AMOEBA+ Classical Potential for Modeling Molecular Interactions

Chengwen Liu, Jean-Philip Piquemal, Pengyu Ren

Submitted date: 14/05/2019 • Posted date: 15/05/2019

Licence: CC BY-NC-ND 4.0

Citation information: Liu, Chengwen; Piquemal, Jean-Philip; Ren, Pengyu (2019): AMOEBA+ Classical Potential for Modeling Molecular Interactions. ChemRxiv. Preprint.

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 the Symmetry-Adapted Perturbation Theory (SAPT). Additionally, a new water model developed based on the AMOEBA+ framework captures various liquid phase properties in molecular dynamics simulations while remains 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.

File list (4)

amoebaplus_SI_1.pdf (1.76 MiB)	view on ChemRxiv • download file
manuscript.R1.pdf (2.09 MiB)	view on ChemRxiv • download file
amoebaplus_SI_2.xlsx (122.26 KiB)	view on ChemRxiv • download file
amoebaplus_SI_3.zip (662.95 KiB)	view on ChemRxiv • download file

Supporting Information for

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, TX, 78712, USA

[‡] Laboratoire de Chimie Théorique, Sorbonne Universités, UPMC, UMR7616 CNRS, Paris, France

Table of Contents

amoebaplus_SI_1.pdf (this file)	
- Computational details	Procedure to derive multipoles employing Iterative Stockholder Analysis
- Figure S1	Validation of CT model on organic dimers with different orientations
- Figure S2	Energy components from SAPT2+ and AMOEBA+ for other water dimers
- Figure S3-S6	Detailed extrapolation to obtain size-corrected self-diffusion constant
- Table S1-S2	The CP, CT, and vdW parameters of AMOEBA+ for organic molecules
- Table S3-S9	Liquid properties calculated with four AMOEBA+ water models
- Table S10-S16	Gas phase properties calculated with four AMOEBA+ water models
- Table S17-S18	Sensitivity of six liquid properties to 22 AMOEBA+ water parameters
amoebaplus_SI_2.xlsx (a separate file)	
- Training set S108	AMOEBA+ vs. SAPT2+ calculated interaction energy for the S108 set
- Testing sets	AMOEBA+ vs. SAPT2+(3) calculated interaction energy for the testing sets
amoebaplus_SI_3.tar (a separate file)	
- AMOEBAplus.prm	AMOEBA+ non-bonded parameters for organic molecules in S108 set
- waterDMA_WH.prm	AMOEBA+ water parameters with DMA multipoles and W-H mixing rule
- waterISA_WH.prm	AMOEBA+ water parameters with ISA multipoles and W-H mixing rule
- waterDMA_HHG.prm	AMOEBA+ water parameters with DMA multipoles and HHG mixing rule
- waterISA_HHG.prm	AMOEBA+ water parameters with ISA multipoles and HHG mixing rule
- S108Set_txyz	Cartesian coordinates of S108×7 dimers in tinker format
- TestingSet_txyz	Cartesian coordinates of testing set dimers in tinker format

* To whom correspondence should be addressed: pren@mail.utexas.edu

Computation details

Here we detail the procedure to derive Tinker-compatible multipole parameters employing the Iterative Stockholder Analysis (ISA) approach.¹⁻³ Example will be taken on water molecule for illustration purpose, but the whole procedure can also be applied to other small molecules as tested. Starting from the MP2-optimized geometry, we calculated the electron density of water monomer at MP2/aug-cc-pvtz level of theory. This was achieved in Gaussian software via the following keywords in route section of its input file

```
#p MP2/aug-cc-pvtz sp scf(fermi, tight) density(current)
```

Then the binary checkpoint file generated by Gaussian was converted to formatted checkpoint file using formchk program of Gaussian software

```
formchk Water-MP2-aTZ.chk Water-MP2-aTZ.fchk
```

The formatted checkpoint file was used by hi-multipoles.py script, which was obtained via <http://molmod.ugent.be/software>, to generate the ISA multipoles using the following command

```
python hi-multipoles.py Water-MP2-aTZ.fchk isa
```

The resulted multipoles (in global frame) were then converted to the GDMA output format via a python script (reformatISA.py), which can be obtained via <https://github.com/leucinw/ComputTools/>

```
python reformatISA.py isa_multipoles.txt
```

Then poledit.x program of Tinker was employed to generate the parameter file (Water-MP2.key), which includes the atom definition, multipoles in *local frame* and polarization definition.

```
poledit.x 1 Water-MP2.isa < Water-peditin.txt
```

Note here that the local frame was defined in the above Water-peditin.txt file. Finally, a perl script of Poltype (avgmpol.pl), which can be obtained via <https://github.com/pren/poltype>, was used to average the parameters with the same chemical environment (two hydrogen atoms here)

```
avgmpols.pl Water-MP2.key Water-MP2.xyz Water-groups.txt Water-MP2.key_2 1
```

HiPart scripts were developed by T. Verstraelen of Ghent University to perform population analysis based on Gaussian output thus serve as post-processing tools. These scripts are currently part of the Horton software (<http://theochem.github.com/horton>). We are grateful that the author of HiPart provided helpful online documentation (<http://molmod.github.io/hipart/index.html>).

Figure S1. Comparison of QM target and AMOEBA+ computed CT energy for organic dimers of different configurations and intermolecular distances. The target CT energy is defined as $E_{Ind}^{SAPT2+} - E_{Pol}^{AMOEBA+}$. Uracil-Uracil dimers at (a) BP and (b) π - π configurations; Benzene-Benzene dimers at (c) T-shaped and (d) π - π configurations; Pyridine-Pyridine dimers at (e) BP and (f) π - π configurations; Benzene-Pyridine dimers at (g) T-shaped and (h) π - π configurations.

Figure S2. Intermolecular interaction energy breakdown of (a) Smith05 dimer (structure shown right) at 7 intermolecular separations, and (b) hydrogen-bonding dimer with various typical flap angle θ . The definition of flap angle is the same as that in the main text **Figure 4** and also shown here on the right.

Figure S3. Detailed extrapolation to obtain the size-corrected self-diffusion constant (D_0) of water for AMOEBA+ model with DMA multipole and W-H combining rule. Red values are the final D_0 used to plot **Figure 12** of the main text.

Figure S4. Detailed extrapolation to obtain the size-corrected self-diffusion constant of water (D_0) for AMOEBA+ model with DMA multipole and HHG combining rule.

Figure S5. Detailed extrapolation to obtain the size-corrected self-diffusion constant of water (D_0) for AMOEBA+ model with ISA multipole and HHG combining rule.

Figure S6. Detailed extrapolation to obtain the size-corrected self-diffusion constant of water (D_0) for AMOEBA+ model with ISA multipole and W-H combining rule.

Table S1. Charge transfer and charge penetration parameters for the 38 molecules in S108 dimer set. The same atom class definition was used for both models. Complete parameters that can be used to calculate intermolecular interactions for these molecules are provided in a separate file (AMOEBAplus.prm).

Atom Class	a_{CT} (10^3 kcal·mol ⁻¹)	b_{CT} (Å ⁻¹)	α_{CP} (kcal·mol ⁻¹)	Description
1	3.4122	4.7712	3.2484	H (non-polar)
2	3.2983	4.6666	3.4437	H (aromatic)
3	3.7994	4.8850	3.2632	H (polar, water)
4	3.4217	4.5829	3.5898	C (sp3)
5	3.3271	3.0769	3.2057	C (sp2, aromatic)
6	3.1996	3.2889	3.1286	C (sp2, carbonyl etc.)
7	3.8170	3.2891	4.0135	N (sp3)
8	3.2119	4.3884	3.6358	N (sp2, aromatic)
9	3.9313	3.4544	3.7071	N (sp2, other)
10	3.4761	3.6034	4.1615	O (sp3, hydroxyl, water)
11	3.8387	3.3773	4.3778	O (sp2, aromatic)
12	3.7062	3.4186	3.7321	O (sp2, carbonyl)
13	3.7274	4.7750	2.7476	P (phosphate)
14	3.7201	2.1887	3.3112	S (sulfide)
15	3.5957	2.3621	2.6247	S (sulfur (IV), DMSO)
16	3.7345	4.0632	4.4675	F (organofluorine)
17	3.6666	2.4359	3.4749	Cl (organochlorine)
18	3.9554	2.2128	3.6696	Br (organobromine)

Table S2. van der Waals parameters for the 38 molecules in S108 dimer set. Atom class definition is the same as the vdw2016 model.⁴ The shape parameters, δ and γ in Buffered-14-7 potential are 0.07 and 0.12, respectively. A universal reduction factor of 0.98 was used for hydrogen atoms in both sets of parameters. Complete parameters that can be used to calculate intermolecular interactions for these molecules are provided in a separate file (AMOEBAplus.prm).

Atom Class	CUBIC-MEAN/W-H		CUBIC-MEAN/HHG		Description
	σ (Å)	ϵ (kcal·mol ⁻¹)	σ (Å)	ϵ (kcal·mol ⁻¹)	
1	3.23507	0.01165	3.31697	0.01284	H (nonpolar)
2	3.23456	0.01107	3.27027	0.01173	H (nonpolar, Alkane)
3	3.31389	0.00236	3.30318	0.00489	H (polar, NH/N aromatic)
4	3.33986	0.00245	3.37122	0.00509	H (polar, OH)
5	3.27806	0.00458	3.39984	0.00663	H (polar, SH)
6	3.33793	0.00790	3.26789	0.01303	H (aromatic, CH)
7	3.86757	0.05887	3.90314	0.03718	C (sp ³)
8	4.06174	0.01683	4.07743	0.02285	C (sp ³ , Alkane)
9	4.12470	0.04213	4.10225	0.05313	C (sp ² , Ethene)
10	4.06698	0.02547	3.99045	0.03178	C (sp ² , C=O)
11	4.04991	0.14012	4.07341	0.12700	C (sp)
12	4.23808	0.06176	4.02310	0.10100	C (aromatic, C-C)
13	4.01139	0.08393	3.99298	0.08121	C (aromatic, C-X)
14	3.88265	0.15802	3.94151	0.17840	N (sp ³)
15	3.91729	0.19853	4.00554	0.18787	N (sp ²)
16	3.76342	0.17121	3.93353	0.14630	N (aromatic)
17	3.81319	0.08478	3.81292	0.07840	O (sp ³ , hydroxyl, water)
18	4.08321	0.04778	3.77237	0.07120	O (sp ² , carbonyl)
19	3.84147	0.12949	3.83140	0.17024	O (O ⁻ in AcO ⁻)
20	3.85272	0.09799	3.81685	0.13469	O (O ⁻ in HPO ₄ ²⁻)
21	3.82250	0.09947	3.77323	0.14079	O (O ⁻ in H ₂ PO ₄ ⁻)
22	3.67652	0.04418	3.79092	0.02386	O (O in H ₃ PO ₄)
23	5.05360	0.06215	4.97998	0.02695	P (phosphate)
24	4.63755	0.20530	4.59947	0.16559	S (sulfide, R-SH)
25	4.62154	0.19518	4.61946	0.14138	S (sulfur IV, DMSO)
26	3.50163	0.09704	3.44498	0.12637	F (organofluorine)
27	4.42228	0.20606	4.42190	0.19471	Cl (organochlorine)
28	4.64273	0.28982	4.65922	0.28751	Br (organobromine)

Table S3. Liquid water density at the temperature ranging from 265 K to 369 K and under 1 atmosphere pressure calculated with four AMOEBA+ water models. DMA/W-H model is the one presented in the main text. Results from AMOEBA14 model are also provided for comparison.

T (K)	Expt.	W-H rule		HHG rule		AMOEBA14 ^a
		DMA	ISA	DMA	ISA	
265.15	998.6	996.5	997.7	996.9	997.7	999.9
273.15	999.8	999.3	1000.1	1000.0	1000.2	1001.3
277.15	1000.0	1000.0	1000.7	1000.6	1000.7	1001.5
281.15	999.8	1000.4	1000.7	1000.8	1000.8	1001.3
289.15	998.9	1000.3	1000.1	1000.1	1000.0	1001.1
298.15	997.0	998.3	997.9	998.2	998.3	997.9
309.15	993.7	994.7	993.7	994.8	994.8	994.7
321.15	988.9	989.1	988.3	989.2	989.0	989.9
333.15	983.2	982.4	981.3	982.1	982.5	983.4
345.15	976.6	974.2	973.0	973.8	974.6	975.9
353.15	971.8	968.4	967.1	967.7	968.6	970.3
369.15	961.2	955.4	953.6	954.3	955.3	958.0

^a These values are taken from the AMOEBA14 paper.⁵ A smaller water box was used in AMOEBA14 simulations. The simulation conditions are also different from those of the current work.

Table S4. Liquid enthalpy of vaporization at the temperature ranging from 265 K to 369 K and under 1 atmosphere calculated with four AMOEBA+ water models. DMA/W-H model is the one presented in the main text. Results from AMOEBA14 model are also provided for comparison.

T (K)	Expt.	W-H rule		HHG rule		AMOEBA14 ^a
		DMA	ISA	DMA	ISA	
265.15	45.4	46.3	45.9	46.7	46.6	46.2
273.15	45.1	45.9	45.5	46.3	46.1	45.8
277.15	44.9	45.6	45.3	46.0	45.9	45.6
281.15	44.7	45.4	45.1	45.8	45.7	45.4
289.15	44.4	45.0	44.6	45.3	45.2	45.0
298.15	44.0	44.5	44.2	44.8	44.8	44.5
309.15	43.5	43.9	43.6	44.2	44.2	43.9
321.15	43.0	43.3	43.0	43.6	43.5	43.3
333.15	42.5	42.7	42.4	42.9	42.9	42.7
345.15	41.9	42.0	41.8	42.3	42.3	42.1
353.15	41.6	41.6	41.4	41.9	41.8	41.6
369.15	40.8	40.8	40.5	41.0	41.0	40.8

^a These values are taken from the AMOEBA14 paper.⁵ A smaller water box was used in AMOEBA14 simulations. The simulation conditions are also different from those of the current work.

Table S5. Liquid thermal expansion coefficient at the temperature ranging from 265 K to 369 K and under 1 atmosphere calculated with four AMOEBA+ water models. DMA/W-H model is the one presented in the main text. Results from AMOEBA14 model are also provided for comparison.

T (K)	Expt.	W-H rule		HHG-rule		AMOEBA14 ^a
		DMA	ISA	DMA	ISA	
265.15	-2.4	-5.0	-3.5	-6.0	-3.8	-2.5
273.15	-0.7	-2.2	-2.1	-2.1	-2.0	-1.0
277.15	0.0	-1.5	-0.7	-1.0	-0.8	-0.1
281.15	0.6	-0.6	0.1	0.1	0.3	0.8
289.15	1.6	1.3	1.6	1.6	1.5	2.3
298.15	2.6	2.7	3.2	2.3	2.5	2.4
309.15	3.5	4.2	4.2	3.9	4.2	3.5
321.15	4.4	5.0	5.1	5.3	4.8	4.8
333.15	5.2	6.7	6.7	6.5	6.5	5.9
345.15	6.0	7.1	7.2	7.6	7.3	7.0
353.15	6.4	8.0	8.1	8.1	8.1	7.4
369.15	7.3	9.3	9.4	9.4	9.5	8.4

^a These values are taken from the AMOEBA14 paper.⁵ A smaller water box was used in AMOEBA14 simulations. The simulation conditions are also different from those of the current work.

Table S6. Liquid water isothermal compressibility at the temperature ranging from 265 K to 369 K and under 1 atmosphere calculated with four AMOEBA+ water models. DMA/W-H model is the one presented in the main text. Results from AMOEBA14 model are also provided for comparison.

T (K)	Expt.	W-H rule		HHG rule		AMOEBA14 ^a
		DMA	ISA	DMA	ISA	
265.15	55.9	60.1	62.3	63.6	68.3	55.5
273.15	52.2	55.9	58.6	59.9	64.2	52.1
277.15	50.7	53.9	56.7	57.3	61.4	50.5
281.15	49.3	52.3	55.3	56.2	59.2	49.2
289.15	46.9	51.6	53.7	53.0	56.6	47.4
298.15	45.2	49.0	51.4	51.6	53.0	46.6
309.15	44.4	49.5	51.2	50.7	53.2	45.3
321.15	44.2	49.2	51.2	50.7	53.1	45.0
333.15	44.5	49.6	51.6	53.0	55.0	45.7
345.15	45.3	51.3	53.2	53.8	55.9	47.0
353.15	46.1	54.3	54.1	56.3	57.4	48.4
369.15	48.3	58.1	58.7	60.2	60.9	51.5

^a These values are taken from the AMOEBA14 paper.⁵ A smaller water box was used in AMOEBA14 simulations. The simulation conditions are also different from those of the current work.

Table S7. Liquid isobaric heat capacity at the temperature ranging from 265 K to 369 K and under 1 atmosphere calculated with four AMOEBA+ water models employing fluctuation formula (**Eq. 22** in main text). DMA/W-H model is the one presented in the main text. Results from AMOEBA14 model are also provided for comparison.

T (K)	Expt.	W-H rule		HHG rule		AMOEBA14^a	Quantum correction
		DMA	ISA	DMA	ISA		
265.15	18.3	28.6	27.7	28.6	27.8	21.3	-2.5
273.15	18.2	28.0	26.7	28.1	27.3	20.7	-2.4
277.15	18.1	27.9	27.0	28.1	27.1	20.5	-2.4
281.15	18.1	27.2	26.4	27.4	26.9	20.5	-2.4
289.15	18.0	27.1	26.7	27.8	26.9	20.9	-2.3
298.15	18.0	26.7	26.4	26.9	26.3	20.5	-2.2
309.15	18.0	26.6	26.4	26.5	26.6	19.5	-2.1
321.15	18.0	25.8	25.9	26.8	26.7	20.0	-2.0
333.15	18.0	26.2	25.8	26.4	26.2	19.5	-1.9
345.15	18.0	25.6	25.6	25.7	26.0	20.2	-1.7
353.15	18.1	25.6	25.4	26.2	26.4	19.3	-1.7
369.15	18.1	25.8	25.5	26.1	26.3	19.5	-1.5

^a These values are taken from the AMOEBA14 paper.⁵ A smaller water box was used in AMOEBA14 simulations. The simulation conditions are also different from those of the current work.

Table S8. Liquid isobaric heat capacity at the temperature ranging from 265 K to 369 K and under 1 atmosphere calculated with four AMOEBA+ water models employing differentiation equation (**Eq. 23** in main text). DMA/W-H model is the one presented in the main text. Results from AMOEBA14 model are also provided for comparison.

T (K)	DMA/W-H	ISA/W-H	DMA/HHG	ISA/HHG	Quantum correction
269.15	20.6	19.8	20.8	20.2	-2.5
275.15	20.6	19.5	20.7	19.8	-2.4
279.15	20.1	19.3	20.5	19.6	-2.4
285.15	19.8	19.2	19.9	19.7	-2.3
293.65	19.6	19.2	20.4	19.7	-2.2
303.65	19.5	18.9	19.9	19.7	-2.1
315.15	19.2	19.0	19.8	19.7	-2.0
327.15	19.2	19.1	19.7	19.6	-1.9
339.15	19.6	19.4	20.1	19.9	-1.8
349.15	19.3	19.5	20.0	20.2	-1.7
361.15	20.0	19.9	20.3	20.2	-1.6

Table S9. Liquid static dielectric constant at the temperature ranging from 265 K to 369 K and under 1 atmosphere calculated with four AMOEBA+ water models. DMA/W-H model is the one presented in the main text. Results from AMOEBA14 model are also provided for comparison.

T (K)	Expt.	W-H rule		HHG rule		AMOEBA14^a
		DMA	ISA	DMA	ISA	
265.15	91.2	103.3	90.9	96.5	88.6	92.8
273.15	87.9	94.4	90.9	92.7	84.3	91.2
277.15	86.3	90.8	89.2	89.4	83.6	89.5
281.15	84.7	88.2	87.8	88.1	83.0	87.6
289.15	81.7	87.4	84.9	83.9	80.3	83.7
298.15	78.4	80.4	78.9	77.5	76.0	79.4
309.15	74.6	74.2	76.0	72.4	73.2	76.1
321.15	70.6	68.8	71.4	70.9	69.4	71.0
333.15	66.8	65.5	67.6	65.7	65.3	65.7
345.15	63.2	59.8	63.9	62.4	61.4	62.1
353.15	60.9	56.9	62.1	59.0	59.8	59.7
369.15	56.6	53.9	57.8	53.3	55.7	54.2

^a These values are taken from the AMOEBA14 paper.⁵ A smaller water box was used in AMOEBA14 simulations. The simulation conditions are also different from those of the current work.

Table S10. Monomer properties predicted by four AMOEBA+ models compared to AMOEBA14 and experimental values. DMA/W-H model is the one presented in the main text. Ideal monomer geometry (where O-H bond length is 0.9572 Å, and angle $\angle\text{HOH}$ is 104.52 degree) was used to calculate these properties.

Property		Expt.	W-H rule		HHG-rule	
			DMA	ISA	DMA	ISA
Dipole (Debye)	d_z	1.855	1.918	1.883	1.970	1.886
Quadrupole (Debye·Å)	Q_{xx}	2.630	3.193	2.941	2.998	2.927
	Q_{yy}	-2.500	-2.679	-2.648	-2.754	-2.751
	Q_{zz}	-0.130	-0.514	-0.294	-0.244	-0.176
Polarizability (Å ² ·s ⁴ ·kg ⁻¹)	α_{xx}	1.528	1.589	1.618	1.530	1.586
	α_{yy}	1.415	1.213	1.237	1.160	1.207
	α_{zz}	1.468	1.329	1.352	1.276	1.323

Table S11. Dimer properties at equilibrium geometry predicted by four AMOEBA+ water models. D_e (in kcal·mol⁻¹) is the dissociation energy. μ_{tot} (in Debye) is the total dipole moment. Other geometrical properties are illustrated in the main text in **Figure 4**. Optimized dimer geometry was used for AMOEBA+ unless explicitly noted.

Property	DMA/W-H	ISA/W-H	DMA/HHG	ISA/HHG
D_e	4.80	4.91	4.92	4.92
r_{O-O}	2.892	2.833	2.879	2.864
α	14.38	25.76	9.35	18.92
θ	79.7	88.8	77.9	84.3
μ_{tot}	1.79 (2.85 ^a)	0.947 (2.802 ^a)	2.062 (2.882 ^a)	1.414 (2.801 ^a)

^a Values calculated using AMOEBA+ model with MP2-optimized geometry

Table S12. Energy decomposition of SAPT2+ for 10 hydrogen-bonding water dimers. Total interaction energy from SAPT2+ and CCSD(T)/aug-cc-pv5z are provided. All energy values are in kcal·mol⁻¹.

$r_{O\cdots O}$ (Å)	E_{elst}	E_{ind}	$E_{exch-rep}$	E_{dtsr}	E_{int}^{SAPT2+}	$E_{int}^{CCSD(T)}$
2.37	-26.2	-14.5	51.6	-7.8	3.0	4.4
2.57	-16.5	-7.3	26.0	-5.1	-2.9	-2.5
2.77	-10.7	-3.8	12.9	-3.3	-4.8	-4.7
2.87	-8.8	-2.7	9.1	-2.7	-5.1	-5.0
2.97	-7.2	-2.0	6.4	-2.2	-5.0	-5.0
3.07	-6.1	-1.5	4.5	-1.8	-4.8	-4.8
3.19	-5.0	-1.0	2.9	-1.4	-4.5	-4.6
3.47	-3.3	-0.5	1.1	-0.8	-3.5	-3.5
3.78	-2.3	-0.2	0.4	-0.5	-2.6	-2.6
3.98	-1.9	-0.1	0.2	-0.3	-2.2	-2.2

Table S13. Interaction energy calculated with two AMOEBA+ models, DMA/W-H and ISA/W-H, for hydrogen-bonding dimer at 10 intermolecular distances. All energy values are in kcal·mol⁻¹.

$r_{O...O}$ (Å)	$E_{elst}^{DMA/W-H}$	$E_{pol}^{DMA/W-H}$	$E_{ct}^{DMA/W-H}$	$E_{ind}^{DMA/W-H}$	$E_{vdw}^{DMA/W-H}$	$E_{int}^{DMA/W-H}$
2.37	-25.8	-4.3	-9.1	-13.4	46.8	7.6
2.57	-16.7	-2.6	-3.9	-6.5	20.9	-2.4
2.77	-11.2	-1.6	-1.7	-3.3	9.6	-4.9
2.87	-9.3	-1.2	-1.1	-2.3	6.5	-5.1
2.97	-7.8	-1.0	-0.7	-1.7	4.4	-5.1
3.07	-6.6	-0.8	-0.5	-1.2	2.9	-4.9
3.19	-5.5	-0.6	-0.3	-0.9	1.8	-4.5
3.47	-3.7	-0.3	-0.1	-0.4	0.5	-3.6
3.78	-2.6	-0.2	0.0	-0.2	0.0	-2.8
3.98	-2.1	-0.1	0.0	-0.1	-0.1	-2.3
$r_{O...O}$ (Å)	$E_{elst}^{ISA/W-H}$	$E_{pol}^{ISA/W-H}$	$E_{ct}^{ISA/W-H}$	$E_{ind}^{ISA/W-H}$	$E_{vdw}^{ISA/W-H}$	$E_{int}^{ISA/W-H}$
2.37	-26.2	-4.0	-10.8	-14.9	50.3	9.3
2.57	-16.6	-2.5	-4.8	-7.2	22.2	-1.6
2.77	-10.9	-1.5	-2.1	-3.6	10.2	-4.3
2.87	-9.0	-1.2	-1.4	-2.6	6.9	-4.6
2.97	-7.5	-0.9	-0.9	-1.8	4.7	-4.6
3.07	-6.3	-0.7	-0.6	-1.3	3.2	-4.5
3.19	-5.2	-0.5	-0.4	-0.9	1.9	-4.2
3.47	-3.5	-0.3	-0.1	-0.4	0.5	-3.4
3.78	-2.4	-0.2	0.0	-0.2	0.1	-2.6
3.98	-2.0	-0.1	0.0	-0.1	0.0	-2.1

Table S14. Interaction energy calculated by two AMOEBA+ models, DMA/HHG and ISA/HHG, for hydrogen-bonding dimer at 10 intermolecular distances. All energy values are in kcal·mol⁻¹.

$r_{O...O}$ (Å)	$E_{elst}^{DMA/HHG}$	$E_{pol}^{DMA/HHG}$	$E_{ct}^{DMA/HHG}$	$E_{ind}^{DMA/HHG}$	$E_{vdw}^{DMA/HHG}$	$E_{int}^{DMA/HHG}$
2.37	-25.0	-4.2	-9.2	-13.4	45.9	7.4
2.57	-16.6	-2.6	-4.0	-6.5	20.5	-2.6
2.77	-11.2	-1.5	-1.7	-3.2	9.4	-5.0
2.87	-9.4	-1.2	-1.1	-2.3	6.4	-5.3
2.97	-7.9	-0.9	-0.7	-1.7	4.3	-5.2
3.07	-6.7	-0.7	-0.5	-1.2	2.9	-5.0
3.19	-5.6	-0.6	-0.3	-0.8	1.8	-4.6
3.47	-3.8	-0.3	-0.1	-0.4	0.5	-3.7
3.78	-2.7	-0.2	0.0	-0.2	0.0	-2.8
3.98	-2.2	-0.1	0.0	-0.1	-0.1	-2.3
$r_{O...O}$ (Å)	$E_{elst}^{ISA/HHG}$	$E_{pol}^{ISA/HHG}$	$E_{ct}^{ISA/HHG}$	$E_{ind}^{ISA/HHG}$	$E_{vdw}^{ISA/HHG}$	$E_{int}^{ISA/HHG}$
2.37	-24.2	-4.1	-10.7	-14.9	48.7	9.7
2.57	-16.0	-2.5	-4.7	-7.2	21.6	-1.6
2.77	-10.8	-1.5	-2.1	-3.6	9.9	-4.5
2.87	-9.0	-1.2	-1.4	-2.6	6.7	-4.8
2.97	-7.6	-0.9	-0.9	-1.8	4.6	-4.8
3.07	-6.4	-0.7	-0.6	-1.3	3.1	-4.7
3.19	-5.3	-0.5	-0.4	-0.9	1.9	-4.4
3.47	-3.6	-0.3	-0.1	-0.4	0.5	-3.5
3.78	-2.5	-0.2	0.0	-0.2	0.0	-2.7
3.98	-2.1	-0.1	0.0	-0.1	0.0	-2.2

Table S15. The binding energy of Smith dimers predicted by four AMOEBA+ models. Results from CCSD(T)/CBS and AMOEBA14 are provided for comparison. All energy values are in kcal·mol⁻¹.

(H ₂ O) ₂	CCSD(T)	W-H rule		HHG rule		AMOEBA14 ^a
		DMA	ISA	DMA	ISA	
Smith1	-4.97	-4.96	-4.41	-5.11	-4.64	-4.65
Smith2	-4.45	-4.11	-3.36	-4.20	-3.61	-4.22
Smith3	-4.42	-4.00	-3.18	-4.07	-3.42	-4.19
Smith4	-4.25	-4.75	-4.29	-4.47	-4.20	-3.54
Smith5	-4.00	-4.08	-3.55	-3.69	-3.48	-3.06
Smith6	-3.96	-3.90	-3.36	-3.48	-3.28	-2.92
Smith7	-3.26	-3.69	-3.72	-3.81	-3.77	-2.49
Smith8	-1.30	-1.39	-1.56	-1.58	-1.51	-1.02
Smith9	-3.05	-3.22	-3.46	-3.53	-3.53	-2.37
Smith10	-2.18	-2.34	-2.51	-2.47	-2.35	-1.96
RMSE		0.28	0.65	0.36	0.56	0.62

^a These values are taken from the AMOEBA14 paper.⁵

Table S16. The binding energy (BE) of trimer to 20-mer water clusters predicted by four AMOEBA+ models. Results from AMOEBA14 and high level QM are provided for comparison, where the BEs of small clusters ($n=3\sim6$) were calculated at CCSD(T)/CBS level of theory and the BEs of large clusters ($n>6$) were calculated at MP2/CBS level of theory. Deviations from QM data are reported for two water models.

(H ₂ O) _n ($n=3\sim20$)	MP2	W-H rule		HHG rule		AMOEBA14 ^a
		DMA	ISA	DMA	ISA	
Trimer	-15.74	-16.07	-15.97	-16.10	-16.08	-15.38
Tetramer	-27.40	-28.26	-28.54	-28.37	-28.38	-27.43
Pentamer	-35.93	-36.41	-36.01	-36.35	-35.72	-35.78
Prism	-45.92	-46.03	-47.12	-47.23	-48.24	-45.18
Cage	-45.67	-46.10	-46.51	-46.70	-47.18	-45.83
Bag	-44.30	-44.65	-44.88	-44.79	-44.99	-44.52
Cyclic Chair	-44.12	-44.88	-43.92	-44.66	-43.26	-43.53
Book1	-45.20	-45.87	-45.81	-46.08	-45.85	-45.08
Book2	-44.90	-45.25	-45.27	-45.54	-45.53	-45.06
CyclicBoat1	-43.13	-43.54	-42.15	-43.34	-41.82	-42.99
CyclicBoat2	-43.07	-43.51	-42.30	-43.32	-41.89	-43.07
Octamer_S4	-72.70	-73.56	-74.47	-74.26	-74.97	-72.22
Octamer_D2d	-72.70	-73.74	-74.69	-74.48	-75.20	-72.24
11-mer_434	-105.72	-101.65	-101.44	-102.00	-101.54	-101.11
11-mer_515	-105.18	-101.54	-101.92	-101.90	-102.04	-100.99
11-mer_551	-104.92	-101.23	-101.30	-101.63	-101.44	-100.58
11-mer_443	-104.76	-101.55	-101.84	-102.32	-102.57	-101.17
11-mer_4412	-103.97	-100.94	-100.83	-101.54	-101.41	-100.33
16-mer_boat-a	-170.80	-162.50	-163.16	-163.55	-163.88	-160.45
16-mer_boat-b	-170.63	-162.09	-162.66	-163.15	-163.48	-160.30
16-mer_antiboat	-170.54	-161.77	-162.13	-163.12	-163.63	-160.30
16-mer_abab	-171.05	-163.59	-164.54	-165.70	-166.93	-161.20
16-mer_aabb	-170.51	-163.19	-164.03	-165.00	-166.27	-160.89
17-mer_sphere	-182.54	-172.24	-171.90	-173.36	-172.96	-171.53
17-mer_5525	-181.83	-171.00	-171.25	-172.15	-172.51	-170.42
20-mer_Dodecahedron	-200.10	-193.58	-193.50	-193.61	-192.76	-193.81
20-mer_FusedCubes	-212.10	-208.65	-209.86	-211.52	-213.12	-205.77
20-mer_FaceSharingPrisms	-215.20	-205.31	-205.68	-207.28	-208.13	-204.41
20-mer_EdgeSharingPrisms	-218.10	-208.53	-208.03	-209.86	-209.32	-207.06
RMSE		5.46	5.31	4.67	4.62	6.33

^a These values are taken from the AMOEBA14 paper.⁵

Table S17. ForceBalance generated final gradient values of six properties w.r.t. tuning parameters.

	κ_T	ρ	ϵ_0	C_p	ΔH_{vap}	α_V
σ of O	-10.94	-138.5	-78.08	-16.67	-12.22	20.18
ϵ of O	-150.7	-727.7	-554.1	-170.3	-92.26	159.3
σ of H	-15.52	-87.82	-56.82	-15.36	-9.904	14.73
ϵ of H	-1893	-7333	-5570	-1779	-1003	1522
OH bond	61.8	-37.36	12.49	-47.61	14.21	-38.74
HOH angle	2.631	-2.666	-6.578	-0.6526	-0.4658	2.41
charge of O	-6.13	-211.1	-348.4	-104.3	-89.14	123.6
d_z of O	-9.231	47.14	90.33	24.04	16.21	-28.97
Q_{xx} of O	-16.16	12.75	0.7349	12.09	6.489	-12.21
Q_{yy} of O	-32.18	-25.4	-10	0.3564	-5.524	-8.936
d_z of H	-9.4	-46.93	-78.51	-24.82	-25.97	28.36
Q_{xx} of H	-8.066	56.82	29.46	6.548	3.744	-6.751
Q_{yy} of H	-0.2732	-10.85	13.59	8.626	-1.35	-7.184
Q_{zx} of H	-6.599	26.38	32.17	12.35	9.622	-12.21
α of O	-1.755	5.114	10.13	1.675	1.168	-2.816
α of H	6.377	9.53	13.59	4.383	2.718	-2
a_{CT} of O	0.9385	4.633	3.195	1.09	0.5862	-0.9451
b_{CT} of O	-4.296	-32.38	-20.17	-5.87	-3.607	5.761
a_{CT} of H	1.338	5.693	3.853	1.112	0.648	-0.9276
b_{CT} of H	-6.168	-33.6	-21.15	-5.467	-3.552	5.01
a_{CP} of O	-4.078	-75.29	-51.31	-7.382	-4.822	8.949
a_{CP} of H	-4.818	-64.17	-47.39	-8.521	-5.899	8.551

Table S18. Normalized fractions of each value in **Table S17** with the sum over each row values. The absolute values in **Table S17** smaller than 60 are thought to be small and the corresponding blocks in this table are filled with zeros for clarity consideration.

	κ_T	ρ	ε_0	C_p	ΔH_{vap}	α_V
σ of O	0.0	0.5	0.3	0.0	0.0	0.0
ε of O	0.1	0.4	0.3	0.1	0.0	0.1
σ of H	0.0	0.4	0.0	0.0	0.0	0.0
ε of H	0.1	0.4	0.3	0.1	0.1	0.1
OH bond	0.3	0.0	0.0	0.0	0.0	0.0
HOH angle	0.0	0.0	0.0	0.0	0.0	0.0
charge of O	0.0	0.2	0.4	0.1	0.1	0.1
d_z of O	0.0	0.0	0.4	0.0	0.0	0.0
Q_{xx} of O	0.0	0.0	0.0	0.0	0.0	0.0
Q_{yy} of O	0.0	0.0	0.0	0.0	0.0	0.0
d_z of H	0.0	0.0	0.4	0.0	0.0	0.0
Q_{xx} of H	0.0	0.0	0.0	0.0	0.0	0.0
Q_{yy} of H	0.0	0.0	0.0	0.0	0.0	0.0
Q_{zx} of H	0.0	0.0	0.0	0.0	0.0	0.0
α of O	0.0	0.0	0.0	0.0	0.0	0.0
α of H	0.0	0.0	0.0	0.0	0.0	0.0
a_{CT} of O	0.0	0.0	0.0	0.0	0.0	0.0
b_{CT} of O	0.0	0.0	0.0	0.0	0.0	0.0
a_{CT} of H	0.0	0.0	0.0	0.0	0.0	0.0
b_{CT} of H	0.0	0.0	0.0	0.0	0.0	0.0
a_{CP} of O	0.0	0.5	0.0	0.0	0.0	0.0
a_{CP} of H	0.0	0.5	0.0	0.0	0.0	0.0

Reference

1. Lillestolen, T. C.; Wheatley, R. J., Redefining the atom: atomic charge densities produced by an iterative stockholder approach. *Chem Commun (Camb)* **2008**, (45), 5909.
2. Lillestolen, T. C.; Wheatley, R. J., Atomic charge densities generated using an iterative stockholder procedure. *J. Chem. Phys.* **2009**, *131* (14), 144101.
3. 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.
4. Qi, R.; Wang, Q.; Ren, P., General van der Waals potential for common organic molecules. *Bioorg Med Chem* **2016**, *24* (20), 4911.
5. 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.

amoebaplus_SI_1.pdf (1.76 MiB)

[view on ChemRxiv](#) • [download file](#)

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, TX 78712, USA

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

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 the Symmetry-Adapted Perturbation Theory (SAPT). Additionally, a new water model developed based on the AMOEBA+ framework captures various liquid phase properties in molecular dynamics simulations while remains 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.

Keywords: Non-bonded Interactions, Charge Penetration, Charge Transfer, Polarizable Water Model, Symmetry-Adapted Perturbation Theory

* To whom correspondence should be addressed: pren@mail.utexas.edu

I. 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 Debye which increases by about 50% to 2.4 ~ 3.0 Debye 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 (Non-empirical 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 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 molecules²³ and complex proteins⁴⁶ 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 SAPT⁵³ provide detailed components of interaction energy including electrostatics, induction, exchange-repulsion, and dispersion. However, classical force fields 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 in with 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 electrostatic potentials (ESP) 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 formula for treating CP have been proposed by several research groups.⁵⁵⁻⁶⁰ We have demonstrated, 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 towards 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* energy decomposition analysis (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 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 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 charge transfer from SAPT induction energy.⁶⁷ To overcome the uncertain 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 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, 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.

II. Methodology

2.1 AMOEBA+ potential energy terms

The total potential energy of AMOEBA+ model can be expressed as the sum of the bonded and non-bonded energy terms

$$E_{total} = E_{bonded} + E_{non-bonded} \quad (1)$$

where E_{bonded} and $E_{non-bonded}$ can be expressed as

$$E_{bonded} = E_{bond} + E_{angle} + E_{b-a} + E_{oop} + E_{torsion} \quad (2)$$

$$E_{non-bonded} = E_{electrostatics}^{CP-corrected} + E_{polarization} + E_{charge-transfer} + E_{VDW} \quad (3)$$

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 to AMOEBA+ water model, the bonded terms including bond stretching, angle bending, and the Urey-Bradley, as with the current AMOEBA water models.^{16, 69} The main improvements of AMOEBA+ over AMOEBA FF on functional form lie in the non-bonded interactions, as will be described in detail next.

Electrostatics. The current AMOEBA model employs the point atomic multipoles truncated at quadrupoles to compute the electrostatic interactions. In AMOEBA the atomic multipole moments are usually derived with distributed multipole analysis (DMA) approach⁷⁰ and then optimized

against high-level *ab initio* electrostatic potential (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-I⁷²⁻⁷³ 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 charge penetration 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

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

where the Z_i and Z_j are nuclei charges and q_i and q_j are electron charges. Thus $Z_i + q_i = c_i^{DMA}$ 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 ($c_i^{DMA} c_j^{DMA} / r$) in common FFs. In AMOEBA+, damping is systematically applied to monopole, dipole and quadrupole. The electrostatic energy of AMOEBA+ is then the sum of 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

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

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 are added. The effect of CP damping is short-ranged and dies off at typical real-space Ewald cutoffs. For example, the magnitude of CP 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.

Polarization. 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:

$$\mu_i^{ind} = \alpha_i (\sum_j T_{ij}^{damped} M_j + \sum_j T_{ij}^{damped} \mu_j^{ind}) \quad (6)$$

where T_{ij} is the multipole-multipole interaction matrix for *direct* induction and T_{ij}^{11} only includes the dipole field related terms for *mutual* induction. Thole damping scheme⁷⁹ 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

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

where $u(r) = r_{ij}/(\alpha_i\alpha_j)^{\frac{1}{6}}$ is the polarizability-normalized distance between sites i and j ; and r_{ij} is the actual distance; 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 3-body energy (E_{3B}) can be achieved by modifying the functional form for damping the *permanent* field to

$$f_{MB}^{direct}(r) = 1 - e^{-au^{\frac{3}{2}}(r)} \quad (8)$$

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 AMOEBA. More details of the modified polarization model can be found in a previous publication.⁶⁸

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 SAPT2+⁸¹ induction energy and the polarization energy as described above: $E_{Induction}^{SAPT2+} - E_{Polarization}^{AMOEBA+}$. SIBFA has an elaborated many-body charger transfer term using 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 discrete charge transfer approach⁸³⁻⁸⁵ and electron density overlap approximation method³⁸ where the monomer density is approximated by exponential functions. We noted the CT

with exponential summation only between hydrogen and oxygen atoms in water⁸⁵ 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:

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

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 hetero-atoms as

$$a_{ij} = \sqrt{a_i a_j} \quad (10)$$

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

In practice, CT interaction is very short-ranged and a 6 Å cutoff is used for 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.

Van der Waals. AMOEBA+ model retains the same vdW functional form used in the current AMOEBA, the Halgren's buffered-14-7 potential ⁹²

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

where ε_{ij} is the potential well depth, $\sigma_{ij} = r_{ij}/r_{ij}^0$ 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.

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

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

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

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

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 code⁸⁶ and CUDA platform supported in 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 CT energy and forces. CP was incorporated as pairwise corrections to the multipole-multipole interactions in the real space of 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 AMOEBA+ energy and gradient evaluation, comparing to AMOEBA, is increased slightly by ~ 6%, which includes the cost of added CP and CT, as well as modified (*direct*) polarization energy and gradient. The detailed computational cost of both 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 AMOEBA-based MD engine⁸⁸ 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-HP⁸⁹ 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 CPU-GPU hybrid platform, is in progress. The Tinker-CPU code is available at TinkerTools GitHub site as AMOEBA+ branch.

Table 1. Evaluation of computational cost of AMOEBA+ model.^{a,b} Total wall time (in second) of 500 energy and gradient calculations for each potential energy component was evaluated by TIMER program in Tinker.

	AMOEBA	AMOEBA+	Difference	Difference %	Change in AMOEBA+
E_{elst}	9.7	11.2	+1.5	+15.2%	Added charge penetration correction
E_{pol}	40.9	41.9	+1.1	+2.6%	Split the <i>direct</i> and <i>mutual</i> induction
E_{ct}	/	0.9	+0.9	/	Added new charge transfer 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 @ 2.70GHz.

^b Important settings: The $36 \text{ \AA}^3$ simulation box contains 1600 water molecules; cutoffs for vdW and Ewald real space and charge transfer were 12, 7 and $6 \text{ \AA}$, respectively; neighbor-list method was used; induced dipoles were converged to $10\text{E-}4$. 500 repetitions were run to evaluate the energy and force.

2.3 Parameterization of organic molecules based on SAPT

All SAPT2+ data, Cartesian coordinate structure and parameter files are also included on the SI. As described in our previous work, the S108×7 dimer set was systematically constructed and the interaction energy was decomposed using 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 of the equilibrium distances. All the 38 molecules in S108 set are listed in

Figure 1.

Figure 1. Molecular structures of the 38 molecules in S108 dataset. These molecules cover nine chemical elements: hydrogen (white), oxygen (red), carbon (gray), nitrogen (blue), sulfur (yellow), phosphorus (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 analogs or fragments of nucleic acids.

The interaction energy components of SAPT2+ can be expressed as

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

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

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

The atomic multipoles were derived for the molecules in S108 dataset 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 have been derived for S101 dataset⁴² and here we directly expanded to S108 set without further optimization.

We modified the *direct* damping factor from the original 0.75 in our previous publication⁶⁸ to 0.70 in this work. It was found that this change leads to better agreement in many-body energy of water clusters with MP2 values and only has a subtle influence on organic compounds. After subtracting polarization energy from total induction energy of SAPT2+, we parametrized CT model with the remaining energy as the target. In the parametrization of the vdW model, we experimented 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 deviation (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 a 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 AMOEBA+ water model based on *ab initio* and experimental data

As an application of AMOEBA+ potential, the AMOEBA+ water model has been parameterized 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 iAMOEBA⁹⁵ 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 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 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 tool⁹⁷ was used to derive the initial DMA multipoles. We detail the computational procedure for deriving ISA multipoles in SI. In FB optimization, ANALYZE program of Tinker was used to compute the energy of water clusters and liquid boxes from MD simulations. DYNAMIC_OMM program was used to perform liquid phase NPT simulations. The box dimension was $\sim 26 \times 26 \times 26 \text{ \AA}^3$ containing 590 water molecules. The vdW cutoff was set to 10 $\text{\AA}$ in FB run with a long-range correction. For simulations to correct the size effect on self-diffusion constant, 12 $\text{\AA}$ was used in the larger water boxes. Long-range electrostatic interaction was treated using particle mesh Ewald (PME) with 7 $\text{\AA}$ real space cutoff, which was also used as the cutoff for calculating CP correction. 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 $1.0 \text{ kg}\cdot\text{m}^{-3}$ 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 w.r.t. 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). All of these data can be found in the “studies/015_amoeba_tinker” directory in ForceBalance program, except for the “Hydrogen-bonding dimer” set, of which the interaction energy was decomposed using SAPT2+ for dimers in 10 intermolecular separations.

System and property		Data point	Data type	Weight ^a	
Clusters	Hydrogen-bonding dimer	10	SAPT2+ interaction energy	1.0	
	Smith dimer	10	CCSD(T) binding energy	1.0	
	Trimer to hexamer	11	CCSD(T) binding energy	1.0	
	Octamer to 20-mer	18	MP2 binding energy	0.5	
Liquid	Density	12	Experiment	1.0	3.0 ^b
	Enthalpy of vaporization	12	Experiment		2.0 ^b
	Thermal expansion coefficient	12	Experiment		1.0 ^b
	Static dielectric constant	12	Experiment		1.0 ^b
	Isobaric heat capacity	12	Experiment		0.1 ^b
	Isothermal compressibility	12	Experiment		1.0 ^b

^a Weights applied in ForceBalance optimization among different targets and properties within the target. Here each cluster system is a target, and liquid is another target. ^b Weights applied to the six liquid properties. Weights were also applied to properties at different temperatures (not shown).

III. Results and discussion

We previously obtained SAPT2+ intermolecular energy components for the S108×7 dataset.⁵⁸ The parameters of CP, CT, and vdW obtained in this study are summarized in **Table S1-S2**. In total AMOEBA+ has 18 atom classes for CP and CT, 28 atom classes for vdW, and nine 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, AMOEBA+ model can accurately capture three energy components. We then developed an AMOEBA+ model for 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 non-bonded interactions of AMOEBA+ (electrostatics, CT and vdW) were parameterized 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 SAPT induction energy. The performance of AMOEBA+ for energy components are 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)

Table 3. Statistics evaluation of the AMOEBA+ model on each energy component and total interactions for S108×7 dimer set compared to SAPT2+ data. All energy values are in kcal·mol⁻¹.

Data points	Statistics ^a	E_{elst}	E_{ind}	E_{vdW} ^b		E_{inter} ^c
<i>short distances</i> (0.70~0.80× 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
<i>medium distances</i> (0.90~1.10× 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
<i>all distances</i>	MUE	0.73	0.51	1.43	(1.62)	1.39
	MSE	0.12	0.15	-0.16	(-0.43)	0.11

(0.70~1.10× r_{eq})	RMSE	1.34	0.92	2.67	(2.97)	2.78
------------------------	------	------	------	------	--------	------

^a MUE: mean unsigned error; MSE: mean signed error; RMSE: root mean square error.

^b Values from W-H and HHG combining rule (in parentheses) for ϵ_{ij} are provided.

^c Sum of three components where the E_{vdW} was calculated with W-H combining rule for ϵ_{ij} .

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 S101 dataset.⁶⁰ In this work, we directly applied CP on the S108 dataset without re-optimizing the parameters. Using transferable, expanded element-based parameters, the CP-corrected point multipoles are able to reproduce 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 panel in **Table 3** and blue dots in **Figure 2**), with an RMSE of 0.61 kcal·mol⁻¹. It is worth to mention that for several organic molecules 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.

Figure 2. Correlation plot of the intermolecular energy and three decomposed components for 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 (d) Van der Waals energy, where the AMOEBA+ values are calculated from parameters with 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.

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 S108 data set. **Figure 3** shows water dimers of very different configurations at various intermolecular separations. It is encouraging that the absolute errors w.r.t. the QM target energy, $E_{Ind.}^{SAPT2+} - E_{Pol.}^{AMOEB+}$, 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. 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 5th and 6th 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 and d**), the deviations are *at the most* 0.5 kcal·mol⁻¹ in 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 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 well 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.

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 rule 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.

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

3.1.2 Validation of AMOEBA+ on additional data set

We have 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 our training set but only the structures with intermolecular separations not present in S108 are included. 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 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.

Table 4. Validation of the accuracy and transferability of AMOEBA+ on additional SAPT data. RMSEs (in kcal·mol⁻¹) of AMOEBA+ against SAPT2+(3) are reported for individual energy components.

Database	Relative intermolecular distances	Data points	RMSE ^e			
			<i>E_{ele}</i>	<i>E_{ind}</i>	<i>E_{vdW}</i>	<i>E_{inter}</i>
S14 ^a	1.2, 1.5, 2.0	42	0.65	0.18	0.21	0.82
S59 ^b	1.25, 1.5, 2.0	177	0.24	0.14	0.21	0.44
Ionic ^c	1.0, 1.05, 1.1, 1.25, 1.5, 2.0	54	1.21	1.63	1.69	2.00
X6 ^d	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 Subset of S22×5;¹⁰³ ^b Subset of S66×8 with 1.2 and above distances;¹⁰⁵ ^c Subset of “Ionic” with 1.0 and above distances;¹⁰⁶ ^d Subset of X40 with 1.0 and above distances;¹⁰⁴ ^e SAPT2+(3) energy decomposition data are taken from reference 107.

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 resulted 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 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-phase

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 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 the detailed results in SI.

Parameters of AMOEBA+ water model are shown in **Table 5**. The initial non-bonded parameters were derived within the S108 dataset, 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 MP2/6-311G** level of theory. The dipole and quadrupole parameters were further optimized by fitting to the ESP at 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 AMOEBA+ model, after FB optimization, converges to that of the AMOEBA03. Please note that the atomic multipoles cannot be uniquely determined. DMA and ISA methods could lead to dramatically different atomic multipole values and yet both produce accurate electrostatic potential 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 give electrostatic interaction energy consist with SAPT decomposition. In addition, the vdW parameters changed slightly from those initial values derived by fitting to SAPT repulsion-dispersion energy, so did the valence parameters. This again shows that, to achieve the desired chemical accuracy in 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 w.r.t. the each parameter (**Table S17 and S18**) and found that 1) almost all of the properties are sensitive to the vdW and multipole parameters; 2) 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) charge penetration parameters have a strong effect on liquid density.

Table 5. Initial and optimized parameters of 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
	<i>direct</i> damping factor	none	0.70	0.70
	<i>mutual</i> 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

3.2.1 Gas phase cluster properties

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 vapor¹⁰⁹ 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 Debye, lower than the *ab initio* and experimental values, mostly due to the larger HOH angle. The molecular dipole moment increases to 1.918 Debye when the “ideal” geometry is used. AMOEBA+ has larger quadrupole moments but smaller polarizability comparing to the AMOEBA14. In **Table 7**, the dimer properties at equilibrium geometry are given. 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 (**Table S10 and S11**).

Table 6. Monomer properties of AMOEBA+ compared to AMOEBA14 and experimental values. AMOEBA+ values were given for both model-optimized and “ideal” geometry, with the latter given in parenthesis.

Property		AMOEBA14 ^a	AMOEBA+ ^b	<i>ab initio</i> ^c	Expt. ^d
Dipole (Debye)	d_z	1.808	1.778 (1.918)	1.840	1.855 ^d
Quadrupole (Debye·Å)	Q_{xx}	2.626	3.314 (3.193)	2.57	2.63 ^e
	Q_{yy}	-2.178	-2.628 (-2.679)	-2.42	-2.50 ^e
	Q_{zz}	-0.045	-0.684 (-0.514)	-0.14	-0.13 ^e
Polarizability (Å ² ·s ⁴ ·kg ⁻¹)	α_{xx}	1.767	1.583 (1.589)	1.47	1.53 ^f
	α_{yy}	1.308	1.209 (1.213)	1.38	1.42 ^f
	α_{zz}	1.420	1.308 (1.329)	1.42	1.47 ^f

^a Reference 69; ^b In parentheses are the values calculated with the “ideal” geometry, where O-H bond length is 0.9572 Å, and angle $\angle$ HOH is 104.52 degree; ^c Reference 110; ^d Reference 111; ^e Reference 112; ^f Reference 113.

Table 7. Dimer properties at equilibrium geometry predicted by AMOEBA+ compared to AMOEBA14 and experimental values. D_e (in kcal·mol⁻¹) is the dissociation energy with monomer optimized. μ_{tot} (in Debye) is the total dipole moment. Other geometrical properties are illustrated in **Figure 4**. Optimized dimer geometry was used for AMOEBA+ unless explicitly noted.

Property	AMOEBA14 ^a	AMOEBA+	<i>ab initio</i>	Expt.
D_e	4.64	4.80	4.98 ^b	5.44±0.7 ^d
r_{O-O}	2.908	2.892	2.907 ^b	2.976 ^e
α	4.61	14.38	5.5 ^b	-1±10 ^e
θ	64.9	79.7	56.9 ^b	57±10 ^e
μ_{tot}	2.20	1.79 (2.85 ^f)	2.76 ^c	2.643 ^e

^a Reference 69; ^b Reference 114; ^c Reference 115; ^d Reference 116; ^e Reference 117; ^f Value calculated using AMOEBA+ model with MP2-optimized geometry.

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

Energy Decomposition Analysis. The primary goal of AMOEBA+ potential over the current AMOEBA is to improve the physical representation of individual energy components, reduce error cancellation, 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. Interaction energy at CCSD(T)/aug-cc-pv5z level of theory was also calculated and provided here. Even with liquid-optimization, the electrostatics, induction and vdW energies are still well reproduced in all distances w.r.t the SAPT2+ EDA. The errors are noticeable mostly at very close contact. At 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**. 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 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 a cyclic geometry with non-typical O···H interactions, and the hydrogen-bonding dimer with various flap angles, shows good transferability of AMOEBA+ water model in predicting interaction energy components (**Figure S2**).

Figure 5. Intermolecular interaction energy of hydrogen-bonding dimer at 10 separation distances. (a) Energy components from SAPT2+ and AMOEBA+ model. (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.

Binding Energy of Water Clusters. 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 dimer to 20-mers are compared with high-level QM results as well as AMOEBA14 in **Table 8** and **Table 9**. Compared to AMOEBA14 model, AMOEBA+ offers better BEs for *eight out of ten* Smith dimers. RMSE of AMOEBA+ water dimer BE with respect to CCSD(T) is reduced from 0.62 kcal·mol⁻¹ of AMOEBA14 to 0.28. The Smith dimer set is usually thought 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 trimer to octamers. However, AMOEBA+ does noticeably better than AMOEBA14 for almost all larger clusters from 11-mer to 20-mers. Overall, an improvement of ~0.6 kcal·mol⁻¹ in terms of RMSE is achieved by AMOEBA+ over AMOEBA14 model (**Table 9**). It is interesting that both DMA-WH and DMA-HHG models predict the BEs of Smith dimers better than ISA-WH and ISA-HHG models (**Table S14**). Considering the greater monopole and smaller quadrupole in the ISA-based multipoles, this could suggest 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**).

Table 8. The binding energy of Smith dimers predicted by AMOEBA+ and AMOEBA14 models compared to QM data at CCSD(T)/CBS. 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.

(H ₂ O) ₂	CCSD(T) ^a	AMOEBA+	AMOEBA14 ^b
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 Reference 126; ^b Reference 69.

Table 9. The binding energy of trimer to 20-mer water clusters predicted by AMOEBA+ and AMOEBA14 models compared to QM data, where the BEs of small clusters ($n=3\sim6$) were calculated at CCSD(T)/CBS and the BEs of large clusters ($n>6$) were calculated at MP2/CBS. 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.

(H ₂ O) _n	Geometry	QM	AMOEBA+	AMOEBA14 ^g
$n=3$ ^a	Cyclic	-15.74	-16.07	-15.38
$n=4$ ^a	Cyclic	-27.40	-28.26	-27.43
$n=5$ ^a	Cyclic	-35.93	-36.41	-35.78
$n=6$ ^b	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
$n=8$ ^c	S4	-72.70	-73.56	-72.22
	D2d	-72.70	-73.74	-72.24
$n=11$ ^d	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
$n=16$ ^e	Boat-a	-170.80	-162.50	-160.45
	Boat-b	-170.63	-162.09	-160.30
	Anti-boat	-170.54	-161.77	-160.30
	ABAB	-171.05	-163.59	-161.20
	AABB	-170.51	-163.19	-160.89
$n=17$ ^e	Sphere	-182.54	-172.24	-171.53
	5525	-181.83	-171.00	-170.42
$n=20$ ^f	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 Reference 16; ^b Reference 118. ^c Reference 119; ^d Reference 120; ^e Reference 121; ^f Reference 122; ^g Reference 69.

3.2.2 Liquid properties

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

Density. One of the anomalous properties of water is its maximum density at 4 °C. The density-temperature profile simulated by 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 SI). For example, the density predicted by AMOEBA+ at 277 K is $1000.0 \text{ kg}\cdot\text{m}^{-3}$, in excellent agreement with the experimental value ($1000.0 \text{ kg}\cdot\text{m}^{-3}$), and AMOEBA14 gives $1001.5 \text{ kg}\cdot\text{m}^{-3}$. At ambient temperature (298 K), the average density by AMOEBA+ is $998.3 \text{ kg}\cdot\text{m}^{-3}$ and AMOEBA14 gives $997.9 \text{ kg}\cdot\text{m}^{-3}$. We note that a larger time step (2 fs) results in an average density of $997.7 \text{ kg}\cdot\text{m}^{-3}$ by AMOEBA+, compared to the experimental density of $997.0 \text{ kg}\cdot\text{m}^{-3}$. Other three models parametrized in this work (DMA/HHG, ISA/W-H, ISA/HHG) show similar performance (**Table S3**).

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

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 as obtained from the stochastic dynamics simulation of a monomer at specific temperatures using a time step of 0.1 fs. The AMOEBA+ ΔH_{vap} at 298 K is 44.5 kJ·mol⁻¹, in agreement with the 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 interaction energy of water molecules and have a strong influence on the self-

diffusion constant of water (see later discussion), the slope of the ΔH_{vap} - T curve is also important as it defines the heat capacity.

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

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

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

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.

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

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

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

To compare with experimental κ_T at selected temperatures, we ran the MD simulations using the AMOEBA+ water final parameters. Here we used on a small water box (216 molecules) and the Nose-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.

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

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

$$\epsilon_0 = 1 + \frac{4\pi}{3k_B T \langle V \rangle} (\langle M^2 \rangle - \langle M \rangle \cdot \langle M \rangle) \quad (21)$$

where $\langle M \rangle$ is the ensemble is 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 atmosphere, whereas the experimental measurement is 78.4.

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

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

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

In the above equation, N is the number of molecules in the simulation box, k_B is the Boltzmann constant and T is 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 intra-molecular corrections are usually necessary when compare simulations with experiments.¹²⁵ Another approach for calculating C_p independent of fluctuation is to use the differentiation equation¹²⁶

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

Where the first term is the differentiation of liquid potential energy w.r.t. 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 ΔH_{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).

Figure 11. Liquid water heat capacity at temperatures ranging from 265 to 369 K and under atmospheric pressure (1 atm) calculated with 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 fluctuation formula are in 0.6~1.4 cal·mol⁻¹·K⁻¹ and not shown in the plot. Experimental data are taken from reference 124.

Radial Distribution Function. The experimental radial distribution functions (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 scattering¹²⁷ or neutron diffraction¹²⁸ 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 since many of water properties are encoded in its structure.¹²⁹ **Figure 12** provides RDFs of O···O, O···H and H···H pairs from AMOEBA+ simulation at 298 K and under 1 atmosphere. 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 on describing RDFs of water by explicit fitting using ForceBalance.¹³⁰ Nonetheless, our previous work also showed that the Buckingham potential has difficulty to capture SAPT2+ exchange-repulsion energy.¹³¹ Features beyond the first peaks are also consistent with the neutron diffraction curves, which is essential since the tetrahedral structure of water is characterized by the second and third peaks.¹²⁸ The 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.

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

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

$$D_0 = \lim_{t \rightarrow \infty} \frac{d}{dt} \langle |r(t) - r(t_0)|^2 \rangle \quad (24)$$

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 the cubic boxes of 18 Å, 40 Å, 60 Å and 90 Å, where N_{water} is 216, 2210, 7500 and 25150, respectively. The time length of NPT production MD simulations are 6 ns, 5 ns, 4 ns and 2 ns, respectively. The size-independent D_0 was obtained by a linear fit to $D \sim \frac{1}{L}$ and extrapolating to $\frac{1}{L} = 0$. As seen from **Figure 13**, AMOEBA+ water D_0 matches very well with the experimental data. AMOEBA+ D_0 at room temperature and under 1 atmosphere is $2.23 \times 10^5 \text{ cm}^2 \cdot \text{s}^{-1}$ while the corresponding experimental value is $2.30 \times 10^5 \text{ cm}^2 \cdot \text{s}^{-1}$. This agreement is better than widely used non-polarizable water models, such as SPC¹⁴¹ and TIPnP ($n=3,4$ and 5)¹³² as well as polarizable models, such as AMOEBA models^{16, 69} and AMOEBA variants.⁹⁵⁻⁹⁶ As discussed above, W-H vdW combining rule performs slightly better than HHG on capturing the vdW component of 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 HHG rule provides, HHG-based water models underestimate D_0 of water at a series of temperatures. For example, the D_0 with the HHG combining rule, with either DMA or ISA based multipoles, is systematically slower than experimental value (298 K) by 8.0~12.6% (**Figure S4 and S5**). In contrast, models employing W-H rule are able to predict appropriate D_0 for various temperatures.

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

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 AMOEBA03 water model even beyond the ambient conditions.⁴² 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 SAPT method. By extending this approach from water to other chemical species, we will be able to arrive at a general and transferable force field for molecular simulations.

IV. Conclusion

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 over-parameterization 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, short-range charge penetration and charge transfer, by extending the polarizable atomic multipole-based AMOEBA model. The key improvements are the inclusion of short-range charge penetration and charge transfer effects and modification of direct polarization damping and vdW combining rules. We adopted a unique approach to separate the CT energy out of 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 AMOEBA+ potential were further validated on the testing database generated by other researchers. Direct application *ab initio* potential to condensed-phase is limited by the accuracy of quantum mechanical methods, as well

as approximations and simplifications made in classical potentials. Thus the most effective approach to extend *ab initio* potentials to condensed-phase is to directly incorporate accurate liquid thermodynamic properties that are widely available. Following this strategy, we have 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 lead 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-HP⁸⁹ 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 force fields.

Supporting Information

Materials not presented in the main text, including the atomic multipoles, atomic polarizabilities, CP, CT, and vdW parameters for organic molecules (AMOEBAplus.prm), validation of CT model on organic dimers with different orientations, EDA results and Cartesian coordinates of organic S108 and testing set dimers, EDA testing of AMOEBA+ on additional water dimers, detailed extrapolation plot to obtain the size-corrected self-diffusion constants for four water models, the numerical results of gas and liquid phase properties obtained with four water models, the analysis of parameters sensitivity to six liquid properties, and complete parameters of four AMOEBA+ water models. This information is available free of charge via the DOI of this article at: <https://pubs.acs.org/doi/abs/10.1021/acs.jctc.xxxx>.

Acknowledgment

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 Jr, A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry 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* **2010**, *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. *The Journal of Physical Chemistry 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. *Annual Review of biophysics* **2019**, *48*.

16. Ren, P.; Ponder, J. W., Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry 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.
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. *The Journal of Physical Chemistry 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. *Proceedings of the National Academy of Sciences* **2018**, *115* (32), 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. *The Journal of Physical Chemistry 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 USA* **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. *Physical Chemistry Chemical Physics* **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*. Second edition. ed.; Oxford University Press: Oxford, 2013; p xi.
55. Freitas, 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. *Physical Chemistry Chemical Physics* **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.
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 (Camb)* **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), 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), 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), 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), 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), 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), 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), 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), 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.
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. Rezáč, 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. *Physical Chemistry Chemical Physics* **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 H₂O, HDO, and D₂O. *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.; Tarakeswar, 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 H₂O and D₂O 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 (H₂O)₁₆ and (H₂O)₁₇ : A New Global Minimum for (H₂O)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.; Pitara, 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), 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. *Bioorganic & Medicinal Chemistry* **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.

Table of Contents graphics

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, TX 78712, USA

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

manuscript.R1.pdf (2.09 MiB)

[view on ChemRxiv](#) • [download file](#)

Other files

amoebaplus_SI_2.xlsx (122.26 KiB)	view on ChemRxiv • download file
amoebaplus_SI_3.zip (662.95 KiB)	view on ChemRxiv • download file