·mol^{-1} for the individual energy components. The errors on the neutral sets (S14, S59, and X6) are well within 1.0 kcal·mol^{-1}, and those on the charged molecules (Ionic set) are higher (1.2−2.0 kcal·mol^{-1}), as expected (see Table 4).

This set of initial AMOEBA+ parameters for organic molecules were derived purely based on the S108 SAPT2+ data. Due to various limitations in the ab initio methods and approximations in classical models, the resulting potential is unlikely to provide chemical accuracy when applied to condensed-phase simulations. Further optimization using the condensed-phase properties is needed, which requires extensive molecular simulations. As an example, we will next demonstrate the refinement and application of the AMOEBA+ potential to water, where the parameters were further optimized based on selected liquid-phase properties.

3.2. AMOEBA+ Water Model for Gas and Liquid Phases. Using water as an example, we illustrate an approach where we integrate experimental measurements with ab initio data in training AMOEBA+ parameters. In addition, to systematically evaluate different multipole methods (DMA vs ISA) and combining rules for ε in a vdW potential (HHG vs W−H), we optimized four sets of water parameters, i.e., DMA/W−H, ISA/W−H, DMA/HHG, and ISA/HHG combinations. For clarity, hereafter we will report the results from the "DMA/W−H" model and refer it to AMOEBA+ in the remaining sections. The discussion will also be made on the other three models, for which we provided detailed results in the SI.

Table 4. Validation of the Accuracy and Transferability of AMOEBA+ on Additional SAPT Data[a]

database	relative intermolecular distances	data points	RMSE[f] E_{ele}	E_{ind}	E_{vdW}	E_{inter}
S14[b]	1.2, 1.5, 2.0	42	0.65	0.18	0.21	0.82
S59[c]	1.25, 1.5, 2.0	177	0.24	0.14	0.21	0.44
Ionic[d]	1.0, 1.05, 1.1, 1.25, 1.5, 2.0	54	1.21	1.63	1.69	2.00
X6[e]	1.0, 1.05, 1.1, 1.25, 1.5, 2.0	36	0.35	0.05	0.40	0.53
overall	1.0−2.0	309	0.60	0.69	0.74	0.97

[a]RMSEs (in kcal·mol^{-1}) of AMOEBA+ against SAPT2+(3) are reported for individual energy components. [b]Subset of S22×5.[103] [c]Subset of S66×8 with 1.2 and above distances.[105] [d]Subset of "Ionic" with 1.0 and above distances.[106] [e]Subset of X40 with 1.0 and above distances.[104] [f]SAPT2+(3) energy decomposition data are taken from ref 107.

Parameters of the AMOEBA+ water model are shown in Table 5. The initial nonbonded parameters were derived within the S108 data set, as described above, and the initial bonded parameters were taken from the AMOEBA03 water model.[16] The multipoles in S108 were first derived using DMA at the MP2/6-311G** level of theory. The dipole and quadrupole parameters were further optimized by fitting to the ESP at the MP2/aug-cc-pvtz level.[58] This procedure resulted in a partial charge of −0.38 e on oxygen. The atomic multipoles in AMOEBA03 were generated via DMA at the MP2/aug-cc-pvtz with the experimental monomer geometry, with a partial charge of −0.52 e on oxygen.[16] Interestingly, the partial charge on oxygen for the AMOEBA+ model, after FB optimization, converged to that of the AMOEBA03. Please note that the atomic multipoles could not be uniquely determined. DMA and ISA methods could lead to dramatically different atomic multipole values, and yet both produce accurate ESP around the molecules, at least at the far distances (the deficiency in the short distances is addressed by CP correction). As will be shown next, both the initial and optimized electrostatic parameters gave electrostatic interaction energy consist with that of SAPT decomposition. In addition, the vdW parameters changed slightly from those initial values derived by fitting to the SAPT repulsion−dispersion energy and so did the valence parameters. This again shows that to achieve the desired chemical accuracy in the condensed phase we need to incorporate experimental measurements in classic potential parameter optimization. To understand the sensitivity of liquid properties to these parameters, we analyzed the final numerical gradient values of the six thermodynamic properties with respect to each parameter (Tables S17 and S18) and found that (1) almost all of the properties are sensitive to the vdW and multipole parameters; (2) the static dielectric constant is strongly affected by atomic dipole moments, which is not surprising as the static dielectric constant is calculated from the cell dipole fluctuation (see the following section); (3) isothermal compressibility is sensitive to the equilibrium OH bond length; and (4) CP parameters have a strong effect on liquid density.

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

Table 5. Initial and Optimized Parameters of the AMOEBA+ Water Model

term	parameter	unit	initial	optimal
multipole	O monopole	e	−0.382800	−0.558246
	O dipole Z	e·bohr	0.054770	−0.144923
	O quadrupole XX	e·bohr^2	0.698660	0.451599
	O quadrupole YY	e·bohr^2	−0.604710	−0.280108
	O quadrupole ZZ	e·bohr^2	−0.093950	−0.171491
	H monopole	e	0.191400	0.279123
	H dipole Z	e·bohr	−0.200970	−0.230060
	H quadrupole XX	e·bohr^2	0.038810	0.215207
	H quadrupole YY	e·bohr^2	0.022140	−0.029761
	H quadrupole ZZ	e·bohr^2	−0.060950	0.185446
	H quadrupole XZ	e·bohr^2	0.000000	0.191100
CP	O damping factor	none	4.1615	4.0483
	H damping factor	none	3.2632	3.2748
CT	O parameter a_{CT}	10^3 kcal·mol^{-1}	3.4761	3.2003
	O parameter b_{CT}	Å^{-1}	3.6034	3.7188
	H parameter a_{CT}	10^3 kcal·mol^{-1}	3.7994	2.9436
	H parameter b_{CT}	Å^{-1}	4.8850	4.7135
polarization	O polarizability	Å	0.837	0.948
	H polarizability	Å	0.496	0.416
	direct damping factor	none	0.70	0.70
	mutual damping factor	none	0.39	0.39
vdW	O vdW diameter	Å	3.813189	3.808992
	O vdW epsilon	kcal·mol^{-1}	0.084785	0.061361
	H vdW diameter	Å	3.339858	3.340781
	H vdW epsilon	kcal·mol^{-1}	0.002449	0.004571
	H vdW reduction	none	0.980000	0.983604
bonded	O−H bond length	Å	0.96	0.94
	bond force constant	kcal·mol^{-1}·Å^{-2}	556.85	556.85
	H−O−H angle	degree	108.50	108.81
	angle force constant	kcal·mol^{-1}·rad^{-2}	48.70	48.70
	U−B H−H length	Å	−7.60	−7.60
	U−B force constant	kcal·mol^{-1}·Å^{-2}	1.53	1.53

Table 6. Monomer Properties of AMOEBA+ Compared to Those of AMOEBA14 and Experimental Values[a]

property		AMOEBA14[b]	AMOEBA+[c]	ab initio[d]	expt.
dipole (Debye)	d_z	1.808	1.778 (1.918)	1.840	1.855[e]
quadrupole (Debye·Å)	Q_{xx}	2.626	3.314 (3.193)	2.57	2.63[f]
	Q_{yy}	−2.178	−2.628 (−2.679)	−2.42	−2.50[f]
	Q_{zz}	−0.045	−0.684 (−0.514)	−0.14	−0.13[f]
polarizability (Å^2·s^4·kg^{-1})	α_{xx}	1.767	1.583 (1.589)	1.47	1.53[g]
	α_{yy}	1.308	1.209 (1.213)	1.38	1.42[g]
	α_{zz}	1.420	1.308 (1.329)	1.42	1.47[g]

[a]AMOEBA+ values are given for both model-optimized and "ideal" geometries, with the latter given in parentheses. [b]Reference 69. [c]In parentheses are the values calculated with the "ideal" geometry, where the O−H bond length is 0.9572 Å and the angle ∠HOH is 104.52°. [d]Reference 110. [e]Reference 111. [f]Reference 112. [g]Reference 113.

3.2.1. Gas-Phase Cluster Properties. 3.2.1.1. Monomer and Equilibrium Dimer. Multipole moments are the most basic properties of a water molecule and are largely responsible for interesting water properties across multiple phases.[108] In Table 6, AMOEBA+ molecular multipole moments calculated at the model-optimized geometry and an "ideal" water geometry determined by rotation−vibration spectra of water vapor[109] are compared with experimental values and those from a previous AMOEBA14 model. With the model-optimized geometry, AMOEBA+ gives a total dipole moment of 1.778 D, lower than the ab initio and experimental values, mostly due to the larger HOH angle. The molecular dipole moment increases to 1.918 D when the "ideal" geometry is used. AMOEBA+ has larger quadrupole moments but smaller polarizability compared to those of AMOEBA14. In Table 7, the dimer properties at equilibrium geometry are given (see also Figure 4). The dimer dissociation energy of AMOEBA+ agrees better with the ab initio value than AMOEBA14, which is also true for other separation distances, as shown in Figure 5a. The total dipole moment of the water dimer calculated with AMOEBA+ based on the MP2-optimized geometry agrees well with both experimental and ab initio values. It is noted that the water models employing DMA multipoles better predict the monomer and equilibrium dimer properties than those using ISA multipoles, using either HHG or W−H vdW mixing rules (Tables S10 and S11).

3.2.1.2. Energy Decomposition Analysis. The primary goal of the AMOEBA+ potential over the current AMOEBA is to

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

Table 7. Dimer Properties at Equilibrium Geometry Predicted by AMOEBA+ Compared to AMOEBA14 and Experimental Values[a]

property	AMOEBA14[b]	AMOEBA+	ab initio	expt.
D_e	4.64	4.80	4.98[c]	5.44 ± 0.7[e]
r_{O-O}	2.908	2.892	2.907[c]	2.976[f]
α	4.61	14.38	5.5[c]	−1 ± 10[f]
θ	64.9	79.7	56.9[c]	57 ± 10[f]
μ_{tot}	2.20	1.79 (2.85[g])	2.76[d]	2.643[f]

[a] D_e (in kcal·mol^{-1}) is the dissociation energy with the monomer optimized. μ_{tot} (in Debye) is the total dipole moment. Other geometrical properties are illustrated in Figure 4. The optimized dimer geometry was used for AMOEBA+, unless explicitly noted. [b]Reference 69. [c]Reference 114. [d]Reference 115. [e]Reference 116. [f]Reference 117. [g]Value calculated using the AMOEBA+ model with the MP2-optimized geometry.

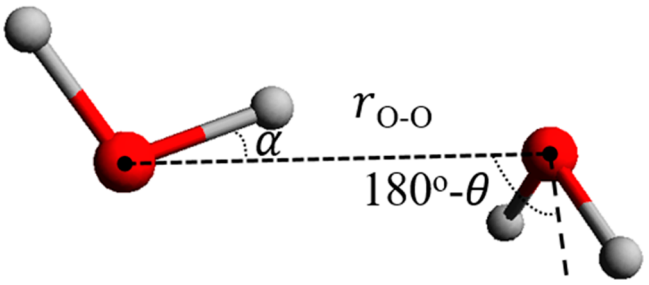


Figure 4. Illustration of geometrical parameters for the equilibrium water dimer. Angle θ is the angle between the extended O···O vector and the bisector of the ∠HOH angle of the hydrogen-bond acceptor molecule.

improve the physical representation of individual energy components, reduce error cancelation, and improve the transferability. Figure 5 shows the comparison of the hydrogen-bonding water dimer interaction energy and its components from AMOEBA+ and SAPT2+, using the final (liquid) optimized AMOEBA+ parameters. Here we defined the intermolecular interaction energy at the cluster geometry without relaxing the monomers. The interaction energy at the CCSD(T)/aug-cc-pv5z level of theory was also calculated and is provided here. Even with liquid optimization, the electrostatics, induction, and vdW energies are still well reproduced in all distances with respect to the SAPT2+ EDA. The errors are noticeable mostly at very close contact. At a O···O distance of 2.37 Å, the errors are +0.4, + 0.9, and +3.0 kcal·mol^{-1} for the three components, respectively. The total dimer interaction energy values from two ab initio methods, SAPT2+ and CCSD(T), show nontrivial disagreements at short distances (O···O distances < 2.77 Å), as shown in Figure 5b. The AMOEBA+ total interaction energy closely follows those of SAPT2+ and CCSD(T) at the equilibrium distance and beyond. It is also clear that AMOEBA14 underestimates the dimer interaction energy for all distances. The deviation between the AMOEBA+ and SAPT2+ vdW energy at the very short distance (e.g., 2.37 Å) may imply the uncertainty of exchange−repulsion by SAPT2+, which is also less repulsive than CCSD(T). Additional validation on the Smith05 dimer, which possesses cyclic geometry with nontypical O···H interactions, and the hydrogen-bonding dimer with various flap angles shows good transferability of the AMOEBA+ water model in predicting interaction energy components (Figure S2).

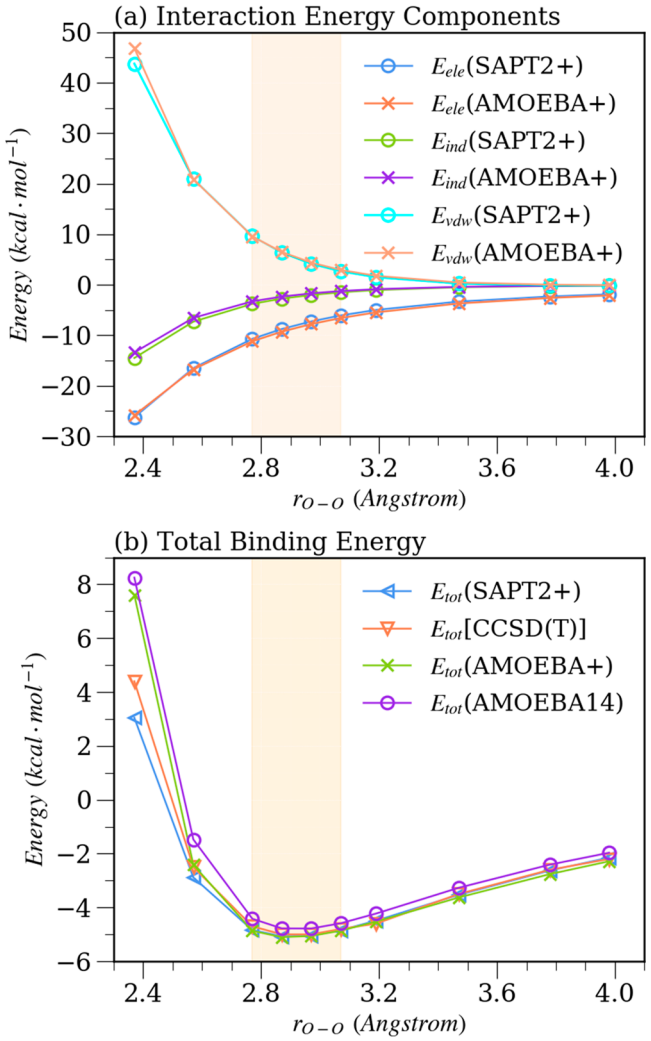


Figure 5. Intermolecular interaction energy of the hydrogen-bonding dimer at 10 separation distances. (a) Energy components from SAPT2+ and AMOEBA+ models. (b) Total interaction energy calculated by two QM methods compared to AMOEBA+ and AMOEBA14 water models. The orange shadows indicate the equilibrium distance region of the water dimer.

3.2.1.3. Binding Energy of Water Clusters. The binding energy (BE) of a molecular cluster is defined as the energy required to separate each monomer from the cluster, with the cluster and monomer optimized to their own minimum energy. The AMOEBA+ BEs of clusters ranging from dimers to 20-mers are compared with high-level QM results as well as AMOEBA14 in Tables 8 and 9. Compared to the AMOEBA14 model, AMOEBA+ offers better BEs for *eight out of ten* Smith dimers. The RMSE of the AMOEBA+ water dimer BE with respect to CCSD(T) is reduced from 0.62 kcal·mol^{-1} of AMOEBA14 to 0.28 kcal·mol^{-1}. The Smith dimer set is usually thought of as a test for "anisotropy" of a water model as different H-bond configurations are present. Both AMOEBA+ and AMOEBA14 perform similarly well on clusters from trimers to octamers. However, AMOEBA+ does noticeably better than AMOEBA14 for almost all larger clusters from 11-mers to 20-mers. Overall, an improvement of ∼0.6 kcal·mol^{-1} in terms of the RMSE is achieved by AMOEBA+ over the AMOEBA14 model (Table 9). It is interesting that both DMA/W−H and DMA/HHG models predict the BEs of Smith

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

Table 8. BE of Smith Dimers Predicted by AMOEBA+ and AMOEBA14 Models Compared to QM Data at CCSD(T)/CBS[a]

$(H_2O)_2$	CCSD(T)[b]	AMOEBA+	AMOEBA14[c]
Smith01	−4.97	**−4.96**	−4.65
Smith02	−4.45	−4.11	**−4.22**
Smith03	−4.42	−4.00	**−4.19**
Smith04	−4.25	**−4.75**	−3.54
Smith05	−4.00	**−4.08**	−3.06
Smith06	−3.96	**−3.90**	−2.92
Smith07	−3.26	**−3.69**	−2.49
Smith08	−1.30	**−1.39**	−1.02
Smith09	−3.05	**−3.22**	−2.37
Smith10	−2.18	**−2.34**	−1.96
RMSE		**0.28**	0.62

[a]Deviations (RMSE) from CCSD(T) data are reported for two water models. Values in bold indicate better agreement to CCSD(T) reference. All energy values are in kcal·mol^{-1}. [b]Reference 126. [c]Reference 69.

Table 9. BE of Trimer to 20-mer Water Clusters Predicted by AMOEBA+ and AMOEBA14 Models Compared to QM Data, Where the BEs of Small Clusters (n = 3−6) Were Calculated at CCSD(T)/CBS and the BEs of Large Clusters (n > 6) Were Calculated at MP2/CBS[a]

$(H_2O)_n$	geometry	QM	AMOEBA+	AMOEBA14[h]
n = 3[b]	cyclic	−15.74	**−16.07**	−15.38
n = 4[b]	cyclic	−27.40	−28.26	**−27.43**
n = 5[b]	cyclic	−35.93	−36.41	**−35.78**
n = 6[c]	prism	−45.92	**−46.03**	−45.18
	cage	−45.67	−46.10	**−45.83**
	bag	−44.30	−44.65	**−44.52**
	cyclic chair	−44.12	−44.88	**−43.53**
	book1	−45.20	−45.87	**−45.08**
	book2	−44.90	−45.25	**−45.06**
	cyclic boat1	−43.13	−43.54	**−42.99**
	cyclic boat2	−43.07	−43.51	**−43.07**
n = 8[d]	S4	−72.70	−73.56	**−72.22**
	D2d	−72.70	−73.74	**−72.24**
n = 11[e]	434	−105.72	**−101.65**	−101.11
	515	−105.18	**−101.54**	−100.99
	551	−104.92	**−101.23**	−100.58
	443	−104.76	**−101.55**	−101.17
	4412	−103.97	**−100.94**	−100.33
n = 16[f]	boat-a	−170.80	**−162.50**	−160.45
	boat-b	−170.63	**−162.09**	−160.30
	antiboat	−170.54	**−161.77**	−160.30
	ABAB	−171.05	**−163.59**	−161.20
	AABB	−170.51	**−163.19**	−160.89
n = 17[f]	sphere	−182.54	**−172.24**	−171.53
	5525	−181.83	**−171.00**	−170.42
n = 20[g]	dodecahedron	−200.10	−193.58	**−193.81**
	fused cubes	−212.10	**−208.65**	−205.77
	face sharing prisms	−215.20	**−205.31**	−204.41
	edge sharing prisms	−218.10	**−208.53**	−207.06
RMSE			**5.46**	6.33

[a]Deviations from QM data are reported for two water models. Values in bold indicate better agreement to QM reference. All energy values are in kcal·mol^{-1}. [b]Reference 16. [c]Reference 118. [d]Reference 119. [e]Reference 120. [f]Reference 121. [g]Reference 122. [h]Reference 69.

dimers better than ISA/W−H and ISA/HHG models (Table S14). Considering the greater monopole and smaller quadrupole in the ISA-based multipoles, this could suggest that ISA multipoles lack in anisotropy in comparison with DMA. Nonetheless, all four water models, regardless of the multipole methods and vdW combining rules, all give better BEs than AMOEBA14 for large clusters (Table S15).

3.2.2. Liquid Properties. Thermodynamic properties of water calculated by the AMOEBA+ model include density, enthalpy of vaporization, thermal expansion coefficient, isothermal compressibility, static dielectric constant, and isobaric heat capacity. Structural (radial distribution function (RDF)) and dynamic properties (self-diffusion constant) that were not included as the parametrization targets but only used as validation are also presented.

3.2.2.1. Density. One of the anomalous properties of water is its maximum density at 4 °C. The density−temperature profile simulated by the AMOEBA+ model shows excellent agreement with experiment, with small deviations up to 0.6% at very high temperature (369 K). Compared to the AMOEBA14 model, AMOEBA+ shows better agreement with experiment in the near-density-maximum temperatures and similar agreement in the ambient temperature (numerical data shown in the SI). For example, the density predicted by AMOEBA+ at 277 K is 1000.0 kg·m^{-3} (see Figure 6), in excellent agreement with the

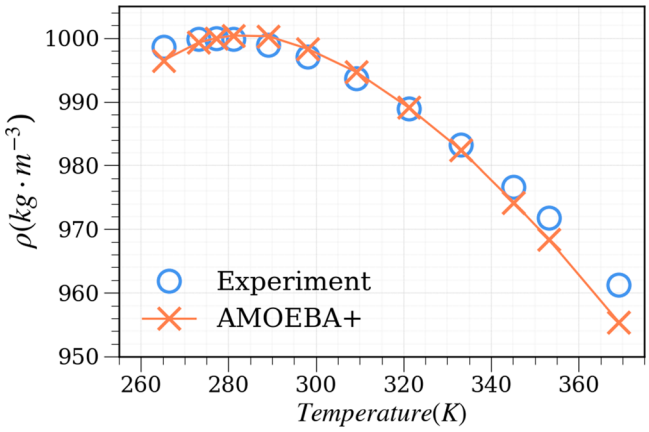


Figure 6. Liquid water density at temperatures ranging from 265 to 369 K and under atmospheric pressure (1 atm) calculated with the AMOEBA+ model. The standard deviations for the simulated density are 0.15−0.47 kg·m^{-3} and not shown on the plot. Experimental data are taken from ref 123.

experimental value (1000.0 kg·m^{-3}), and AMOEBA14 gives 1001.5 kg·m^{-3}. At ambient temperature (298 K), the average density by AMOEBA+ is 998.3 kg·m^{-3}, and AMOEBA14 gives 997.9 kg·m^{-3}. We note that a larger time step (2 fs) results in an average density of 997.7 kg·m^{-3} by AMOEBA+, compared to the experimental density of 997.0 kg·m^{-3}. The other three models parametrized in this work (DMA/HHG, ISA/W−H, ISA/HHG) show similar performance (Table S3).

3.2.2.2. Enthalpy of Vaporization. The enthalpy of vaporization is computed from the potential energy difference between liquid and gas phases, assuming water vapor to be an ideal gas

$$\Delta H_{vap} = -\Delta E + \Delta PV = -E_{liq} + E_{gas} + RT \quad (18)$$

The potential energy of the gas is obtained from the stochastic dynamics simulation of a monomer at specific temperatures

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

using a time step of 0.1 fs. The AMOEBA+ ΔH_{vap} at 298 K is 44.5 kJ·mol^{-1} (see Figure 7), in agreement with the

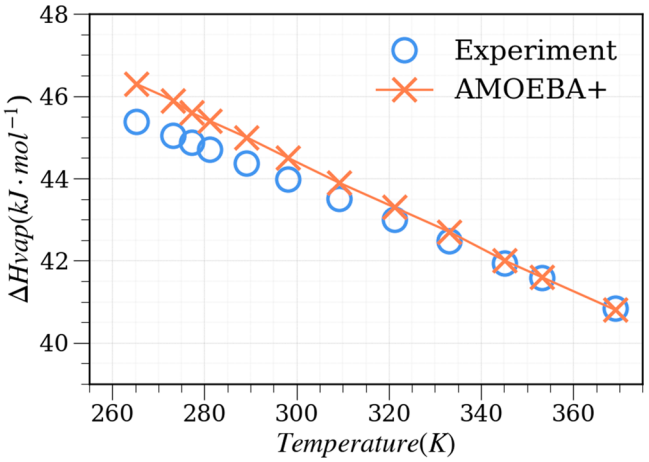


Figure 7. Liquid water enthalpy of vaporization at temperatures ranging from 265 to 369 K and under atmospheric pressure (1 atm) calculated with the AMOEBA+ model. The standard deviations for the simulated ΔH_{vap} are 0.03−0.06 kJ·mol^{-1} and not shown on the plot. Experimental data are taken from ref 124.

experimental value (44.0 kJ·mol^{-1}). The ΔH_{vap} values by AMOEBA+ and AMOEBA14 are within 0.1 kJ·mol^{-1}, and both are slightly larger (+0.8−0.9 kJ·mol^{-1}) at low temperature (265 K) than the experimental value. While the absolute value of ΔH_{vap} reflects the magnitude of the interaction energy of water molecules and has a strong influence on the self-diffusion constant of water (see later discussion), the slope of the $\Delta H_{vap}-T$ curve is also important as it defines the heat capacity.

3.2.2.3. Thermal Expansion Coefficient. The thermal expansion coefficient measures the volume change of water in response to the temperature. It was calculated using the fluctuation formula

$$\alpha = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_{p,N} = \frac{1}{k_B T^2}\frac{\langle HV\rangle - \langle H\rangle\langle V\rangle}{\langle V\rangle} \tag{19}$$

AMOEBA+ α values are consistent with the experimental data at temperatures below 300 K but begin to deviate as the temperature rises (Figure 8). The transition of α from negative to positive follows the same trend as the experiment.

3.2.2.4. Isothermal Compressibility. The isothermal compressibility measures the volume change of water in response to the pressure. It was calculated from the fluctuation equation

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_{T,N} = \frac{1}{k_B T}\frac{\langle V^2\rangle - \langle V\rangle^2}{\langle V\rangle} \tag{20}$$

To compare with experimental κ_T at selected temperatures, we ran the MD simulations using the AMOEBA+ water final parameters. Here we used a small water box (216 molecules) and the Nosé−Hoover integrator in Tinker (CPU) code to calculate the κ_T values, which again agree well with the experimental measurements at low temperatures but are overestimated above 290 K (see Figure 9).

3.2.2.5. Static Dielectric Constant. The static dielectric constant (ε_0) is primarily determined by the electrostatics and polarization component of the water model (Table S16). The

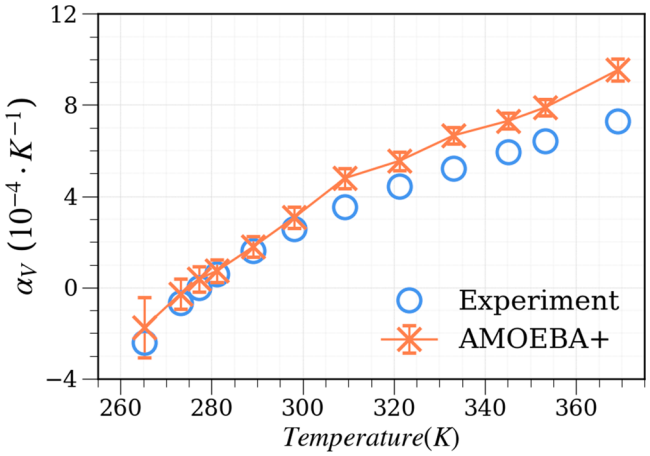


Figure 8. Liquid water thermal expansion coefficient at temperatures ranging from 265 to 369 K and atmospheric pressure (1 atm) calculated with the AMOEBA+ model. These values are obtained from a small box simulation with a Nosé−Hoover integrator. Experimental data are taken from ref 123.

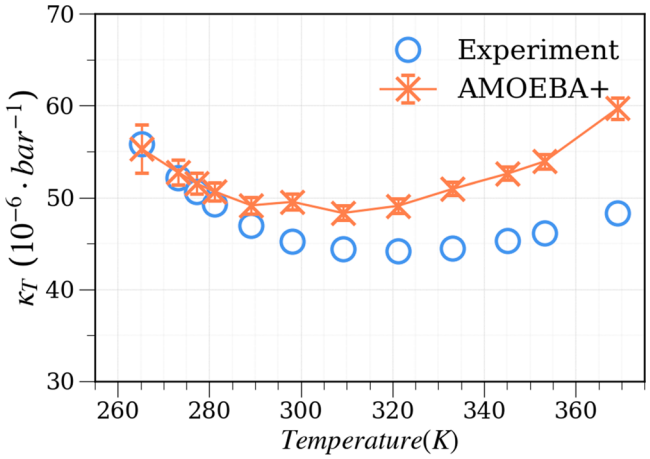


Figure 9. Liquid water isothermal compressibility at temperatures ranging from 265 to 369 K and under atmospheric pressure (1 atm) calculated with the AMOEBA+ model. Experimental data are taken from ref 123.

ε_0 is calculated from the fluctuation of the total dipole moment of the simulation box

$$\varepsilon_0 = 1 + \frac{4\pi}{3k_B T\langle V\rangle}(\langle M^2\rangle - \langle M\rangle\cdot\langle M\rangle) \tag{21}$$

where $\langle M\rangle$ in the ensemble is the average of the total box dipole moment; V is the volume of the simulation box, and T is the temperature. The simulated ε_0 at all temperatures matches reasonably well with the experimental measurements, with the uncertainties increase quickly for low temperatures due to slow dynamics (Figure 10). The ε_0 for AMOEBA+ is 80.4 at 298 K and under 1 atm, whereas the experimental measurement is 78.4.

3.2.2.6. Isobaric Heat Capacity. The isobaric heat capacity is related to liquid enthalpy fluctuation via

$$C_p = \frac{1}{Nk_B T}(\langle H_{liq}{}^2\rangle - \langle H_{liq}\rangle^2) \tag{22}$$

In the above equation, N is the number of molecules in the simulation box, k_B is the Boltzmann constant, and T is

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

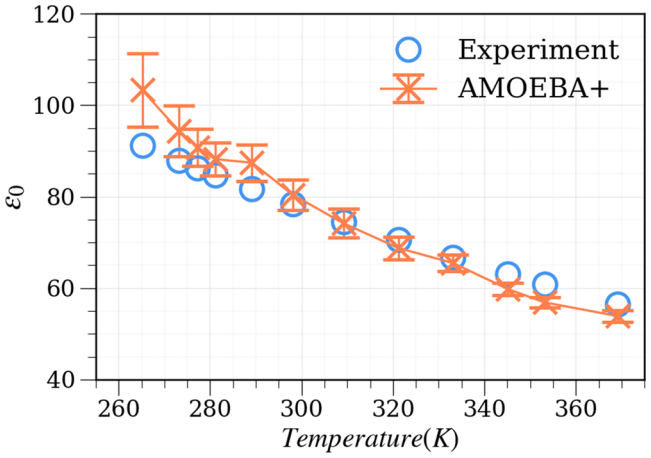


Figure 10. Liquid water static dielectric constant at temperatures ranging from 265 to 369 K and under atmospheric pressure (1 atm) calculated with the AMOEBA+ model. Experimental data are taken from ref 124.

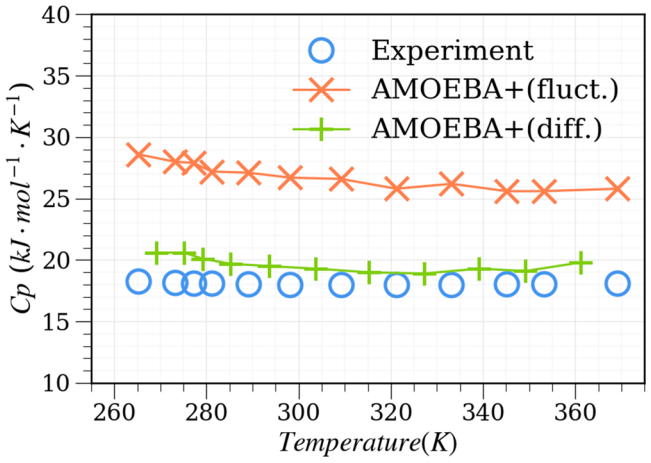


Figure 11. Liquid water heat capacity at temperatures ranging from 265 to 369 K and under atmospheric pressure (1 atm) calculated with the AMOEBA+ model. C_p values were calculated using both the fluctuation formula and differential equation. Quantum corrections were included in both approaches. The standard deviations for the C_p calculated by the fluctuation formula are 0.6−1.4 cal·mol^{-1}·K^{-1} and not shown in the plot. Experimental data are taken from ref 124.

temperature. It is noted that classical approximation of the intra- and intermolecular vibration overestimates the heat capacity compared with a quantum oscillator model. Thus, the temperature-dependent inter- and intramolecular corrections are usually necessary when comparing simulations with experiments.[125] Another approach for calculating C_p independent of fluctuation is to use the differentiation equation[126]

$$C_p = \left(\frac{\partial\langle H_{liq}\rangle}{\partial T}\right)_P = \left(\frac{\partial\langle U_{liq}\rangle}{\partial T}\right)_P + \frac{9}{2}R \quad (23)$$

where the first term is the differentiation of the liquid potential energy with respect to the temperature and the second term is the kinetic energy contribution from nine degrees of freedom. Using eq 23, we simply calculated C_p of a certain temperature from the Δ_{vap} values of two neighboring temperatures. At 298.15 K, the C_p calculated by the differential equation, after addition of quantum corrections (in the range of −2.5 to −1.6 cal·mol^{-1}·K^{-1}),[125] is ∼19.4 cal·mol^{-1}·K^{-1}, in closer agreement with the experimental value (18.0 cal·mol^{-1}·K^{-1}; see Figure 11).

3.2.2.7. Radial Distribution Function. The experimental RDFs are not included in our parametrization; thus, this comparison serves as a test of our model. RDFs in the experiment can be derived from X-ray scattering[127] or neutron diffraction[128] techniques and reflect the structural features of liquid water. An ideal water model is expected to reproduce the RDF peak positions and the intensity of the experimental RDFs because many of water's properties are encoded in its structure.[129] Figure 12 provides RDFs of O···O, O···H, and H···H pairs from AMOBA+ simulation at 298 K and under 1 atm. The first peak of $g_{OO}(r)$ from AMOEBA+ resides at 2.75 Å (Figure 12a), in between the values derived from neutron diffraction (2.73 Å)[128] and X-ray scattering (2.80 Å) experiments.[127] The intensity of the first and second peaks as well as the position of the second peak of AMOEBA+ $g_{OO}(r)$ agrees slightly better with the neutron diffraction than the X-ray scattering data. The higher first peak of $g_{OO}(r)$ by AMOEBA[16,69,95] and AMOEBA+ water models may be a result of the buffered-14−7 vdW functional form. A recent study shows that the exponential Buckingham potential has advantage over Lennard-Jones in describing RDFs of water by

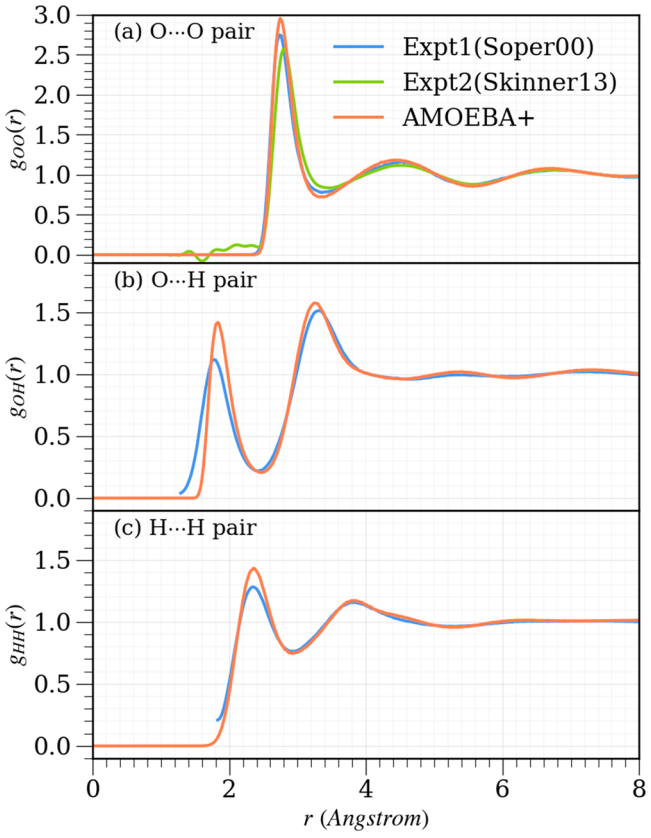


Figure 12. AMOEBA+ predicted RDFs at 298.15 K and 1 atm compared to experimental measurement. (a) O···O pair, (b) O···H pair, and (c) H···H pair. RDFs between intramolecular pairs were not calculated, and the experimental data below 1.65 (b) and 1.80 Å (c) are not shown on the plot.

explicit fitting using FB.[130] Nonetheless, our previous work also showed that the Buckingham potential has difficulty capturing the SAPT2+ exchange−repulsion energy.[131] Features beyond the first peaks are also consistent with the neutron diffraction curves, which is essential because the tetrahedral structure of

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

water is characterized by the second and third peaks.[128] AMOEBA+ predicts the first and second peaks of $g_{OH}(r)$ at 1.84 and 3.25 Å, respectively, vs the experimental data (1.77 and 3.33 Å). The positions of the first and second peaks of $g_{HH}(r)$ (2.36 and 3.85 Å) predicted by AMOEBA+ are in excellent agreement with the neutron scattering measurement (2.34 and 3.84 Å). As seen from Figure 12c, the height of the second peak from AMOEBA+ overlaps with the experiment.

3.2.2.8. Self-Diffusion Constant. The self-diffusion constant (D_0) was not used as a parametrization target; thus, this comparison serves as another validation. It reflects the dynamic feature of water, which is crucial for kinetics processes such as chemical reactions in solution. It was evaluated with the Einstein equation

$$D_0 = \lim_{t\to\infty} \frac{\mathrm{d}}{\mathrm{d}t}\langle |r(t) - r(t_0)|^2 \rangle \tag{24}$$

It has been shown that the D_0 could be underestimated in a limited size of water box in simulations. To correct the size effect, we computed the D_0 with cubic boxes of 18, 40, 60, and 90 Å, where N_{water} is 216, 2210, 7500, and 25150, respectively. The time length of *NPT* production MD simulations are 6, 5, 4, and 2 ns, respectively. The size-independent D_0 was obtained by a linear fit to $D \approx \frac{1}{L}$ and extrapolating to $\frac{1}{L} = 0$. As seen from Figure 13, the AMOEBA+ water D_0 matches very well

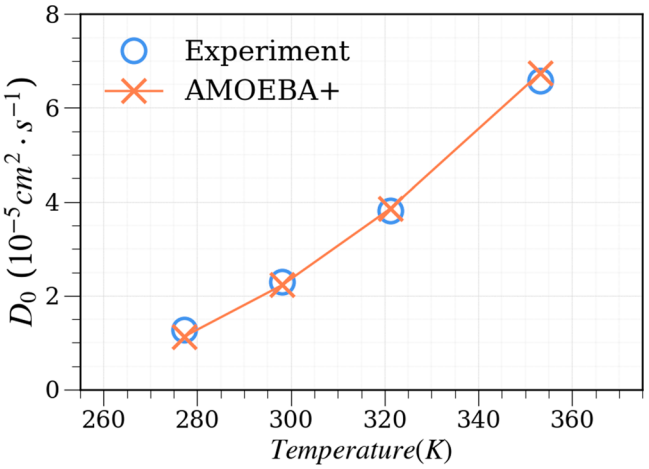


Figure 13. Size-corrected water self-diffusion constant as a function of temperatures under 1-atm pressure compared to experiment. Detailed extrapolation over different system sizes is provided in Figures S3–S6.

with the experimental data. AMOEBA+ D_0 at room temperature and under 1 atm is 2.23 × 10^5 $cm^2 \cdot s^{-1}$, while the corresponding experimental value is 2.30 × 10^5 $cm^2 \cdot s^{-1}$. This agreement is better than widely used nonpolarizable water models, such as SPC[141] and TIP*n*P (n = 3–5),[132] as well as polarizable models, such as AMOEBA models[16,69] and AMOEBA variants.[95,96] As discussed above, the W–H vdW combining rule performs slightly better than HHG capturing the vdW component of the SAPT2+ energy for the organic S108 sets (Table 3). It is seen here that seemingly small differences in potential energy can have noticeable impacts on the dynamic and thermodynamic property of liquid water. Due to less repulsive energy that the HHG rule provides, HHG-based water models underestimate D_0 of water at a series of temperatures. For example, D_0 with the HHG combining rule, with either DMA- or ISA-based multipoles, is systematically slower than the experimental value (298 K) by 8.0–12.6% (Figures S4 and S5). In contrast, models employing the W–H rule are able to predict appropriate D_0 for various temperatures.

3.2.2.9. Summary of AMOEBA+ and AMOEBA Water Models. The first AMOEBA water model was published in 2003, and the parametrization relied particularly on high-level ab initio data of small water clusters and limited liquid properties (density and heat of vaporization) at only room temperature.[16] It was encouraging that various bulk properties can be reasonably reproduced by the AMOEBA03 water model even beyond the ambient conditions.[42] The AMOEBA14 model was a parameter optimization effort explicitly utilizing a wide range of liquid properties at various temperatures, with every single parameter allowed to be optimized. Nevertheless, the gas-phase properties were sacrificed somewhat during the optimization. More importantly, the physical meaning of the parameter, e.g., atomic charges, after optimization becomes unclear. When mixing this water model with other molecules, the "electrostatic" interactions may no longer be truly electrostatic and errors summing over different energy components may be amplified instead of canceled. To address this challenge, we have developed a new potential energy function that allows us to systematically examine and model the individual components of intermolecular forces by explicitly accounting for CP, CT, and many-body dependence of polarization effects using relatively simple, computationally tractable empirical functions. The resulting AMOEBA+ water model shows comparable performance on liquid properties and improvements on gas-phase cluster properties over the AMOEBA14 model. Most importantly, the AMOEBA+ water potential also provides interaction energy components that are consistent with the SAPT method. By extending this approach from water to other chemical species, we will be able to arrive at a general and transferable FF for molecular simulations.

4. CONCLUSIONS

Classical potentials based on isotropic and additive atomic interactions have been widely used over the past few decades to simulate small molecules to large proteins in computers. While computationally attractive, it is well understood that severe approximations are made in the underlying physics. The inaccuracy can be hidden due to error cancellation in overparameterization or insufficient sampling in molecular simulations. To fundamentally advance the accuracy and transferability of classical potentials, we propose a new classical potential, AMOEBA+, to model the essential intermolecular forces, including permanent electrostatics, repulsion and dispersion, as well as many-body polarization and short-range CP and CT, by extending the polarizable atomic multipole-based AMOEBA model. The key improvements are the inclusion of short-range CP and CT effects and modification of direct polarization damping and vdW combining rules. We adopted a unique approach to separate the CT energy out of the SAPT induction energy by first determining the many-body polarization energy independent of ab initio EDA such as SAPT. The AMOEBA+ framework was then successfully demonstrated on a set of common organic molecules, where we showed that a classical potential with limited, element-based parameters reproduced SAPT2+ intermolecular energies and their components for homo- and heterodimers at various configurations and separations. The accuracy and transferability of the AMOEBA+ potential were further validated on the testing database generated by other researchers. Direct

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122–4139

application of the ab initio potential to the condensed phase is limited by the accuracy of QM methods, as well as approximations and simplifications made in classical potentials. Thus, the most effective approach to extend ab initio potentials to the condensed phase is to directly incorporate accurate liquid thermodynamic properties that are widely available. Following this strategy, we derived a new water model based on the AMOEBA+ framework by incorporating both ab initio data and experimental thermodynamic properties. The optimization against liquid properties led to subtle changes in the parameters; however, the agreement with initial gas-phase SAPT2+ EDA remains excellent for a relatively simple classical potential. To develop a general AMOEBA+ FF beyond water, we will extend this approach to other molecular species in the future, where the AMOEBA+ parameters derived in this study will be further refined by using additional cluster QM data and well-established experimental properties. This will require extensive molecular simulations that can be achieved by using high-performance Tinker simulation platforms, including Tinker-HP[89] and Tinker-OpenMM.[88] We believe that the new classical mechanics model and approach presented in this work has the potential to fundamentally advance the general applicability of classical FFs.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.9b00261.

Atomic multipoles, atomic polarizabilities, CP, CT, and vdW parameters of AMOEBA+ for organic molecules, validation of the CT model on organic dimers with different orientations, energy components for other water dimers, detailed extrapolation plot to obtain the size-corrected self-diffusion constants for four water models, numerical results of gas and liquid phase properties obtained with four water models, analysis of parameters' sensitivity to six liquid properties, and complete parameters of four AMOEBA+ water models (PDF)

Calculated interaction energy for the S108 set and calculated interaction energy for the testing sets (XLSX)

vdW parameters for organic molecules and Cartesian coordinates of organic S108 and testing set dimers (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: pren@mail.utexas.edu.

ORCID

Chengwen Liu: 0000-0002-3930-7793

Jean-Philip Piquemal: 0000-0001-6615-9426

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful for support from the National Institutes of Health (R01GM106137 and R01GM114237).

■ REFERENCES

(1) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696.

(2) Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T. Further along the Road Less Traveled: AMBER ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.* **2016**, *12* (8), 3926.

(3) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121* (16), 4023.

(4) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **2013**, *3* (2), 198.

(5) MacKerell, A. D., Jr; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586.

(6) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2009**, *31* (4), 671.

(7) Lopes, P. E. M.; Guvench, O.; MacKerell, A. D. Current status of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* **2015**, *1215*, 47.

(8) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40* (7), 843.

(9) Oostenbrink, C.; Soares, T. A.; Van der Vegt, N. F.; Van Gunsteren, W. F. Validation of the 53A6 GROMOS force field. *Eur. Biophys. J.* **2005**, *34* (4), 273.

(10) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657.

(11) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225.

(12) 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.

(13) Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *J. Phys.: Condens. Matter* **2009**, *21* (33), 333102.

(14) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114* (8), 2549.

(15) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable force fields for biomolecular simulations: Recent advances and applications. *Annu. Rev. Biophys.* **2019**, *48*, 371.

(16) Ren, P.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003**, *107* (24), 5933.

(17) Silvestrelli, P. L.; Parrinello, M. Water molecule dipole in the gas and in the liquid phase. *Phys. Rev. Lett.* **1999**, *82* (16), 3308.

(18) Gubskaya, A. V.; Kusalik, P. G. The total molecular dipole moment for liquid water. *J. Chem. Phys.* **2002**, *117* (11), 5290.

(19) Lamoureux, G.; MacKerell, A. D.; Roux, B. t. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003**, *119* (10), 5185.

(20) 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.

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

(21) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; Mackerell, A. D. Force Field for Peptides and Proteins based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013**, *9* (12), 5430.
(22) Patel, S.; Brooks, C. L. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004**, *25* (1), 1.
(23) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7* (10), 3143.
(24) Xie, W.; Gao, J. The Design of a Next Generation Force Field: The X-POL Potential. *J. Chem. Theory Comput.* **2007**, *3* (6), 1890.
(25) Wang, Z.-X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. Strike a balance: optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J. Comput. Chem.* **2006**, *27* (6), 781.
(26) Jorgensen, W. L.; Jensen, K. P.; Alexandrova, A. N. Polarization Effects for Hydrogen-Bonded Complexes of Substituted Phenols with Water and Chloride Ion. *J. Chem. Theory Comput.* **2007**, *3* (6), 1987.
(27) Geerke, D. P.; van Gunsteren, W. F. On the Calculation of Atomic Forces in Classical Simulation Using the Charge-on-Spring Method To Explicitly Treat Electronic Polarization. *J. Chem. Theory Comput.* **2007**, *3* (6), 2128.
(28) Lamoureux, G.; Roux, B. t. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **2003**, *119* (6), 3025.
(29) Holt, A.; Karlström, G. Improvement of the NEMO potential by inclusion of intramolecular polarization. *Int. J. Quantum Chem.* **2009**, *109* (6), 1255.
(30) Hermida-Ramón, J. M.; Brdarski, S.; Karlström, G.; Berg, U. Inter- and intramolecular potential for the N-formylglycinamide-water system. A comparison between theoretical modeling and empirical force fields. *J. Comput. Chem.* **2003**, *24* (2), 161.
(31) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. *J. Comput. Chem.* **2002**, *23* (16), 1515.
(32) Holt, A.; Karlström, G. Inclusion of the quadrupole moment when describing polarization. The effect of the dipole-quadrupole polarizability. *J. Comput. Chem.* **2008**, *29* (12), 2033.
(33) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry. *J. Phys. Chem. A* **2001**, *105* (2), 293.
(34) Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. V. Accurate first principles model potentials for intermolecular interactions. *Annu. Rev. Phys. Chem.* **2013**, *64*, 553.
(35) Piquemal, J.-P.; Chevreau, H.; Gresh, N. Toward a Separate Reproduction of the Contributions to the Hartree-Fock and DFT Intermolecular Interaction Energies by Polarizable Molecular Mechanics with the SIBFA Potential. *J. Chem. Theory Comput.* **2007**, *3* (3), 824.
(36) Gresh, N.; Claverie, P.; Pullman, A. Computations of intermolecular interactions: Expansion of a charge-transfer energy contribution in the framework of an additive procedure. Applications to hydrogen-bonded systems. *Int. J. Quantum Chem.* **1982**, *22* (1), 199.
(37) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3* (6), 1960.
(38) Piquemal, J.-P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. Towards a force field based on density fitting. *J. Chem. Phys.* **2006**, *124* (10), 104101.
(39) Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A. Generalization of the Gaussian electrostatic model: extension to arbitrary angular momentum, distributed multipoles, and speedup with reciprocal space methods. *J. Chem. Phys.* **2006**, *125* (18), 184101.
(40) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. X-Pol Potential: An Electronic Structure-Based Force Field for Molecular Dynamics Simulation of a Solvated Protein in Water. *J. Chem. Theory Comput.* **2009**, *5* (3), 459.
(41) 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.
(42) Ren, P.; Ponder, J. W. Temperature and pressure dependence of the AMOEBA water model. *J. Phys. Chem. B* **2004**, *108* (35), 13427.
(43) Grossfield, A.; Ren, P.; Ponder, J. W. Ion solvation thermodynamics from simulation with a polarizable force field. *J. Am. Chem. Soc.* **2003**, *125* (50), 15671.
(44) Jiao, D.; King, C.; Grossfield, A.; Darden, T. A.; Ren, P. Simulation of Ca2+ and Mg2+ solvation using polarizable atomic multipole potential. *J. Phys. Chem. B* **2006**, *110* (37), 18553.
(45) Wu, J. C.; Piquemal, J.-P.; Chaudret, R.; Reinhardt, P.; Ren, P. Polarizable molecular dynamics simulation of Zn(II) in water using the AMOEBA force field. *J. Chem. Theory Comput.* **2010**, *6* (7), 2059.
(46) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046.
(47) Jing, Z.; Qi, R.; Liu, C.; Ren, P. Study of interactions between metal ions and protein model compounds by energy decomposition analyses and the AMOEBA force field. *J. Chem. Phys.* **2017**, *147* (16), 161733.
(48) Jing, Z.; Liu, C.; Qi, R.; Ren, P. Many-body effect determines the selectivity for Ca2+ and Mg2+ in proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (32), No. E7495.
(49) Qi, R.; Jing, Z.; Liu, C.; Piquemal, J.-P.; Dalby, K. N.; Ren, P. Elucidating the Phosphate Binding Mode of PBP: The Critical Effect of Buffer Solution. *J. Phys. Chem. B* **2018**, *122* (24), 6371.
(50) Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (17), 6290.
(51) Zhang, J.; Yang, W.; Piquemal, J.-P.; Ren, P. Modeling Structural Coordination and Ligand Binding in Zinc Proteins with a Polarizable Potential. *J. Chem. Theory Comput.* **2012**, *8* (4), 1314.
(52) Bell, D. R.; Qi, R.; Jing, Z.; Xiang, J. Y.; Mejias, C.; Schnieders, M. J.; Ponder, J. W.; Ren, P. Calculating binding free energies of host-guest systems using the AMOEBA polarizable force field. *Phys. Chem. Chem. Phys.* **2016**, *18* (44), 30261.
(53) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes. *Chem. Rev.* **1994**, *94* (7), 1887.
(54) Stone, A. J. *The theory of intermolecular forces*. 2nd ed.; Oxford University Press: Oxford, 2013; p xi.
(55) Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J. Evaluation of charge penetration between distributed multipolar expansions. *J. Chem. Phys.* **2000**, *112* (17), 7300.
(56) Werneck, A. S.; Filho, T. M. R.; Dardenne, L. E. General methodology to optimize damping functions to account for charge penetration effects in electrostatic calculations using multicentered multipolar expansions. *J. Phys. Chem. A* **2008**, *112* (2), 268.
(57) Wang, B.; Truhlar, D. G. Including charge penetration effects in molecular modeling. *J. Chem. Theory Comput.* **2010**, *6* (11), 3330.
(58) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. General model for treating short-range electrostatic penetration in a molecular mechanics force field. *J. Chem. Theory Comput.* **2015**, *11* (6), 2609.
(59) Piquemal, J.-P.; Gresh, N.; Giessner-Prettre, C. Improved Formulas for the Calculation of the Electrostatic Contribution to the Intermolecular Interaction Energy from Multipolar Expansion of the Electronic Distribution. *J. Phys. Chem. A* **2003**, *107* (48), 10353.
(60) Rackers, J. A.; Wang, Q.; Liu, C.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. An optimized charge penetration model for use with the AMOEBA force field. *Phys. Chem. Chem. Phys.* **2017**, *19* (1), 276.
(61) Slipchenko, L. V.; Gordon, M. S. Electrostatic energy in the effective fragment potential method: theory and application to benzene dimer. *J. Comput. Chem.* **2007**, *28* (1), 276.

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

(62) Tafipolsky, M.; Engels, B. Accurate Intermolecular Potentials with Physically Grounded Electrostatics. *J. Chem. Theory Comput.* **2011**, *7* (6), 1791.
(63) Hohenstein, E. G.; Duan, J.; Sherrill, C. D. Origin of the surprising enhancement of electrostatic energies by electron-donating substituents in substituted sandwich benzene dimers. *J. Am. Chem. Soc.* **2011**, *133* (34), 13244.
(64) Mulliken, R. S. Molecular Compounds and their Spectra II. *J. Am. Chem. Soc.* **1952**, *74* (3), 811.
(65) Mao, Y.; Ge, Q.; Horn, P. R.; Head-Gordon, M. On the Computational Characterization of Charge-Transfer Effects in Noncovalently Bound Molecular Complexes. *J. Chem. Theory Comput.* **2018**, *14* (5), 2401.
(66) Stone, A. J. Natural bond orbitals and the nature of the hydrogen bond. *J. Phys. Chem. A* **2017**, *121* (7), 1531.
(67) Deng, S.; Wang, Q.; Ren, P. Estimating and modeling charge transfer from the SAPT induction energy. *J. Comput. Chem.* **2017**, *38* (26), 2222.
(68) Liu, C.; Qi, R.; Wang, Q.; Piquemal, J.-P.; Ren, P. Capturing Many-body Interactions with Classical Dipole Induction Models. *J. Chem. Theory Comput.* **2017**, *13* (6), 2751.
(69) Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. Revised parameters for the AMOEBA polarizable atomic multipole water model. *J. Phys. Chem. B* **2015**, *119* (29), 9423.
(70) Stone, A. J. Distributed multipole analysis: stability for large basis sets. *J. Chem. Theory Comput.* **2005**, *1* (6), 1128.
(71) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor Chim Acta* **1977**, *44* (2), 129.
(72) de la Lande, A.; Clavaguéra, C.; Köster, A. On the accuracy of population analyses based on fitted densities. *J. Mol. Model* **2017**, *23* (4), 99.
(73) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **2007**, *126* (14), 144111.
(74) Lillestolen, T. C.; Wheatley, R. J. Redefining the atom: atomic charge densities produced by an iterative stockholder approach. *Chem. Commun. (Cambridge, U. K.)* **2008**, *45*, 5909.
(75) Lillestolen, T. C.; Wheatley, R. J. Atomic charge densities generated using an iterative stockholder procedure. *J. Chem. Phys.* **2009**, *131* (14), 144101.
(76) Misquitta, A. J.; Stone, A. J.; Fazeli, F. Distributed Multipoles from a Robust Basis-Space Implementation of the Iterated Stockholder Atoms Procedure. *J. Chem. Theory Comput.* **2014**, *10* (12), 5405.
(77) Popelier, P. L. A. QCTFF: On the construction of a novel protein force field. *Int. J. Quantum Chem.* **2015**, *115* (16), 1005.
(78) Vandenbrande, S.; Waroquier, M.; Speybroeck, V. V.; Verstraelen, T. The monomer electron density force field (MEDFF): A physically inspired model for noncovalent interactions. *J. Chem. Theory Comput.* **2017**, *13* (1), 161.
(79) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59* (3), 341.
(80) van Duijnen, P. T.; Swart, M. Molecular and atomic polarizabilities: thole's model revisited. *J. Phys. Chem. A* **1998**, *102* (14), 2399.
(81) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J. Chem. Phys.* **2014**, *140* (9), No. 094106.
(82) Chaudret, R.; Gresh, N.; Parisel, O.; Piquemal, J.-P. Many-body exchange-repulsion in polarizable molecular mechanics. I. Orbital-based approximations and applications to hydrated metal cation complexes. *J. Comput. Chem.* **2011**, *32* (14), 2949.
(83) Lee, A. J.; Rick, S. W. The effects of charge transfer on the properties of liquid water. *J. Chem. Phys.* **2011**, *134* (18), 184507.
(84) Korchowiec, J.; Uchimaru, T. New energy partitioning scheme based on the self-consistent charge and configuration method for subsystems: Application to water dimer system. *J. Chem. Phys.* **2000**, *112* (4), 1623.
(85) Kumar, R.; Wang, F.-F.; Jenness, G. R.; Jordan, K. D. A second generation distributed point polarizable water model. *J. Chem. Phys.* **2010**, *132* (1), No. 014309.
(86) 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.
(87) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13* (7), No. e1005659.
(88) Harger, M.; Li, D.; Wang, Z.; Dalby, K.; Lagardère, L.; Piquemal, J.-P.; Ponder, J.; Ren, P. Tinker-OpenMM: Absolute and relative alchemical free energies using AMOEBA on GPUs. *J. Comput. Chem.* **2017**, *38* (23), 2047.
(89) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. 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.
(90) Rackers, J. A.; Liu, C.; Ren, P.; Ponder, J. W. A physically grounded damped dispersion model with particle mesh Ewald summation. *J. Chem. Phys.* **2018**, *149* (8), No. 084115.
(91) Leonard, A. N.; Simmonett, A. C.; Pickard, F. C.; Huang, J.; Venable, R. M.; Klauda, J. B.; Brooks, B. R.; Pastor, R. W. Comparison of Additive and Polarizable Models with Explicit Treatment of Long-Range Lennard-Jones Interactions Using Alkane Simulations. *J. Chem. Theory Comput.* **2018**, *14* (2), 948.
(92) Kolář, M.; Kubař, T.; Hobza, P. On the role of London dispersion forces in biomolecular structure determination. *J. Phys. Chem. B* **2011**, *115* (24), 8038.
(93) Demerdash, O.; Wang, L.-P.; Head-Gordon, T. Advanced models for water simulations. *WIREs Comput. Mol. Sci.* **2018**, *8* (1), No. e1355.
(94) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building force fields: an automatic, systematic, and reproducible approach. *J. Phys. Chem. Lett.* **2014**, *5* (11), 1885.
(95) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic improvement of a classical molecular model of water. *J. Phys. Chem. B* **2013**, *117* (34), 9956.
(96) Qi, R.; Wang, L.-P.; Wang, Q.; Pande, V. S.; Ren, P. United polarizable multipole water model for molecular mechanics simulation. *J. Chem. Phys.* **2015**, *143* (1), No. 014504.
(97) Wu, J. C.; Chattree, G.; Ren, P. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012**, *131* (3), 1138.
(98) Humphreys, D. D.; Friesner, R. A.; Berne, B. J. A Multiple-Time-Step Molecular Dynamics Algorithm for Macromolecules. *J. Phys. Chem.* **1994**, *98* (27), 6885.
(99) Qian, X.; Schlick, T. Efficient multiple-time-step integrators with distance-based force splitting for particle-mesh-Ewald molecular dynamics simulations. *J. Chem. Phys.* **2002**, *116* (14), 5971.
(100) Bussi, G.; Zykova-Timan, T.; Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* **2009**, *130* (7), No. 074101.
(101) Azar, R. J.; Head-Gordon, M. An energy decomposition analysis for intermolecular interactions from an absolutely localized molecular orbital reference at the coupled-cluster singles and doubles level. *J. Chem. Phys.* **2012**, *136* (2), No. 024103.
(102) Khaliullin, R. Z.; Head-Gordon, M.; Bell, A. T. An efficient self-consistent field method for large systems of weakly interacting components. *J. Chem. Phys.* **2006**, *124* (20), 204105.
(103) Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended S22 data set. *J. Chem. Theory Comput.* **2010**, *6* (8), 2365.

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139

(104) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark calculations of noncovalent interactions of halogenated molecules. *J. Chem. Theory Comput.* **2012**, *8* (11), 4285.
(105) Rezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7* (8), 2427.
(106) Řezáč, J.; Hobza, P. Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods. *J. Chem. Theory Comput.* **2012**, *8* (1), 141.
(107) Li, A.; Muddana, H. S.; Gilson, M. K. Quantum Mechanical Calculation of Noncovalent Interactions: A Large-Scale Evaluation of PMx, DFT, and SAPT Approaches. *J. Chem. Theory Comput.* **2014**, *10* (4), 1563.
(108) Vega, C.; Abascal, J. L. F. Simulating water with rigid non-polarizable models: a general perspective. *Phys. Chem. Chem. Phys.* **2011**, *13* (44), 19663.
(109) Benedict, W. S.; Gailar, N.; Plyler, E. K. Rotation-vibration spectra of deuterated water vapor. *J. Chem. Phys.* **1956**, *24* (6), 1139.
(110) Maroulis, G. Hyperpolarizability of H2O revisited: accurate estimate of the basis set limit and the size of electron correlation effects. *Chem. Phys. Lett.* **1998**, *289* (3−4), 403.
(111) Clough, S. A.; Beers, Y.; Klein, G. P.; Rothman, L. S. Dipole moment of water from Stark measurements of H2 O, HDO, and D2 O. *J. Chem. Phys.* **1973**, *59* (5), 2254.
(112) Verhoeven, J.; Dymanus, A. Magnetic properties and molecular quadrupole tensor of the water molecule by beam-maser zeeman spectroscopy. *J. Chem. Phys.* **1970**, *52* (6), 3222.
(113) Murphy, W. F. The Rayleigh depolarization ratio and rotational Raman spectrum of water vapor and the polarizability components for the water molecule. *J. Chem. Phys.* **1977**, *67* (12), 5877.
(114) Tschumper, G. S.; Leininger, M. L.; Hoffman, B. C.; Valeev, E. F.; Schaefer, H. F.; Quack, M. Anchoring the water dimer potential energy surface with explicitly correlated computations and focal point analyses. *J. Chem. Phys.* **2002**, *116* (2), 690.
(115) Lee, H. M.; Suh, S. B.; Lee, J. Y.; Tarakeshwar, P.; Kim, K. S. Structures, energies, vibrational spectra, and electronic properties of water monomer to decamer. *J. Chem. Phys.* **2000**, *112* (22), 9759.
(116) Curtiss, L. A.; Frurip, D. J.; Blander, M. Studies of molecular association in H2 O and D2 O vapors by measurement of thermal conductivity. *J. Chem. Phys.* **1979**, *71* (6), 2703.
(117) Odutola, J. A.; Dyke, T. R. Partially deuterated water dimers: Microwave spectra and structure. *J. Chem. Phys.* **1980**, *72* (9), 5062.
(118) Bates, D. M.; Tschumper, G. S. CCSD(T) complete basis set limit relative energies for low-lying water hexamer structures. *J. Phys. Chem. A* **2009**, *113* (15), 3555.
(119) Xantheas, S. S.; Aprà, E. The binding energies of the D2d and S4 water octamer isomers: high-level electronic structure and empirical potential results. *J. Chem. Phys.* **2004**, *120* (2), 823.
(120) Bulusu, S.; Yoo, S.; Aprà, E.; Xantheas, S.; Zeng, X. C. Lowest-energy structures of water clusters (H2O)11 and (H2O)13. *J. Phys. Chem. A* **2006**, *110* (42), 11781.
(121) Yoo, S.; Aprà, E.; Zeng, X. C.; Xantheas, S. S. High-Level Ab Initio Electronic Structure Calculations of Water Clusters (H2 O)16 and (H2 O)17: A New Global Minimum for (H2 O)16. *J. Phys. Chem. Lett.* **2010**, *1* (20), 3122.
(122) Fanourgakis, G. S.; Aprà, E.; Xantheas, S. S. High-level ab initio calculations for the four low-lying families of minima of (H2O)20. I. Estimates of MP2/CBS binding energies and comparison with empirical potentials. *J. Chem. Phys.* **2004**, *121* (6), 2655.
(123) Kell, G. S. Density, thermal expansivity, and compressibility of liquid water from 0.deg. to 150.deg.. Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. *J. Chem. Eng. Data* **1975**, *20* (1), 97.
(124) Wagner, W.; Pruß, A. The IAPWS formulation 1995 for the thermodynamic properties of ordinary water substance for general and scientific use. *J. Phys. Chem. Ref. Data* **2002**, *31* (2), 387.
(125) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120* (20), 9665.
(126) Jorgensen, W. L.; Jenson, C. Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *J. Comput. Chem.* **1998**, *19* (10), 1179.
(127) Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G. M.; Nilsson, A.; Benmore, C. J. Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range. *J. Chem. Phys.* **2013**, *138* (7), No. 074506.
(128) Soper, A. K. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chem. Phys.* **2000**, *258* (2−3), 121.
(129) Brini, E.; Fennell, C. J.; Fernandez-Serra, M.; Hribar-Lee, B.; Lukšič, M.; Dill, K. A. How water's properties are encoded in its molecular structure and energies. *Chem. Rev.* **2017**, *117* (19), 12385.
(130) Wade, A. D.; Wang, L.-P.; Huggins, D. J. Assimilating Radial Distribution Functions To Build Water Models with Improved Structural Properties. *J. Chem. Inf. Model.* **2018**, *58* (9), 1766.
(131) Qi, R.; Wang, Q.; Ren, P. General van der Waals potential for common organic molecules. *Bioorg. Med. Chem.* **2016**, *24* (20), 4911.
(132) Mahoney, M. W.; Jorgensen, W. L. Diffusion constant of the TIP5P model of liquid water. *J. Chem. Phys.* **2001**, *114* (1), 363.

DOI: 10.1021/acs.jctc.9b00261
J. Chem. Theory Comput. 2019, 15, 4122−4139