Atomic Polarizabilities for Interactive Dipole Induction Models

Jacob M. Litman, [†] *Chengwen Liu,* [†] *and Pengyu Ren**

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

[†] J.M.L. and C.L. contributed equally to this work.

[*] Corresponding author. Email: pren@mail.utexas.edu

Abstract

Thole-style mutual induction models for molecular polarization have been adopted by several popular polarizable force fields (FFs) for their simplicity and transferability. The atomic polarizability parameters of these models are typically derived by fitting to *ab initio* or/and experimental molecular polarizabilities. In this work, we improve upon Thole polarizability parameters by employing both high-level quantum mechanics molecular polarizabilities and electrostatic potential (ESP) responses on 3-dimensional grids. Our results indicate that the two approaches to derive atomic polarizability parameters are both effective, while the ESP approaches can also capture the polarization for the atoms with lone pair electrons. The resulted polarizability parameters have been validated on a set of over 7200 molecules covering the most common elements found in organic molecules (C, H, O, N, P, S, F, Cl, Br, and I). These parameters have also been tested on the experimentally measured molecular polarizabilities of 422 molecules. The final set of parameters derived in this work show notable improvement over the current AMOEBA set. The result is a highly transferable, expanded set of atomic polarizabilities defined by the local chemical environment in the form of SMARTS patterns. These parameters can be used directly in molecular mechanics polarizable potential energy functions such as AMOEBA, AMOEBA+ and other Thole-style models.

1. INTRODUCTION

Polarizable force fields (FFs) play important roles in studying the structure, dynamics, and thermodynamics of many biomolecular systems due to their explicit treatment of electronic polarization effect.[1-4] They are also routinely applied in predicting the host-ligand binding affinities that are essential to the drug discovery process.[5-7] Interactive point dipole (IPD) models such as AMOEBA,[8] AMBER,[9] and SIBFA[10] give each atom a point induced dipole, proportional to local electrical field strength and an atomic polarizability parameter α. Other means of estimating polarizability exist, e.g., fluctuating charges[11] and Drude oscillator models,[4, 12] but those use their own parameters (albeit Drude models have been proven equivalent to IPD models after careful conversion[12]). The most widely used parameters are those of Thole[13] and Swart,[14] both of whom fit to modest sets of molecular polarizabilities (both experimental and computational) to obtain physically reasonable atomic polarizability values for the most common elements in organic chemistry.

While effective, these polarizability parameters leave room for improvement. Firstly, per-element atomic polarizabilities ignore the chemical environment that atoms are exposed to. For example, the current AMOEBA model uses element-based polarizability parameters supplemented with aromatic carbon and hydrogen types.[15] Upon examining many organic molecular clusters, significant many-body polarization errors in this model were eliminated by the introduction of carbonyl-specific carbon and oxygen types.[16] Secondly, polarizability parameters were often derived based on a small number of relatively small molecules. For example, in the original paper of Thole,[13] about 20 molecules were employed. The polarizability parameters in the polarizable AMBER FF were derived initially based on 41 small molecules. In their later development, Wang and co-workers increased the number of molecules in training set to several hundreds.[17-18] Thirdly, the polarizability parameters have been mostly fitted to experimental or *ab initio* molecular polarizabilities. Reliance on a single source of data may cause over/underfitting of the model, especially when the data set is small.

Electrostatic potential (ESP) has been widely used by the FF developer community to obtain permanent charges[19] and higher-order moments (such as dipole and quadrupole).[8] Utilizing electrostatic potential responses to derive atomic polarizability parameters has also been demonstrated but without widespread adoption. For example, Kaminski *et al*. applied a series of electrostatic perturbations to the target molecules with dipoles (represented by two charges) at various locations around the molecule.[20-21] The atomic polarizability was then fitted to reproduce the ESP response upon perturbation. Elking *et al*. used a similar approach with a probe charge to derive atomic polarizabilities for their Gaussian induced dipole polarization model.[22] Their results show that the ESP probing procedure outperforms the generic atom type-based molecular polarizability fitting. Although different in representation of polarizability effect, the polarizable

Drude model by Mackerell and co-workers [23] also uses a series of perturbed ESP maps to determine the polarizability parameters.

Here in this work, we combine both approaches, deriving Thole-style polarizability parameters by fitting against both molecular polarizabilities and ESP responses. For six elements (C, H, O, N, S, Cl) we are able to take advantage of 7211 molecular polarizabilities previously calculated at the LR-CCSD/d-aug-cc-pVDZ level.[24] This is supplemented with an additional set of over 300 molecules, including charged molecules and other heavy elements (F, Cl, Br, I, P), with molecular polarizabilities and ESP responses calculated at the MP2/aug-cc-pVTZ level.

By fitting to both sets of data (molecular polarizabilities and ESP responses), we: 1) avoid possible over/underfitting by solely targeting either set of data, and 2) examine the robustness of the model and transferability of the derived parameters by cross-validating on two sets of data. The current element-based atom types have been split to better reflect any given atom's chemical environment, with the new types defined by SMARTS strings. Our results indicate that both individual sets of atomic polarizability parameters, as well as a hybrid model based on both data sets, can be used interchangeably to describe the molecular polarizability and ESP responses. The *final* set of parameters (based on the hybrid model) shows notable improvement over the current AMOEBA model, with the mean percentage molecular polarizability error being reduced from 9.3% to 2.5% and the RMS error from 1.24 $Å^3$ to 0.40 $Å^3$ for over 7200 molecules. Meanwhile, the RMS error of ESP responses has been reduced by 27%.

2. METHOD

2.1 Dipole interaction model

In the AMOEBA and AMOEBA+ models, an induced point dipole is introduced on the polarizable atom according to the electric field felt by that atom. Atomic polarization is obtained via an interactive induction model based on Thole's damped interaction scheme.[13]

$$\mu_i = \alpha_i \left(E_i - \sum_{i \neq j} T_{ij} \mu_j \right) \qquad \textbf{Eq. 1}$$

where α_i is the atomic polarizability, E_i is the field felt by site i due to the multipole moments (charge, dipole and quadrupole) on other sites, and T_{ij} is the dipole field tensor (see e.g. Ref. [8]). This interactive induction scheme requires that an induced dipole produced at any atom i will further induce other polarizable sites. The induced dipole on each polarizable atom is solved iteratively to obtain the converged induced dipole. It is worth mentioning that the dipole interaction matrix is damped to avoid the so-called

polarization catastrophe[25] at very short interatomic distances. With a universal damping factor of 0.39 (with the exception of some metal cations[26]), together with element-based isotropic atomic polarizability parameters, this model remains simple and shows reasonable transferability to describe the molecular polarizability of small organic molecules.[15]

2.2 Fitting targets: molecular polarizability and ESP responses

Molecular polarizability is a by-molecule description of how electron distribution changes in response to an external electric field. Experimental measurements exist for many molecules, as well as computational methods to estimate polarizability from quantum mechanical (QM) calculations.[27] The primary shortcoming is that there is no unique mapping of atomic polarizabilities to molecular polarizability, a problem aggravated when examining large molecules with many atoms contributing to overall polarizability.

Molecular electrostatic potential (ESP) has widely been used to optimize charges[28-29] and multipoles.[30-33] This is the energy a probe charge would feel at a given point in a fixed electrical field, such as that created by a molecule's charge distribution. As such, it is often used as a proxy for electrostatic intermolecular interactions, with charges and multipoles fit to reproduce ESP on a grid surrounding the molecule.

Polarization can be thought of as the response of a molecule's electron distribution (and thus its ESP) in response to an external electric field, such as that of another molecule. We add a small probe charge (+0.125 *e*) located at 4.3 Å away from the atom of interest, and permit the molecule's electron distribution to relax to the presence of this probe charge, and re-calculate ESP. We note here that the combination of +0.125 e charge and 4.3 Å distance can provide large enough electronic response for later numerical fitting, and small enough to still result in linear polarization. The change in ESP from polarization is then the non-additivity in ESP between the probe-molecule system and its components (probe and unpolarized molecule) (**Eq. 2**). We then fit both to these $\Delta\varphi_{pol}$ values (ESP responses) and molecular polarizabilities up to quadrupole. An illustration of this approach is provided in **Figure 1** using molecular surface of a benzene molecule colored by ESP values.

$$\Delta\varphi_{pol} = \varphi_{Mol+pro} - \varphi_{Mol} - \varphi_{Pro} \qquad \textbf{Eq. 2}$$

This approach has two advantages. First, the probe charge is a highly localized electrical disturbance: a dipole or other multipole would be even more strongly localized, but this is not explored in this work. Second, multiple probe locations can be used in independent calculations to probe near multiple locations of interest, e.g. near every non-redundant atom in the system.

2.3 Molecular databases

2.3.1 Molecular polarizability database

The molecular polarizability database includes a *training* set for parametrization and a *validation* set for validating the derived parameters. The *training* set contains 773 molecules in total, of which 700 molecules were chosen from the QM7b dataset according to chemical diversity, and the additional 73 molecules were constructed to expand the element coverage to F, Cl, Br, and I. The molecules in the QM7b set contain up to 7 heavy atoms (i.e., C, N, O, S, and Cl) with various levels of hybridization. The molecular polarizability for the 73 molecules was calculated at the MP2/aug-cc-pvtz level of theory. For the 700 molecules selected from QM7b set, the molecular polarizability data calculated at LR-CCSD/d-aug-cc-pVDZ level of theory was taken from Wilkins *et al*.[24] We show in the **Results** section that the differences between LR-CCSD/d-aug-cc-pVDZ and MP2/aug-cc-pvtz polarizabilities are within the uncertainty of our polarization model. The remaining 6511 molecules in the QM7b dataset were used as the validation set. Additionally, 422 molecules whose experimental molecular polarizabilities are available were added, of which less than 10% overlap with the QM7b set. Structures of these molecules were retrieved from PubChem API by molecular names provided in the publication of Bosque and Sales.[34] Structure optimization was performed using the Gaussian 09 or 16 software package[35] at MP2/6-31G* [36] level of theory; MP2/6-311G* was used for iodine-containing molecules.

2.3.2 ESP probing database

We constructed a set of molecules for ESP probing. Due to the computational cost, a minimal set of molecules that cover all the atom types of the molecular polarizability set was selected. In total, 316 probing structures were generated for 37 small molecules, with each molecule being probed at several different locations outside of the van der Waals sphere by an external charge (+0.125 *e*). The ESP of these probe structures was evaluated at MP2/aug-cc-pvtz[37] level; The def2-tzvppd [38-39] basis set was used for iodine-containing molecules. This basis set was obtained from the basis set exchange library.[40]

2.3.3 Charged molecule database

Like the neutral molecules described above, we also constructed molecular polarizability and ESP probing datasets for charged molecules commonly encountered in (bio)chemistry. In total, we performed QM polarizability calculations for 70 molecules with either positive or negative charges, of which 27 molecules were used to generate the ESP probing data. Polarizability and ESP were calculated at the same level as the neutral molecules.

3. RESULTS AND DISCUSSION

3.1 Atom types for atomic polarizability

String-based specifications are widely used in cheminformatics and computational chemistry as molecular descriptors. Here we adopt SMARTS patterns[41] to define atom types for the atoms in neutral molecules. Every atom matching a given atom type shares one atomic polarizability parameter. In total, we found that 34 atom types can cover the chemistry of over 7600 molecules (both QM and experimental datasets) used in this work. Whereas current AMOEBA polarizability types are generally limited to one aliphatic and one aromatic type (see Supporting Information, SI, **Table S1**), the current set includes multiple aliphatic types for many elements. The exact SMARTS patterns are listed in **Table S2**. For charged molecules, the atom types were defined based on the connectivity of the atoms. A detailed illustration of the atom types for the charged molecules is shown in **Figure S1**.

3.2 Comparison of CCSD and MP2 polarizability

To check whether the literature LR-CCSD/d-aug-cc-pVDZ level[24] was consistent with our own MP2/aug-cc-pVTZ level, we randomly selected 52 molecules from the QM7b database for comparison. Molecular polarizabilities were re-calculated at MP2/aug-cc-pVTZ level, using the QM7b geometry. The result shows that difference between these two levels of QM theory is only 0.05 $Å^3$ on average for the isotropic molecular polarizability (**Table S3**), which is within the uncertainty of a classical polarizability model. This suggests that we can safely use two levels of QM molecular polarizabilities as the fitting target.

3.3 Atomic polarizability derived from fitting molecular polarizability

As described above, the atomic polarizability was first derived by fitting the QM molecular polarizability on the training set data. The resulted atomic polarizability parameters (denoted as *fit-polar* set) are listed in **Table 1**. This *fit-polar* set represents several changes compared to the current parameters of the AMOEBA model (**Table S1**). For example, the aromatic carbon is increased from 1.75 to 2.11 $Å^3$; the *sp* hybridized carbon is now 2.28 $Å^3$; similarly, nitrogen on aromatic systems also increases from 1.07 to 1.65 $Å^3$). However, many other polarizability parameters show only small changes versus the current AMOEBA model.

The performance of this set of parameters (*fit-polar*) is summarized in **Table 2**. The statistics are calculated on the three polarizability eigenvalues (denoted as α_{ii} in **Table 2**) and the isotropic molecular polarizabilities (denoted as α_{iso} in **Table 2**). For the training set, the Root Mean Square Error (RMSE) and Unsigned Mean Percentage Error (UMPE, defined in **Eq. 3**) are 0.85 $Å^3$ and 4.7%, respectively, on the three eigenvalues α_{ii}, and 0.41 $Å^3$ and 2.8%, respectively, on the isotropic polarizability α_{iso}. This performance quality is well maintained (**Table 2**) on the much larger validation set, which demonstrates a robust, transferable model unlikely to be suffering from overfitting problems.

$$\mathrm{UMPE}(\alpha) = \frac{1}{N}\sum_{N}\frac{|\alpha - \alpha_{ref}|}{\alpha_{ref}} * 100\% \qquad \textbf{Eq. 3}$$

where N loops over the number of molecules and α is the molecular polarizability (either α_{ii} or α_{iso}).

3.4 Atomic polarizability derived from fitting electrostatic potential

Atomic polarizability parameters were also derived from fitting the ESP responses as described in the **Methods** section. As seen in **Table 1** (*fit-esp* set), generally, these parameters are slightly smaller than their *fit-polar* counterparts. Exceptionally, for several nitrogen and oxygen types, the *fit-esp* set of atomic polarizability parameters become bigger than the *fit-polar* set. This can easily be attributed to the lone pair of nitrogen and oxygen atoms in these molecules. Fitting to electrostatic potential tends to increase the polarizability to capture the polarization due to external field along the direction of the lone pair.

The performance of this set of parameters has been summarized in **Table 3**. The RMS error of ESP responses is 0.0326 kcal/mol/e. We directly apply this set of parameters to the polarizability dataset and find the resulting parameters are approximately as effective as the *fit-polar* set (**Table 3**).

3.5 Finalization of atomic polarizability

As described above, two sets of atomic polarizability parameters have been derived by fitting to either molecular polarizability or the change of ESP due to an external field. Results show that the two sets of parameters have very similar quality metrics (**Table 3**), which is partly due to very similar parameters for most polarizability types. For the atoms with lone pair electrons such as oxygen and nitrogen, the atomic polarizability must be bigger to capture the polarization of ESP in the direction of lone pairs. These results suggest that the ESP information should not be excluded in the derivation of atomic polarizability, as ESP captures these outliers where molecular polarizability does not. Our final polarizability parameters are a simple average of the two sets of parameters (denorted as *final* set in **Table 1**).

The testing results on the whole molecular polarizability and ESP databases are listed in **Table 3**. As expected, the averaged parameters maintain performance very similar to the component *fit-polar* and *fit-esp* sets. In addition, all sets of parameters outperform the current AMOEBA model. The current AMOEBA gives an RMS error of 1.24 $Å^3$ for the mean polarizability of 7284 molecules in the polarizability set, which is reduced to 0.40 $Å^3$ in this study with updated atomic polarizabilities. Nearly 4-fold improvement can be seen from the unsigned mean percentage error (9.3% vs. 2.5%, **Table 3**, **Figures 2** and **3**). The correlation coefficient (R^2) for the mean molecular polarizability by this work improves from 0.74 to 0.92. As seen from **Table 3** and **Figure 3**, a higher RMS error is obtained when checking the three eigenvalues of the molecular polarizability. Specifically, large errors present for large molecules such as those with highly

conjugated rings and long chains. For these systems, neither the current nor the updated set of parameters result as good performance as for the small molecules. This is well expected and acceptable given that AMOEBA adopts the simple *isotropic* atomic polarizability model[13] and a limited number of atom types are employed in this work. Further improvements may require the use of *anisotropic* atomic polarizabilities[42] and finer atom type definitions.

In addition, the *final* set of parameters have been tested on several gas molecules as special cases (**Table S4**) and 422 organic molecules with experimental molecular polarizability data (**Table S5**). The result shows that one can directly apply these parameters without defining new atom types for the gas molecules. For the organic molecules, the RMSE and UMPE are 0.49 $Å^3$ and 2.9%, respectively. The level of accuracy is similar to the QM set with over 7200 molecules, indicating the transferability of the derived atomic polarizability parameters.

It is believed that the atomic polarizability reflects the spheric "*volume*" of an atom, not unsurprising given that polarizability parameters are defined in units of volume.[43] Correlating the polarizability parameters with atomic and covalent radii suggests polarizability is better correlated with atomic radii (**Figure 4**). Particularly, polarizability for fluorine is 0.33 $Å^3$ on average, smaller than that of hydrogen (0.45 $Å^3$). This trend is in line with the atomic radii of two atoms (0.42 Å for fluorine vs. 0.53 Å for hydrogen). Polarizability for the heavy aromatic atom types is usually higher than non-aromatic ones (e.g., C, N, Cl, Br, and I), which is in line with our chemical intuition that the electrons are more delocalized in the aromatic than non-aromatic systems. Atoms with type "Xpolar" have higher polarizability than those with "Xnonpol" (where X = C, N and O), which is consistent with chemistry that one atom connecting to electronegative atoms (N and O) is more polarizable than the one that connects to a carbon. These results suggest that the polarizability parameters we derived here are physically meaningful, as higher polarizability correlates with larger, more delocalized electron densities.

3.6 Effect of polarizability on the many-body energy

Non-additive many-body interactions are responsible for various condensed phase properties. It has been of interest to examine the agreement of many-body interaction energy by polarizable models and high-level quantum chemical calculations.[16, 44-45] Here we calculate the 3-body energy by AMOEBA model with the current and the *final* set of polarizability parameters and compare them with the high-quality benchmark data obtained at CCSD(T)/CBS level of theory for the 3B-69 dataset.[46] Our results show that the AMOEBA model with the current element-based polarizability parameters captures the 3-body energy quite well with an RMS error of 0.30 kcal/mol (**Table S6**). The agreement can be slightly improved with an RMS error of 0.26 kcal/mol when the *final* set of polarizability parameters are used. These parameters will be further examined in MD simulations of condensed phase in the future.

3.7 Atomic polarizability for charged molecules

Following the same procedure of deriving parameters for neutral molecules, atomic polarizability parameters for common charged molecules have been generated. Due to the formal charge, it is sometimes difficult to use SMARTS strings to define the atom types of these molecules. In this work, we developed a program to enumerate the possible charged fragments to assign the atom types. The parameters and performance for charged molecules are provided in **Table S7** and **S8**, respectively. For the charged molecules, an RMSE of 0.52 $Å^3$ and UMPE of 3.1% for the *isotropic* polarizability are obtained with the *final* set of parameters. The slight reduction in fit quality (compared with the neutral set) may partially be explained by the difficulty in fitting the significantly larger absolute polarizability values in charged groups, particularly negatively charged groups with diffuse electron clouds.

4. CONCLUSION

We have derived three sets of atomic polarizability parameters for use with interactive point dipole models. One set is from direct fitting to the molecular polarizability calculated from high-level quantum chemistry calculations. The second set is from fitting the electrostatic potential change due to an external electric field introduced by a dummy partial charge. The final set is obtained by averaging these values, which fortunately are in close correspondence. Our results clearly show that two sets of approaches to deriving atomic polarizability parameters are closely matched, though the ESP approach is more effective in capturing polarization in the direction of the lone pairs of atoms such as nitrogen and oxygen atoms. With the newly derived atomic polarizability parameters, we show that for the QM set with over 7200 molecules, the RMS isotropic molecular polarizability error is reduced from 1.24 $Å^3$ (current AMOEBA model) to 0.40 $Å^3$. This level of accuracy is achieved through introducing additional polarizability types to Thole's interactive induction scheme to consider both element and local electronic environment of the atoms. It is worth mentioning that the universal damping factor (0.39) determined by AMOEBA model [8] remains the same in this work. Testing of the final parameters on the available experimental polarizability set (422 molecules) results in an RMS error of 0.49 $Å^3$, indicating the good transferability of the derived set of atomic polarizability parameters. Using the same procedure, we have also derived the atomic polarizability parameters for 70 molecular ions, which expands the applicability of the polarizability model.

A Python program was developed to assign polarizability parameters from a SMARTS string, available on Github.[47] The final set of atomic polarizability parameters has also been incorporated into the PolType2 automation tool [48] for AMOEBA-family FF development. The parameters derived in this work will be used

in future AMOEBA[8] and AMOEBA+ [33] force fields, and can easily be used in other Thole-type polarization models, and with slightly more effort, in Drude oscillator models.[12]

5. ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (R01GM106137 and R01GM114237) and National Science Foundation (CHE-1856173).

6. DATA AND SOFTWARE AVAILABILITY

The Tinker software package used in this work can be obtained from TinkerTools Github site (https://github.com/TinkerTools/tinker). The Python program to assign atomic polarizability parameters is accessible to the public via https://github.com/prenlab/amoebaplus_parameter. The QM optimized structures of all the molecules used in this work and the molecular polarizability data can be downloaded from https://github.com/prenlab/amoebaplus_data.

7. SUPPORTING INFORMATION

This material is available free of charge via the Internet at http://pubs.acs.org.

The atom types for the current AMOEBA model and this work (**Figure S1**, **Table S1** and **S2**); comparison of molecular polarizabilities calculated from two levels of QM methods (**Table S3**); performance test on a set of gas molecules (**Table S4**) and experimental data (**Table S5**); the 3-body energy of 3B-69 dataset calculated with AMOEBA using two sets of polarizabilities (**Table S6**); atomic polarizability parameters and performance statistics for the charged molecules (**Table S7** and **S8**).

Figures and Tables

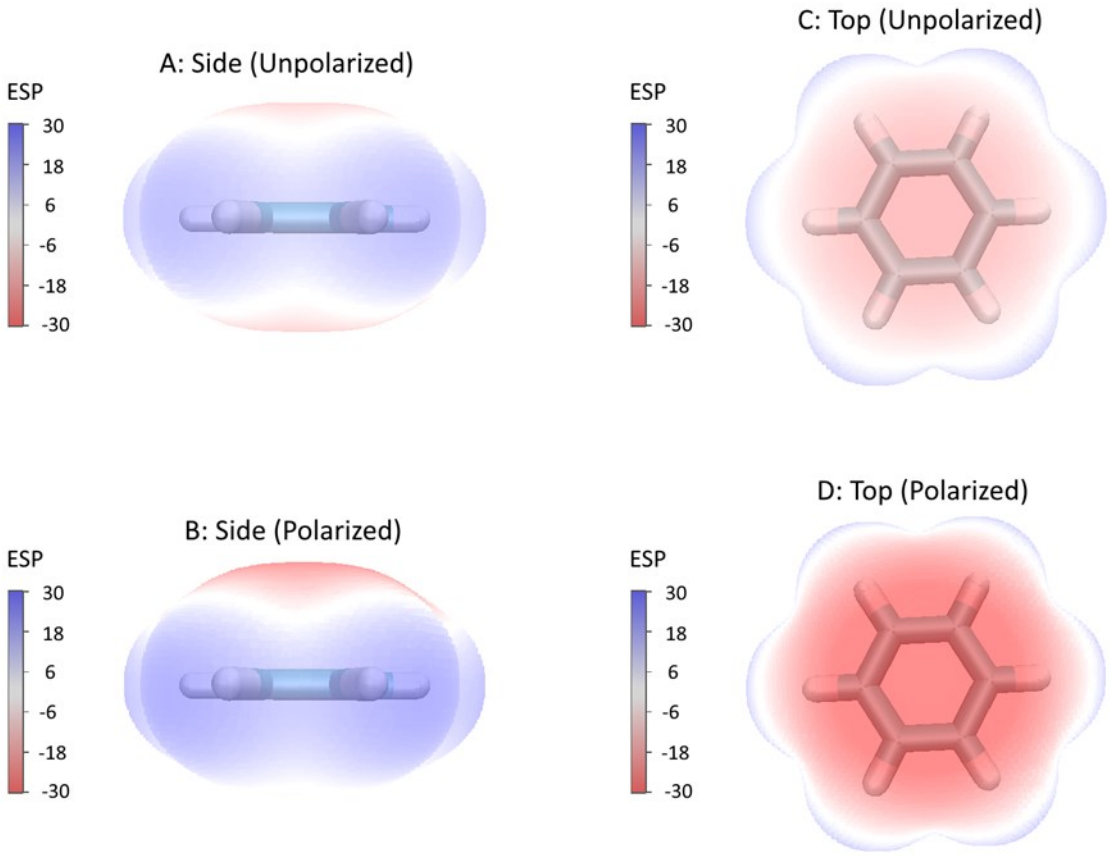


Figure 1. Molecular surface of benzene colored by electrostatic potential (ESP) in kcal/mol/e. (A,C) Unpolarized ESP surface and (B,D) Polarized ESP surface by the addition of a +1 electron charge above the ring (probe charge ESP not directly included). Calculations were performed at the MP2/aug-cc-pVTZ level using Gaussian 16 software.

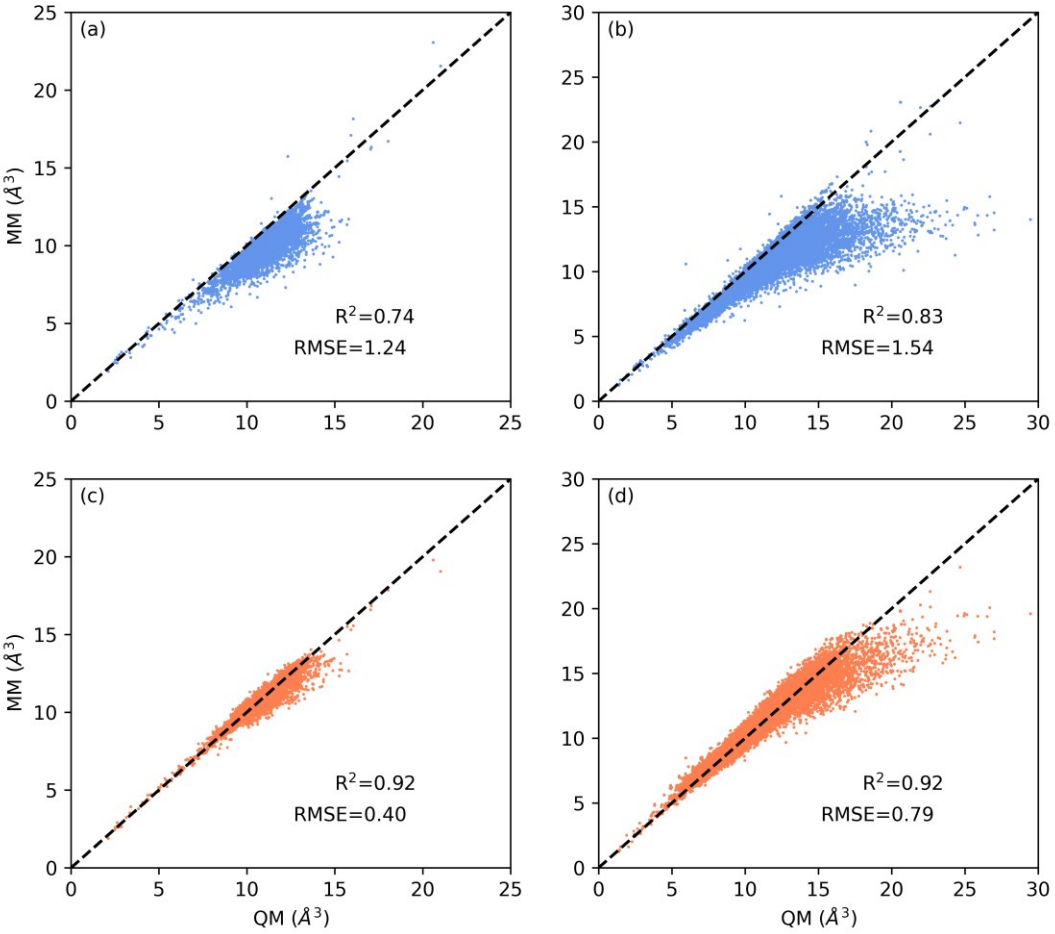


Figure 2. Correlation between QM and MM molecular polarizabilities. The MM model values were calculated with two sets of parameters (blue: current AMOEBA; orange: this work). Both the isotropic molecular polarizability (a) and (c), and three eigenvalues (b) and (d) are plotted. Black dotted lines represent perfect correlation. Correlation coefficient (R^2) and RMSE values (in $Å^3$) are reported.

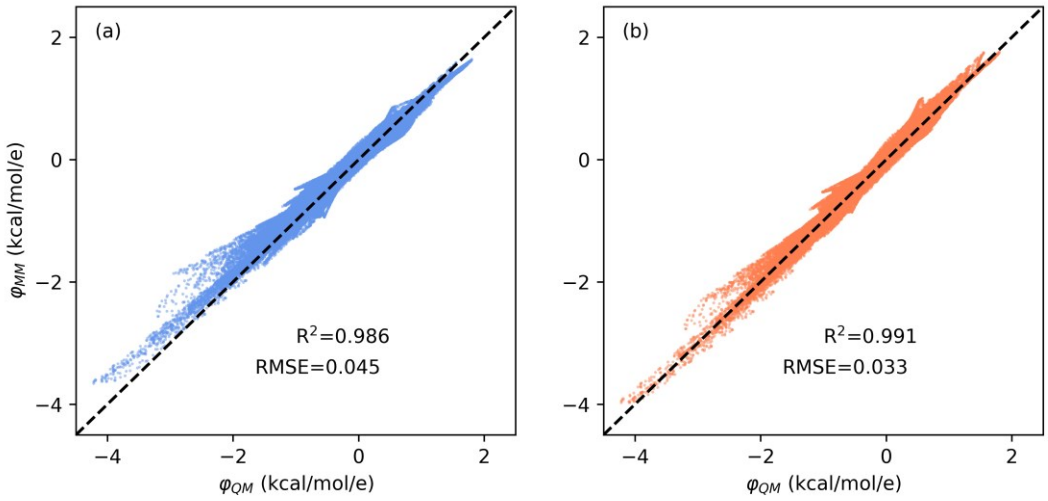


Figure 3. Correlation between the QM and MM electrostatic potential changes due to external field (i.e. ESP responses). The MM values were calculated with two sets of parameters: (a) the current AMOEBA, and (b) this work. Black dotted lines represent perfect correlation. Correlation coefficient (R^2) and RMS error (in kcal/mol/e) values are reported.

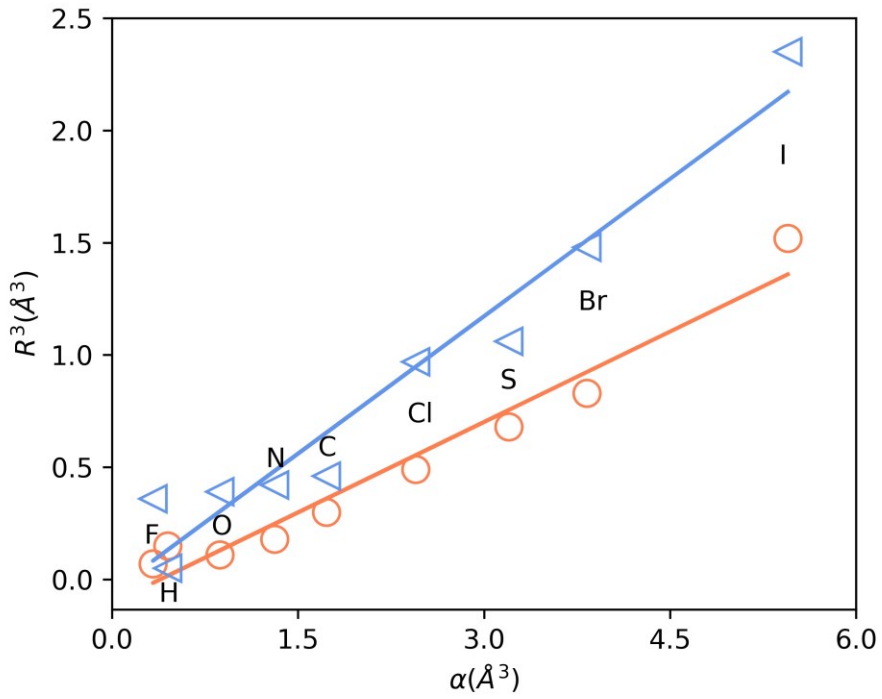


Figure 4. Correlation between the final atomic polarizability parameters and atomic radius (orange circles) or covalent radius (blue triangles). Polarizability parameters are averaged over atom types for each element. Solid lines are the linear fit of the data. Atomic and covalent radii are taken from Ref. 49 and 50, respectively.

Table 1. Atomic polarizabilities (in Å^3) obtained from fitting against molecular polarizabilities (*fit-polar*) and electrostatic potential responses (*fit-esp*) respectively. The *final* parameters are the result of averaging *fit-polar* and *fit-esp* sets.

Index	Atom type	Short description [a]	*fit-polar*	*fit-esp*	*final*
1	HW	Water H	0.43000	0.43190	0.42830 [b]
2	Har	H on aromatic and double bonded atoms	0.45938	0.40413	0.43175
3	Hnonpol	H on carbon atoms	0.47085	0.48982	0.48034
4	Hpolar	H on N, O, S, and halogens	0.47167	0.44294	0.45730
5	Car	Aromatic and double bonded C	2.11333	2.01569	2.06451
6	Cch4	Methane C	1.33467	1.34778	1.34123
7	Cnonpol	Nonpolar C	1.45458	1.37539	1.41499
8	Cpolar	Polar C	1.65038	1.58890	1.61964
9	Csp	Triple bonded C	2.27642	2.12827	2.20235
10	Nnh3	Ammonia N	1.07512	1.23636	1.15574
11	Nar	Aromatic N	1.64607	1.75761	1.70184
12	Nnonpol	N on C atoms	1.23564	1.13368	1.18466
13	Npolar	N on non-carbon atoms	1.44144	1.44604	1.44374
14	Nsp	Triple bonded N	1.15924	1.15674	1.15799
15	Nsp2	Double bonded N	1.19623	1.29033	1.24328
16	OW	Water O	0.97102	0.97393	0.97635 [b]
17	Oar	Aromatic O	0.83226	0.83226	0.83226
18	Ononpol	O on carbon atoms	0.81831	0.80617	0.81224
19	Opolar	O on other atoms	0.85184	0.86322	0.85753
20	Ohacid	O of acid (with H)	0.81831	0.84635	0.83233
21	Osp2c	O on carbonyl	0.85551	0.97202	0.91377
22	O-sulf	O single bonded S	0.83413	0.85888	0.84651
23	O=sulf	O double bonded S	0.83394	0.88365	0.85880
24	Sar	Aromatic S	3.27951	3.13221	3.20586
25	Snonpol	S on carbon atoms	3.54925	3.26305	3.40615
26	Spolar	S on other atoms	2.99663	2.99151	2.99407
27	Fnonar	Nonaromatic F	0.36896	0.32725	0.34810
28	Far	Aromatic F	0.29050	0.32680	0.30865
29	Clar	Aromatic Cl	2.58795	2.54576	2.56686
30	Clnonar	Nonaromatic Cl	2.36596	2.30821	2.33709
31	Brar	Aromatic Br	4.20104	4.06331	4.13218
32	Brnonar	Nonaromatic Br	3.57711	3.47471	3.52591
33	Iar	Aromatic I	5.48163	5.48163	5.48163
34	Inonar	Nonaromatic I	5.42354	5.42354	5.42354

a. Refer to **Table S2** for the full description and SMART patterns

b. Final atomic polarizability parameters of water (atom type: HW and OW) were made consistent with the optimized values using both gas and liquid water properties.[51]

Table 2. Performance of the *fit-polar* set of parameters on the *training* and *validation* polarizability sets. Root mean square error (RMSE, in $Å^3$) and unsigned mean percentage error (UMPE, defined in **Eq. 3**) are reported.

Statistics [a,b]	***Training* set** (773 molecules)	***Validation* set** (6511 molecules)
RMSE (α_{ii})	0.85	0.78
UMPE (α_{ii})	4.7%	4.4%
RMSE (α_{iso})	0.41	0.39
UMPE (α_{iso})	2.8%	2.6%

a. α_{ii}: statistics based on three eigenvalues, where $i = x, y, z$.
b. α_{iso}: statistics based on the isotropic polarizabilities, i.e., $\frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$.

Table 3. Performance of the *fit-polar*, *fit-esp*, *final* and the current AMOEBA sets of polarizability parameters on the ESP and molecular polarizability sets. Root mean square error (RMSE, in kcal/mol/e for ESP and in $Å^3$ for polarizability) and unsigned mean percentage error (UMPE, defined in **Eq. 3**) are reported.

Parameters	**ESP set** (316 probing structures)		**Polarizability set** [a,b] (7284 molecules)	
fit-polar	RMSE	0.0347	RMSE (α_{ii})	0.79
			UMPE (α_{ii})	4.5%
			RMSE (α_{iso})	0.39
			UMPE (α_{iso})	2.6%
fit-esp	RMSE	0.0326	RMSE (α_{ii})	0.81
			UMPE (α_{ii})	4.2%
			RMSE (α_{iso})	0.43
			UMPE (α_{iso})	2.6%
final	RMSE	0.0333	RMSE (α_{ii})	0.79
			UMPE (α_{ii})	4.2%
			RMSE (α_{iso})	0.40
			UMPE (α_{iso})	2.5%
AMOEBA	RMSE	0.0451	RMSE (α_{ii})	1.54
			UMPE (α_{ii})	8.8%
			RMSE (α_{iso})	1.24
			UMPE (α_{iso})	9.3%

a. α_{ii}: statistics based on three eigenvalues, where $i = x, y, z$.
b. α_{iso}: statistics based on the isotropic polarizabilities, i.e., $\frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$.

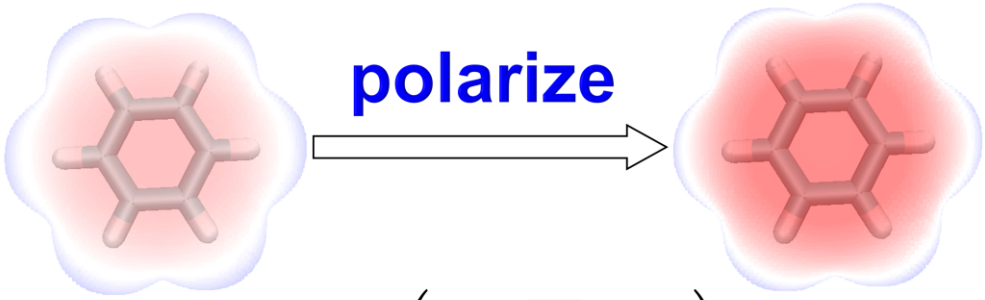
Reference

1. Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P., Polarizable force fields for biomolecular simulations: Recent advances and applications. *Annu. Rev. Biophys* **2019,** *48* 371-394.
2. Melcr, J.; Piquemal, J.-P., Accurate Biomolecular Simulations Account for Electronic Polarization. *Front. Mol. Biosci.* **2019,** *6* 1-8.
3. Inakollu, V. S. S.; Geerke, D. P.; Rowley, C. N.; Yu, H., Polarisable force fields: what do they add in biomolecular simulations? *Curr. Opin. Struct. Biol.* **2020,** *61* 182-190.
4. Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D., An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **2016,** *116* 4983-5013.
5. Shi, Y.; Laury, M. L.; Wang, Z.; Ponder, J. W., AMOEBA binding free energies for the SAMPL7 TrimerTrip host–guest challenge. *J. Comput. Aided Mol. Des.* **2021,** *35* 79-93.
6. Bell, D. R.; Qi, R.; Jing, Z.; Xiang, J. Y.; Mejias, C.; Schnieders, M. J.; Ponder, J. W.; Ren, P., Calculating binding free energies of host–guest systems using the AMOEBA polarizable force field. *Phys. Chem. Chem. Phys.* **2016,** *18* 30261-30269.
7. Khoury, L. E.; Jing, Z.; Cuzzolin, A.; Deplano, A.; Loco, D.; Sattarov, B.; Hédin, F.; Wendeborn, S.; Ho, C.; Ahdab, D. E., Computationally driven discovery of SARS-CoV-2 Mpro inhibitors: from design to experimental validation. *arXiv preprint arXiv:2110.05427* **2021**.
8. Ren, P. Y.; Ponder, J. W., Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003,** *107* 5933-5947.
9. Cieplak, P.; Caldwell, J.; Kollman, P., Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J Comput Chem* **2001,** *22* 1048-1057.
10. 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* 1960-1986.
11. Rick, S. W.; Berne, B. J., Dynamical Fluctuating Charge Force Fields: The Aqueous Solvation of Amides. *J. Am. Chem. Soc.* **1996,** *118* 672-679.
12. Huang, J.; Simmonett, A. C.; Pickard, F. C. t.; MacKerell, A. D., Jr.; Brooks, B. R., Mapping the Drude polarizable force field onto a multipole and induced dipole model. *J Chem Phys* **2017,** *147* 161702.
13. Thole, B. T., Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chem Phys* **1981,** *59* 341-350.
14. van Duijnen, P. T.; Swart, M., Molecular and Atomic Polarizabilities: Thole's Model Revisited. *J Phys Chem A* **1998,** *102* 2399-2407.
15. Ren, P.; Wu, C.; Ponder, J. W., Polarizable Atomic Multipole-based Molecular Mechanics for Organic Molecules. *J Chem Theory Comput* **2011,** *7* 3143-3161.
16. 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* 2751--2761.
17. Wang, J.; Cieplak, P.; Luo, R.; Duan, Y., Development of Polarizable Gaussian Model for Molecular Mechanical Calculations I: Atomic Polarizability Parameterization To Reproduce ab Initio Anisotropy. *J. Chem. Theory Comput.* **2019,** *15* 1146-1158.
18. Wang, J.; Cieplak, P.; Li, J.; Hou, T.; Luo, R.; Duan, Y., Development of Polarizable Models for Molecular Mechanical Calculations I: Parameterization of Atomic Polarizability. *J. Phys. Chem. B* **2011,** *115* 3091-3099.
19. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J Comput Chem* **2004,** *25* 1157-1174.

20. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A., Development of an Accurate and Robust Polarizable Molecular Mechanics Force Field from ab Initio Quantum Chemistry. *J. Phys. Chem. A* **2004,** *108* 621-627.
21. 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* 1515-1531.
22. Elking, D.; Darden, T.; Woods, R. J., Gaussian induced dipole polarization model. *J Comput Chem* **2007,** *28* 1261-1274.
23. Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D., Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2005,** *1* 153-168.
24. Wilkins, D. M.; Grisafi, A.; Yang, Y.; Lao, K. U.; DiStasio, R. A.; Ceriotti, M., Accurate molecular polarizabilities with coupled cluster theory and machine learning. *Proc Natl Acad Sci U S A* **2019,** *116* 3401-3406.
25. Applequist, J.; Carl, J. R.; Fung, K.-K., Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.* **1972,** *94* 2952-2960.
26. Jing, Z.; Liu, C.; Qi, R.; Ren, P., Many-body effect determines the selectivity for Ca(2+) and Mg(2+) in proteins. *Proc Natl Acad Sci U S A* **2018,** *115* E7495-E7501.
27. van Duijnen, P. T.; Swart, M., Molecular and Atomic Polarizabilities: Thole's Model Revisited. *J Phys Chem A* **1998,** *102* 2399-2407.
28. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993,** *97* 10269-10280.
29. Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C., ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2019**.
30. Wu, J. C.; Chattree, G.; Ren, P., Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012,** *131* 1138.
31. Ren, P.; Ponder, J. W., Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *Journal of Computational Chemistry* **2002,** *23* 1497-1506.
32. 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* 2609-2618.
33. Liu, C.; Piquemal, J.-P.; Ren, P., AMOEBA+ Classical Potential for Modeling Molecular Interactions. *J. Chem. Theory Comput.* **2019,** *15* 4122-4139.
34. Bosque, R.; Sales, J., Polarizabilities of Solvents from the Chemical Composition. *Journal of Chemical Information and Computer Sciences* **2002,** *42* 1154-1163.
35. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A. 02*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

36. Ditchfield, R.; Hehre, W. J.; Pople, J. A., Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971,** *54* 724-728.
37. Kendall, R. A.; Jr., T. H. D.; Harrison, R. J., Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992,** *96* 6796-6806.
38. Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005,** *7* 3297-3305.
39. Weigend, F., Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006,** *8* 1057-1065.
40. Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L., New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community. *J Chem Inf Model* **2019,** *59* 4814-4820.
41. Daylight Chemical Information Systems: Santa Fe, New Mexico.
42. Das, A. K.; Demerdash, O. N.; Head-Gordon, T., Improvements to the AMOEBA Force Field by Introducing Anisotropic Atomic Polarizability of the Water Molecule. *J Chem Theory Comput* **2018**.
43. DelloStritto, M.; Klein, M. L.; Borguet, E., Bond-Dependent Thole Model for Polarizability and Spectroscopy. *J. Phys. Chem. A* **2019,** *123* 5378-5387.
44. Lambros, E.; Paesani, F., How good are polarizable and flexible models for water: Insights from a many-body perspective. *J. Chem. Phys.* **2020,** *153* 060901.
45. Heindel, J. P.; Xantheas, S. S., The Many-Body Expansion for Aqueous Systems Revisited: I. Water–Water Interactions. *J. Chem. Theory Comput.* **2020,** *16* 6843-6855.
46. Řezáč, J.; Huang, Y.; Hobza, P.; Beran, G. J. O., Benchmark Calculations of Three-Body Intermolecular Interactions and the Performance of Low-Cost Electronic Structure Methods. *J. Chem. Theory Comput.* **2015,** *11* 3065-3079.
47. Automated Parameter Assignment for AMOEBA+ Model https://github.com/prenlab/amoebaplus_parameter. Accessed: Dec. 10, 2021
48. PolType2: Automated Parameterization of Advanced Force Fields. https://github.com/TinkerTools/poltype2. Accessed: Sep. 8, 2021
49. Clementi, E.; Raimondi, D.-L., Atomic screening constants from SCF functions. J. *Chem. Phys. 1***963,** 3*8* 2686-2689.
50. Vainshtein, B. K.; Friedkin, V. M.; Indenbom, V. L., S*tructure of crystals.* Springer Science & Business Media: 2013.
51. Liu, C.; Piquemal, J.-P.; Ren, P., Implementation of Geometry-Dependent Charge Flux into the Polarizable AMOEBA+ Potential. J. *Phys. Chem. Lett 2***020,** 1*1* 419-426.

TOC graphic



$$\mu_i = \alpha_i \left(E_i - \sum_{i \neq j} T_{ij} \mu_j \right)$$

Supporting Information

Atomic Polarizabilities for Interactive Dipole Induction Models

Jacob M. Litman, [†] *Chengwen Liu,* [†] *and Pengyu Ren**

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

[†] J.M.L and C.L. contributed equally to this work

[*] Corresponding author. Email: pren@mail.utexas.edu

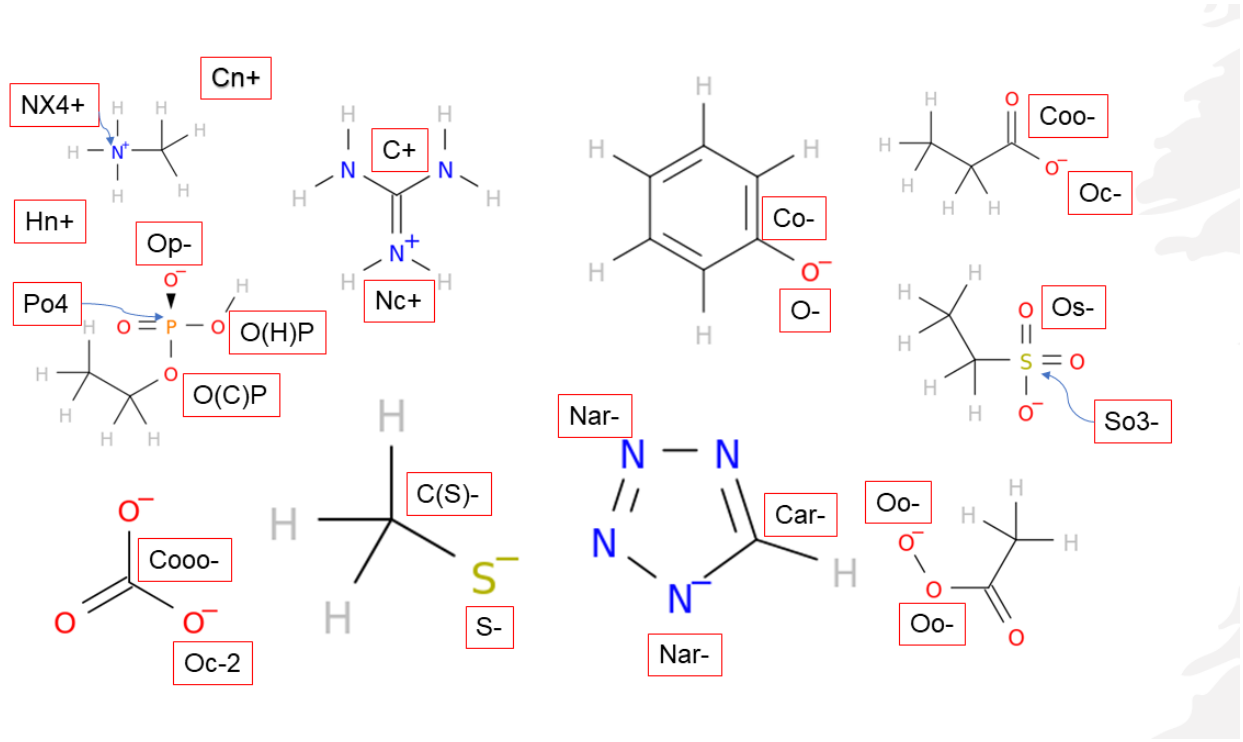


Figure S1. An illustration of the atom types for charged molecules. The same atom type is assigned to chemically equivalent atoms in a molecule (for example three nitrogen atoms are all "Nc+" in guanidinium ion). For the remaining atoms in a molecule, SMARTS pattern matching for the neutral molecules is applied.

Table S1. Atomic polarizabilities of the current AMOEBA model.

Index	SMARTS	α (Å^3)	Description
1	[#1]	0.496	Hydrogen
2	[#1][c,n]	0.696	Hydrogens on aromatic carbon/nitrogen
3	[C]	1.334	Non-aromatic carbon
4	[c]	1.750	Aromatic carbon
5	[#7]	1.073	Nitrogen
6	[#8]	0.837	Oxygen
7	[#16]	2.800	Sulfur
8	[#15]	1.788	Phosphorus
9	[#9]	0.600	Florine
10	[#17]	2.500	Chloride
11	[#35]	3.595	Bromine
12	[#53]	5.705	Iodine

Table S2. Atom types are defined by SMARTS strings for the neutral molecules. The first atom in the smarts string is of interest. The more detailed matching takes the priority and overrides a courser matching for a specific atom.

Element	SMARTS	Atom type	Description
H	[H]	H*	wild matching H
	[H][C]	Hnonpol	H on non-aromatic C
	[H][O]	Hpolar	H on non-aromatic O
	[H][N]	Hpolar	H on non-aromatic N
	[H][S]	Hpolar	H on non-aromatic S
	[H][NH3]	Hpolar	H on ammonia
	[H][F,Cl,Br,I]	Hpolar	H on HX (X=F, Cl, Br, and I)
	[H][c]	Har	H on aromatic C
	[H][n]	Har	H on aromatic N
	[H][o]	Har	H on aromatic O
	[H][C]=[C]	Har	H on C=C system
	[H][OH2]	HW	H on water
C	[C]	C*	wild matching C
	[$([C][C])]	Cnonpol	C on non-aromatic C
	[C][c]	Cnonpol	C connected to aromatic C
	[C][O]	Cpolar	C on non-aromatic O
	[C][o]	Cpolar	C on aromatic O
	[C][N]	Cpolar	C on non-aromatic N
	[C][n]	Cpolar	C on aromatic N
	[C][F,Cl,Br,I]	Cpolar	C connected to F, Cl, Br, and I
	[C]=[N]	Cpolar	C connected to N via double bond
	[C][S]	Cpolar	C on S
	[c]	Car	C on aromatic system
	[CX3](=O)	Car	C on carbonyl O
	[$([CX3]=[CX3])]	Car	C on C=C
	[C]([C]=[*])([C]=[*]	Car	C on conjugated systems
	[CH4]	Cch4	C on methane
	[$([C]#[*])]	Csp	C on triple bond, e.g. C#C, C#N
N	[N]	N*	wild matching Nitrogen
	[N][C]	Nnonpol	N on non-aromatic C
	[N][c]	Nnonpol	N connected to aromatic C via single bond
	[N]=[c]	Nnonpol	N connected to aromatic C via double bond
	[N][S]	Npolar	N connected to S
	[$([N][N])]	Npolar	N connected to non-aromatic N via single bond
	[N]=[N]	Npolar	N connected to non-aromatic N via double bond
	[N][n]	Npolar	N connected to aromatic N via single bond
	[n]	Nar	N on aromatic system
	[N]=[C]	Nar	N on double bond N=C
	[NX3]([CX3](=O))	Nsp2	N on amide

	[N]#[*]	Nsp	N of triple bond, N#N, N#C, etc
	[NX3H3]	Nnh3	N on ammonia
O	[O]	O*	wild matching O
	[O][CX4]	Ononpol	O on non-polar C
	[O][c]	Ononpol	O connected to aromatic C via single bond
	[O]=[c]	Ononpol	O connected to aromatic C via double bond
	[O][N]	Opolar	O connected to non-aromatic N
	[O][n]	Opolar	O connected to aromatic N
	[o]	Oar	O on aromatic system
	[OX2][CX3]	Oar	O connected to carbonyl group
	[OX1]=[CX3]	Osp2c	O on carbonyl group
	[OX2][S]	O-sulf	O on O-S bond
	[OX1]=[S]	O=sulf	O on O=S bond
	[OX2H2]	OW	O on Water
	[$([OH1][C]=[O])]	Ohacid	O of acid with H
S	[S]	S*	wild matching S
	[S][C]	Snonpol	S on non-polar C via single bond
	[S]=[C]	Snonpol	S on non-polar C via double bond
	[S][N]	Spolar	S connected to N
	[S][O]	Spolar	S connected to O
	[S]=[O]	Spolar	S connected to O via double bond
	[s]	Sar	S on aromatic system
P	[P]	P*	wild matching P
	[P](=[O])(O)([O])[O]	Po4	P on PO4
F	[F][A]	Fnonar	F connected to non-aromatic atom
	[F][a]	Far	F connected to aromatic atom
Cl	[Cl][A]	Clnonar	Cl connected to non-aromatic atom
	[Cl][a]	Clar	Cl connected to aromatic atom
Br	[Br][A]	Brnonar	Br connected to non-aromatic atom
	[Br][a]	Brar	Br connected to aromatic atom
I	[I][A]	Inonar	I connected to non-aromatic atom
	[I][a]	Iar	I connected to aromatic atom

Table S3. Comparison of molecular polarizabilities calculated with LR-CCSD/aug-cc-pvdz (CCSD) and MP2/aug-cc-pvtz (MP2) at the same geometry for 52 representative molecules. Both the geometry and CCSD data are taken from reference. Polarizabilities are in $Å^3$.

Index	Molecule [a]	CCSD [a]	MP2	Diff.
1	QM7B_0041.xyz	5.06	5.03	0.02
2	QM7B_0067.xyz	12.75	12.58	0.18
3	QM7B_0100.xyz	12.54	12.52	0.02
4	QM7B_0131.xyz	5.01	4.99	0.02
5	QM7B_0158.xyz	9.65	9.60	0.05
6	QM7B_0191.xyz	10.44	10.42	0.02
7	QM7B_0387.xyz	9.15	8.98	0.18
8	QM7B_0396.xyz	12.88	12.76	0.12
9	QM7B_0425.xyz	8.88	8.73	0.15
10	QM7B_0474.xyz	7.99	7.89	0.10
11	QM7B_0533.xyz	10.44	10.35	0.09
12	QM7B_0616.xyz	12.35	12.30	0.05
13	QM7B_0651.xyz	11.05	11.02	0.03
14	QM7B_0679.xyz	14.89	15.19	-0.30
15	QM7B_0965.xyz	11.25	11.25	0.00
16	QM7B_0967.xyz	8.26	8.19	0.07
17	QM7B_1218.xyz	10.44	10.32	0.12
18	QM7B_1314.xyz	8.04	7.99	0.05
19	QM7B_1353.xyz	9.82	9.78	0.05
20	QM7B_1385.xyz	9.57	9.46	0.11
21	QM7B_1398.xyz	9.03	8.91	0.11
22	QM7B_1460.xyz	10.80	10.69	0.11
23	QM7B_1505.xyz	9.97	9.89	0.08
24	QM7B_1597.xyz	10.16	10.14	0.02
25	QM7B_1626.xyz	9.25	9.14	0.11
26	QM7B_1679.xyz	9.66	9.57	0.09
27	QM7B_1801.xyz	10.90	10.89	0.01
28	QM7B_1828.xyz	10.11	10.09	0.03
29	QM7B_1883.xyz	9.24	9.23	0.01
30	QM7B_2079.xyz	6.33	6.29	0.04
31	QM7B_2116.xyz	10.04	9.97	0.07
32	QM7B_2250.xyz	8.01	7.94	0.07
33	QM7B_2441.xyz	8.73	8.68	0.05
34	QM7B_2515.xyz	9.58	9.46	0.12
35	QM7B_2577.xyz	9.77	9.76	0.00
36	QM7B_2910.xyz	10.91	10.88	0.03
37	QM7B_3066.xyz	9.45	9.40	0.05
38	QM7B_3109.xyz	9.76	9.71	0.04
39	QM7B_3481.xyz	10.12	10.10	0.02
40	QM7B_4044.xyz	9.71	9.61	0.10

41	QM7B_4092.xyz	10.30	10.26	0.04
42	QM7B_4212.xyz	12.65	12.57	0.09
43	QM7B_4256.xyz	9.35	9.30	0.05
44	QM7B_4387.xyz	9.90	9.84	0.06
45	QM7B_4561.xyz	9.81	9.71	0.10
46	QM7B_4828.xyz	10.23	10.19	0.04
47	QM7B_4955.xyz	10.50	10.50	-0.01
48	QM7B_5334.xyz	10.87	10.85	0.02
49	QM7B_5352.xyz	9.95	9.92	0.03
50	QM7B_5397.xyz	10.41	10.34	0.07
51	QM7B_6342.xyz	12.71	12.70	0.01
52	QM7B_6800.xyz	12.39	12.31	0.08
	Mean signed difference			**0.05**

a. The molecular geometry was taken from QM7b dataset without further optimization; CCSD data were taken from Wilkins et al. [1]

Table S4. Mean molecular polarizability of common gas molecules calculated from MP2/aug-cc-pvtz and AMOEBA model with the *final* set of parameters.

Molecule	MP2	Model	Diff.	% Diff
CO_2	2.70	2.84	0.14	5.1
CO	3.09	2.85	-0.24	-7.7
H_2	0.77	0.75	-0.02	-2.1
N_2	1.75	1.88	0.13	7.2
NO_2	2.66	2.62	-0.04	-1.5
O_2	1.39	1.43	0.04	3.2

Table S5. Molecular polarizabilities (in Å^3) from experiment and the improved AMOEBA model. Isotropic molecular polarizabilities are highlighted in bold.

Index	Molecule	Expt.	This work			
		α_{iso}	α_{xx}	α_{yy}	α_{zz}	α_{iso}
1	3-methyltetrahydropyran	**11.71**	10.21	11.32	12.43	**11.32**
2	chloromethyloxirane	**8.19**	7.18	7.57	9.54	**8.10**
3	propylene carbonate	**8.55**	7.65	8.87	10.47	**9.00**
4	1,1,1,3,3,3-hexafluoro-2-propanol	**7.20**	6.25	7.12	7.72	**7.03**
5	1,1,1-trichloroethane	**10.44**	9.50	10.04	10.04	**9.86**
6	1,1,1-trichlorotrifluoroethane	**10.47**	9.58	10.18	10.18	**9.98**
7	1,1,2,2-tetrachloroethane	**12.21**	9.31	13.08	13.25	**11.88**
8	1,1,2-trichloro-1,2,2-trifluoro-ethane	**10.43**	9.14	10.17	10.48	**9.93**
9	1,1,2-trichloroethane	**10.33**	8.21	10.07	11.81	**10.03**
10	1,1,3,3-tetraethylurea	**20.17**	17.27	19.93	23.72	**20.30**
11	1,1,3,3-tetramethylurea	**12.86**	10.53	13.12	15.65	**13.10**
12	1,1-dichloroethylene	**8.36**	5.91	8.63	8.91	**7.82**
13	1,1-dichloroethane	**8.38**	6.86	8.38	8.87	**8.04**
14	1,10-dichlorodecane	**23.04**	18.28	20.53	31.37	**23.39**
15	1,2,3,4-tetrahydronaphthalene	**17.05**	11.54	17.93	21.61	**17.03**
16	1,2,3,4-tetramethylbenzene	**17.81**	11.89	19.88	22.48	**18.08**
17	1,2,3-propanetriol	**8.14**	7.43	7.92	9.58	**8.31**
18	1,2,3-trichloropropane	**12.09**	10.38	11.96	13.21	**11.85**
19	1,2,4-trichlorobenzene	**16.32**	10.78	17.67	21.88	**16.78**
20	1,2-butanediol	**9.32**	8.05	8.91	11.09	**9.35**
21	1,2-diaminoethane	**7.21**	6.31	6.71	7.58	**6.87**
22	1,2-dibromoethane	**10.75**	8.73	9.22	14.15	**10.70**
23	1,2-dibromopropane	**12.54**	10.07	11.62	15.57	**12.42**
24	1,2-dichloro-1,1,2,3,3,3-hexafluoropropane	**10.52**	9.18	10.05	10.71	**9.98**
25	1,2-dichlorobenzene	**14.30**	9.40	16.21	17.95	**14.52**
26	1,2-dichloroethane	**8.43**	6.82	7.33	10.50	**8.22**
27	1,2-difluorobenzene	**10.42**	6.70	12.10	12.27	**10.36**
28	1,2-dimethoxybenzene	**15.69**	10.69	17.81	19.24	**15.91**
29	1,2-dimethoxyethane	**9.60**	8.09	8.36	13.38	**9.94**
30	1,2-Ethanediol	**5.72**	5.06	6.05	6.15	**5.75**
31	1,3,5-trifluorobenzene	**10.24**	6.70	12.26	12.26	**10.40**
32	1,3-butanediol	**9.37**	8.01	9.12	10.90	**9.34**
33	1,3-dibromobenzene	**16.59**	11.90	18.53	23.22	**17.88**
34	1,3-dichlorobenzene	**14.38**	9.50	15.76	18.66	**14.64**
35	1,3-dichloropropane	**10.17**	9.11	9.35	11.22	**9.89**
36	1,3-Dimethyl-2-oxohexahydropyrimidine	**13.80**	10.57	15.08	16.81	**14.15**
37	1,3-dimethylimidazolidin-2-one	**12.14**	9.24	12.70	15.09	**12.34**
38	1,3-dioxolane	**6.79**	5.99	6.80	7.93	**6.91**

39	1,3-propanediol	**7.55**	6.90	7.08	8.70	**7.56**
40	1,4-butanediol	**9.35**	7.97	8.59	11.83	**9.46**
41	1,4-cyclohexadiene	**10.48**	7.88	12.52	12.70	**11.03**
42	1,4-dichlorobutane	**12.02**	10.51	11.04	13.34	**11.63**
43	1,4-difluorobenzene	**10.27**	6.70	12.02	12.36	**10.36**
44	1,4-dioxane	**8.63**	7.38	8.77	10.09	**8.75**
45	1,5-pentanediol	**11.20**	9.96	10.43	12.83	**11.07**
46	1,8-cineole	**18.10**	16.86	17.03	19.01	**17.64**
47	1-bromobutane	**11.26**	9.37	9.90	13.91	**11.06**
48	1-butanol	**8.79**	7.38	7.99	10.56	**8.64**
49	1-chloronaphthalene	**19.37**	11.26	21.80	23.93	**19.00**
50	1-decanol	**19.83**	15.67	16.83	28.04	**20.18**
51	1-heptanol	**14.30**	11.55	12.40	19.08	**14.34**
52	1-hexanol	**12.46**	10.59	11.46	14.48	**12.18**
53	1-hexene	**11.79**	9.61	10.16	14.95	**11.57**
54	1-hexyne	**11.09**	9.07	9.93	14.88	**11.29**
55	1-iodonaphthalene	**22.49**	11.68	22.56	24.44	**19.56**
56	1-iodopropane	**11.50**	8.90	9.32	13.54	**10.59**
57	1-methyl-2-pyrrolidinone	**10.66**	8.46	11.25	12.55	**10.76**
58	1-methylimidazole	**9.24**	6.80	10.18	11.99	**9.66**
59	1-methylpiperidine	**12.62**	10.59	12.38	12.90	**11.96**
60	1-methylnaphthalene	**19.42**	11.23	21.43	23.69	**18.79**
61	1-nitropropane	**8.60**	7.64	8.03	8.90	**8.19**
62	1-octanol	**16.14**	12.92	13.89	22.02	**16.28**
63	1-pentanol	**10.61**	8.83	9.49	13.00	**10.44**
64	1-phenylethanol	**14.62**	10.18	15.57	18.89	**14.88**
65	1-propanol	**6.96**	5.98	6.41	8.02	**6.80**
66	2,2,2-trichloroethanol	**11.00**	10.15	10.49	11.18	**10.61**
67	2,2,2-trifluoroethanol	**5.20**	4.67	5.11	5.60	**5.13**
68	2,2,3,3-tetrafluoro-1-propanol	**7.08**	6.28	6.63	8.14	**7.02**
69	2,2,4,4-tetramethyl-3-pentanone	**17.37**	15.60	16.04	19.16	**16.93**
70	2,2,5,5-tetramethyltetrahydrofuran	**15.40**	13.65	14.51	16.71	**14.96**
71	2,3,4,5,6-pentafluorotoluene	**12.24**	8.09	13.67	15.75	**12.50**
72	2,3-butanediol	**9.33**	8.26	9.11	10.50	**9.29**
73	2,4,6-collidine	**15.48**	10.55	18.43	18.46	**15.81**
74	2,4-dimethyl-3-pentanone	**13.70**	12.07	12.86	15.81	**13.58**
75	2,4-lutidine	**13.40**	9.19	14.78	17.25	**13.74**
76	2,4-pentanedione	**10.99**	9.32	10.24	12.50	**10.69**
77	2,5-dibromotoluene	**18.72**	13.08	19.64	26.84	**19.85**
78	2,6-di-tert-butylpyridine	**24.98**	19.23	24.75	29.71	**24.56**
79	2,6-difluoropyridine	**9.43**	6.42	11.06	11.86	**9.78**
80	2,6-dimethyl-4-heptanone	**17.50**	14.81	16.17	20.73	**17.23**
81	2,6-lutidine	**13.54**	9.20	14.82	17.29	**13.77**
82	2-(hydroxymethyl)furan	**9.84**	7.46	10.34	12.83	**10.21**

83	2-(n-butoxy)ethanol	**13.10**	11.89	12.33	15.14	**13.12**
84	2-(tert-butyl)phenol	**18.60**	13.51	19.26	22.32	**18.37**
85	2-ethoxyethanol	**9.48**	8.07	8.63	12.46	**9.72**
86	2-amino-1-butanol	**10.09**	8.14	9.92	11.60	**9.89**
87	2-aminoethanol	**6.48**	5.69	6.18	7.08	**6.31**
88	2-bromopyridine	**12.78**	9.09	13.61	17.85	**13.52**
89	2-butanol	**8.77**	7.59	8.15	9.95	**8.56**
90	2-butanone	**8.25**	6.79	7.95	10.14	**8.29**
91	2-chloroaniline	**14.05**	8.71	15.42	17.15	**13.76**
92	2-chloroethanol	**7.03**	6.01	6.91	7.89	**6.94**
93	2-chloropyridine	**11.60**	7.86	12.47	15.38	**11.90**
94	2-cyanoethanol	**6.92**	6.13	6.56	9.09	**7.26**
95	2-ethylphenol	**14.60**	10.26	15.90	18.28	**14.81**
96	2-hexanone	**11.95**	9.57	10.93	15.73	**12.07**
97	2-iodopropane	**11.67**	8.89	10.22	12.24	**10.45**
98	2-isopropylphenol	**16.40**	11.84	17.65	20.44	**16.64**
99	2-methoxyethanol	**7.62**	6.57	7.37	9.35	**7.76**
100	2-methoxyphenol	**13.60**	8.88	15.04	17.81	**13.91**
101	2-methyl-1-propanol	**8.81**	7.34	8.76	9.45	**8.52**
102	2-methyl-2-butanol	**10.64**	9.35	10.00	11.45	**10.27**
103	2-methyl-2-nitropropane	**10.30**	9.44	10.12	10.23	**9.93**
104	2-methylbutane	**10.13**	8.26	9.52	10.85	**9.54**
105	2-nitropropane	**8.59**	7.62	8.17	8.84	**8.21**
106	2-pentanol	**10.57**	9.29	9.87	11.68	**10.28**
107	2-pentanone	**10.11**	8.18	9.48	12.86	**10.17**
108	2-picoline	**11.57**	7.83	12.43	14.94	**11.73**
109	allyl alcohol	**6.71**	5.84	6.34	8.14	**6.78**
110	2-propenenitrile	**6.24**	4.85	6.09	8.13	**6.36**
111	2-propyne-1-ol	**5.98**	5.36	5.97	8.18	**6.50**
112	3,3-dimethyl-2-butanone	**11.85**	10.68	11.50	12.85	**11.68**
113	3,4-lutidine	**13.34**	9.09	15.16	16.46	**13.57**
114	3-bromopyridine	**12.76**	9.07	13.50	17.75	**13.44**
115	3-chloro-1,1,1-trifluoropropane	**8.23**	6.98	7.42	10.08	**8.16**
116	3-chlorophenol	**13.60**	8.44	14.70	16.55	**13.23**
117	3-chloropropyne	**7.52**	6.27	7.05	9.83	**7.72**
118	3-ethyl-3-pentanol	**14.44**	12.16	14.00	15.28	**13.81**
119	3-heptanone	**13.70**	11.53	12.23	17.91	**13.89**
120	3-methoxypropanenitrile	**8.79**	7.58	8.15	12.31	**9.35**
121	3-methyl-1-butanol	**10.61**	8.72	10.28	12.09	**10.36**
122	3-methyl-2-butanone	**10.02**	8.78	9.76	11.47	**10.00**
123	3-methyl-2-oxazolidinone	**9.29**	7.56	10.14	11.50	**9.73**
124	3-pentanone	**10.03**	8.17	9.44	12.87	**10.16**
125	3-picoline	**11.50**	7.82	12.34	14.92	**11.69**
126	4-methylphenol	**13.20**	8.41	13.63	17.20	**13.08**

127	5-acetyl-5-methyl-1,3-dioxane	**14.26**	12.56	13.58	17.00	**14.38**
128	5-isopropyl-2-methylphenol	**18.60**	13.67	18.26	24.37	**18.77**
129	5-nonanone	**17.44**	14.86	17.40	18.82	**17.02**
130	N,N,N',N'-tetraethylsulfamide	**21.32**	19.09	21.70	24.45	**21.74**
131	tetramethylguanidine	**13.85**	11.13	13.76	15.96	**13.61**
132	N,N-diethylformamide	**11.56**	9.75	11.75	12.96	**11.49**
133	N,N-dimethylacetamide	**9.69**	7.57	10.27	11.44	**9.76**
134	N,N-dimethylaniline	**16.29**	10.67	16.25	19.99	**15.64**
135	N,N-dimethylbenzylamine	**17.53**	13.38	16.65	22.74	**17.59**
136	N,N-dimethylcyclohexylamine	**16.09**	13.45	15.91	17.25	**15.53**
137	N,N-dimethylformamide	**7.93**	6.25	8.60	9.00	**7.95**
138	N,N-dimethylthioformamide	**11.39**	7.83	10.09	12.34	**10.09**
139	N-methylacetamide	**7.85**	6.20	7.59	9.93	**7.91**
140	N-methylaniline	**14.21**	8.85	14.41	17.88	**13.71**
141	N-methylcyclohexylamine	**14.02**	12.07	13.57	15.48	**13.71**
142	N-methylformamide	**6.01**	4.78	6.09	7.27	**6.05**
143	N-methylformanilide	**15.89**	10.99	16.37	20.79	**16.05**
144	acetic acid	**5.18**	4.32	5.32	6.42	**5.36**
145	acetic anhydride	**8.93**	7.61	9.00	11.83	**9.48**
146	acetone	**6.47**	5.39	6.54	7.45	**6.46**
147	acetonitrile	**4.44**	4.14	4.14	5.76	**4.68**
148	acetophenone	**14.45**	9.30	15.41	18.98	**14.56**
149	allylamine	**7.57**	6.27	6.89	8.84	**7.33**
150	aniline	**12.16**	7.37	12.92	14.66	**11.65**
151	benzaldehyde	**12.70**	7.99	13.72	16.25	**12.65**
152	benzene	**10.44**	6.70	12.05	12.05	**10.27**
153	benzonitrile	**12.57**	8.12	13.34	16.38	**12.61**
154	benzophenone	**22.50**	15.47	22.13	31.60	**23.07**
155	benzoyl chloride	**14.70**	9.19	15.51	19.10	**14.60**
156	benzyl alcohol	**12.89**	8.91	13.79	16.39	**13.03**
157	biacetyl	**8.30**	6.69	8.67	10.75	**8.70**
158	bis(2-methoxyethyl)ether	**13.90**	11.51	12.17	20.98	**14.89**
159	bromobenzene	**13.52**	9.32	14.38	18.37	**14.03**
160	bromoform	**11.87**	9.53	12.73	12.73	**11.67**
161	bromotrichloromethane	**11.67**	10.83	10.83	11.74	**11.13**
162	butanenitrile	**8.12**	6.97	7.43	10.48	**8.29**
163	butyraldehyde	**8.20**	6.83	7.87	10.20	**8.30**
164	carbon disulfide	**8.51**	6.26	6.26	9.48	**7.34**
165	carbon tetrachloride	**10.54**	9.95	9.95	9.95	**9.95**
166	chloroacetonitrile	**6.40**	5.54	6.25	8.07	**6.62**
167	cinnamaldehyde	**17.50**	10.18	16.47	23.33	**16.66**
168	cis-1,2-dichloroethylene	**8.08**	5.92	8.13	9.28	**7.78**
169	cycloheptane	**12.84**	10.47	12.91	13.21	**12.20**
170	cycloheptatriene	**12.46**	9.22	13.78	14.12	**12.37**

171	cyclohexane	**11.04**	9.06	11.16	11.16	**10.46**
172	cyclohexanol	**11.40**	10.11	11.58	11.79	**11.16**
173	cyclohexanone	**11.12**	9.15	11.68	11.89	**10.91**
174	cyclohexene	**10.79**	8.40	11.56	11.57	**10.51**
175	cyclohexylamine	**12.47**	10.65	11.92	12.59	**11.72**
176	cyclohexylbenzene	**20.77**	16.19	18.91	27.07	**20.72**
177	cyclooctane	**14.62**	12.02	14.78	14.95	**13.92**
178	cyclooctatetraene	**14.07**	10.82	15.32	15.32	**13.82**
179	cyclopentane	**9.23**	7.72	9.24	9.25	**8.74**
180	cyclopentanol	**9.72**	8.32	9.67	10.55	**9.51**
181	cyclopropyl methyl ketone	**9.51**	8.02	8.72	11.13	**9.29**
182	di-n-butyl ether	**16.31**	14.54	14.78	18.82	**16.05**
183	di-n-butyl phthalate	**30.75**	22.80	34.49	37.86	**31.72**
184	di-n-butylamine	**16.90**	14.75	15.16	19.99	**16.64**
185	di-tert-butyl ether	**16.27**	14.75	15.05	17.15	**15.65**
186	di-tert-butyl sulfide	**19.31**	16.37	16.60	21.00	**17.99**
187	dibromomethane	**8.75**	7.29	8.03	10.65	**8.66**
188	dichloroacetic acid	**9.10**	8.35	9.16	9.71	**9.07**
189	dicyclopropyl ketone	**12.52**	10.60	10.85	15.13	**12.19**
190	diethanolamine	**10.70**	9.30	10.20	13.20	**10.90**
191	diethoxymethane	**11.40**	9.74	11.01	13.61	**11.45**
192	diethyl carbonate	**11.32**	9.98	10.24	15.40	**11.87**
193	diethyl disulfide	**14.59**	12.11	12.56	17.04	**13.90**
194	diethyl malonate	**15.11**	13.56	16.96	17.39	**15.97**
195	diethyl sulfate	**13.00**	11.23	12.09	17.54	**13.62**
196	diethyl sulfide	**11.38**	8.99	10.00	14.66	**11.22**
197	diethyl sulfite	**12.71**	10.33	12.76	15.99	**13.02**
198	diethylamine	**9.63**	7.82	8.53	11.92	**9.43**
199	diethylene glycol	**10.09**	8.56	9.66	13.14	**10.46**
200	diethyleneglycol dimethyl ether	**13.97**	11.51	12.17	20.98	**14.89**
201	diiodomethane	**12.96**	9.44	10.25	14.11	**11.27**
202	diisopropyl ether	**12.65**	10.74	12.14	14.13	**12.34**
203	diisopropylamine	**13.45**	11.40	12.19	14.99	**12.86**
204	dimethoxymethane	**7.68**	6.63	7.60	8.91	**7.71**
205	dimethyl carbonate	**7.56**	6.43	7.19	11.14	**8.25**
206	dimethyl disulfide	**10.84**	8.93	9.69	12.01	**10.21**
207	dimethyl phthalate	**19.59**	13.72	22.78	25.55	**20.68**
208	dimethyl sulfate	**8.88**	8.36	8.36	12.52	**9.75**
209	dimethyl sulfide	**7.63**	6.29	7.30	8.73	**7.44**
210	dimethyl sulfite	**8.94**	7.47	8.18	12.25	**9.30**
211	dimethyl sulfoxide	**8.03**	6.57	7.79	8.49	**7.62**
212	diphenyl ether	**21.02**	16.14	19.05	29.31	**21.50**
213	diphenylmethane	**21.90**	18.84	19.59	28.66	**22.36**
214	ethanol	**5.13**	4.53	4.99	5.50	**5.01**

215	ethyl acetate	**8.87**	7.85	8.24	11.35	**9.15**
216	ethyl acetoacetate	**12.82**	11.67	12.97	15.35	**13.33**
217	ethyl benzoate	**16.94**	11.83	17.09	23.50	**17.47**
218	ethyl chloroacetate	**10.76**	9.51	10.54	13.18	**11.08**
219	ethyl formate	**7.09**	6.36	6.77	8.62	**7.25**
220	ethyl lactate	**11.34**	9.98	11.00	14.07	**11.68**
221	ethyl propiolate	**10.08**	8.69	9.14	14.20	**10.67**
222	Ethyl trichloroacetate	**14.82**	13.50	13.82	17.27	**14.86**
223	ethyl vinyl ether	**8.74**	6.87	7.99	11.61	**8.82**
224	ethylbenzene	**14.24**	10.00	14.45	17.76	**14.07**
225	ethylene carbonate	**6.60**	5.85	7.09	8.76	**7.23**
226	ethylene glycol dimethyl ether	**9.58**	8.09	8.36	13.38	**9.94**
227	ethylene sulfite	**8.00**	7.13	7.90	9.75	**8.26**
228	ethyl propanoate	**10.60**	9.48	10.12	13.10	**10.90**
229	formamide	**4.22**	3.20	4.27	4.92	**4.13**
230	formic acid	**3.39**	2.86	3.69	4.13	**3.56**
231	furan	**7.35**	5.23	8.12	9.03	**7.46**
232	gamma-butyrolactone	**7.94**	6.79	8.33	9.49	**8.20**
233	hexachloro-1,3-butadiene	**19.75**	16.54	17.61	21.95	**18.70**
234	hexachlorocyclopentadiene	**20.67**	15.40	22.41	22.48	**20.10**
235	hexafluorobenzene	**10.56**	6.72	12.48	12.48	**10.56**
236	hexamethylphosphoric acid triamide	**19.00**	16.85	18.74	18.75	**18.11**
237	hexanenitrile	**11.84**	9.90	10.47	15.56	**11.98**
238	iodoethane	**9.66**	7.52	7.98	10.58	**8.69**
239	iodomethane	**7.68**	6.19	6.19	8.31	**6.90**
240	isoamyl acetate	**14.48**	13.31	13.44	16.59	**14.45**
241	isooctane	**15.63**	13.28	14.35	16.78	**14.80**
242	isopropylbenzene	**16.10**	12.02	15.66	19.94	**15.87**
243	m-xylene	**14.33**	9.44	15.46	17.95	**14.28**
244	mesitylene	**16.25**	10.79	19.11	19.11	**16.33**
245	methanesulfonic acid	**6.70**	6.44	6.73	7.60	**6.93**
246	methyl acetate	**7.00**	5.93	6.86	9.30	**7.36**
247	methyl acrylate	**8.85**	6.63	9.10	11.65	**9.13**
248	methyl benzoate	**15.06**	9.86	15.82	21.18	**15.62**
249	methyl butanoate	**10.67**	9.24	9.96	13.74	**10.98**
250	methyl formate	**5.21**	4.49	5.25	6.72	**5.49**
251	methyl hexanoate	**14.34**	11.69	13.06	19.89	**14.88**
252	methyl methacrylate	**10.60**	7.95	11.24	13.82	**11.00**
253	methyl octanoate	**18.09**	14.56	16.43	25.20	**18.73**
254	methyl pentanoate	**12.53**	10.11	11.34	17.71	**13.05**
255	methyl propionate	**8.82**	7.33	8.37	11.97	**9.22**
256	methyl salicylate	**16.00**	10.12	17.02	21.88	**16.34**
257	methyl trichloroacetate	**12.87**	11.82	12.37	15.02	**13.07**
258	methyl trifluoroacetate	**7.19**	6.07	7.00	9.31	**7.46**

259	hexamethyl phosphoramide	**16.05**	16.85	18.74	18.75	**18.11**
260	morpholine	**9.40**	8.03	9.24	10.52	**9.26**
261	morpholine-4-carbonitrile	**11.22**	10.06	11.92	12.33	**11.44**
262	n-butanoic acid	**8.80**	7.14	8.35	11.61	**9.03**
263	n-butyl acetate	**12.57**	11.10	12.05	15.10	**12.75**
264	n-butyl methyl ether	**10.80**	9.25	10.17	12.17	**10.53**
265	n-butylamine	**9.64**	7.97	8.38	11.27	**9.21**
266	n-decane	**19.33**	16.40	17.63	22.33	**18.79**
267	n-dodecane	**23.03**	19.57	20.88	26.76	**22.40**
268	n-dodecanol	**23.50**	18.42	19.75	34.15	**24.11**
269	n-heptane	**13.81**	11.19	11.99	16.97	**13.38**
270	n-hexadecane	**30.42**	26.09	28.44	36.39	**30.31**
271	n-octane	**15.60**	13.06	13.68	18.66	**15.13**
272	n-pentane	**10.11**	8.26	8.78	12.04	**9.69**
273	n-pentanoic acid	**10.60**	8.53	9.88	14.34	**10.92**
274	n-propyl acetate	**10.72**	9.46	10.25	12.98	**10.90**
275	n-propyl formate	**8.93**	8.09	8.52	10.43	**9.01**
276	n-undecane	**21.14**	18.25	19.04	24.26	**20.52**
277	nitrobenzene	**13.00**	7.88	13.84	16.16	**12.63**
278	nitrocyclohexane	**13.26**	11.41	13.09	13.34	**12.61**
279	nitroethane	**6.76**	5.40	6.46	7.59	**6.48**
280	nitromethane	**4.97**	3.93	4.85	5.31	**4.70**
281	o-xylene	**14.25**	9.34	15.74	17.38	**14.15**
282	p-chloroacetophenone	**16.50**	10.69	16.67	23.16	**16.84**
283	p-methoxybenzaldehyde	**15.90**	9.84	15.58	21.50	**15.64**
284	p-methylacetophenone	**16.40**	10.66	16.61	22.73	**16.67**
285	pentachloroethane	**14.17**	12.32	14.30	14.34	**13.65**
286	pentafluorobenzene	**10.43**	6.71	12.31	12.50	**10.51**
287	pentafluoropyridine	**10.18**	6.43	11.27	12.10	**9.93**
288	pentanenitrile	**9.99**	8.51	9.72	11.94	**10.06**
289	perfluoro(methylcyclohexane)	**13.71**	11.16	13.29	14.91	**13.12**
290	perfluoro-n-heptane	**14.70**	12.10	12.29	18.46	**14.29**
291	perfluoro-n-hexane	**12.71**	10.56	10.75	15.47	**12.26**
292	perfluoro-n-octane	**11.00**	7.04	12.42	13.71	**11.05**
293	perfluorodecalin	**13.91**	8.83	13.98	18.64	**13.82**
294	phenol	**15.80**	12.47	16.49	19.88	**16.28**
295	phenylacethylene	**13.96**	9.89	15.06	18.84	**14.60**
296	phenylacetone	**10.65**	8.83	10.46	11.06	**10.12**
297	propanenitrile	**6.30**	5.57	6.04	7.75	**6.45**
298	propanoic acid	**6.72**	5.74	6.90	8.85	**7.16**
299	propargylamine	**7.29**	5.99	6.48	8.69	**7.05**
300	propionaldehyde	**6.40**	5.42	6.48	7.47	**6.46**
301	propiophenone	**16.20**	10.67	16.67	22.40	**16.58**
302	pyridine	**9.58**	6.44	11.05	11.66	**9.72**

303	pyrimidine	**8.59**	6.17	10.39	10.93	**9.17**
304	pyrrole	**8.23**	5.58	9.40	9.54	**8.17**
305	pyrrolidin-2-one	**8.66**	7.02	9.33	9.96	**8.77**
306	pyrrolidine	**8.77**	7.15	8.97	9.15	**8.42**
307	pyrrolidine-1-carbonitrile	**11.09**	8.65	10.59	12.91	**10.72**
308	quinoline	**16.65**	9.72	17.41	21.71	**16.28**
309	styrene	**14.50**	8.97	14.68	18.65	**14.10**
310	sulfolane	**10.73**	9.67	10.97	11.83	**10.83**
311	tert-butyl alcohol	**8.82**	8.07	8.61	8.82	**8.50**
312	tert-amyl methyl ether	**12.48**	11.06	12.66	12.81	**12.17**
313	tert-butyl ethyl ether	**12.62**	11.39	11.52	14.04	**12.32**
314	tert-butyl methyl ether	**10.65**	9.80	10.06	11.41	**10.42**
315	tert-butylamine	**9.69**	8.76	9.10	9.24	**9.03**
316	tert-butyl hydroperoxide	**9.44**	8.73	9.20	10.02	**9.32**
317	tetraethylene glycol	**18.71**	15.57	19.29	23.91	**19.59**
318	tetrahydrofuran	**7.97**	6.93	7.95	8.73	**7.87**
319	tetrahydrothiophene	**10.41**	8.48	10.89	11.03	**10.13**
320	tetramethylene sulfoxide	**10.77**	9.15	10.68	11.11	**10.31**
321	tetrahydropyran	**9.85**	8.24	9.96	10.61	**9.61**
322	thiophene	**9.72**	6.54	10.75	10.80	**9.36**
323	toluene	**12.40**	8.08	13.23	15.47	**12.26**
324	tri-n-butyl phosphate	**27.55**	23.01	27.52	30.89	**27.14**
325	tri-n-butylamine	**24.43**	20.24	23.47	27.55	**23.75**
326	tri-n-propyl phosphate	**22.24**	18.10	22.21	24.97	**21.76**
327	trichloroacetonitrile	**10.48**	10.14	10.14	10.58	**10.29**
328	trichloroethylene	**10.08**	7.14	10.03	11.78	**9.65**
329	triethanolamine	**15.10**	13.38	15.35	16.84	**15.19**
330	triethylamine	**13.48**	10.96	13.36	14.49	**12.94**
331	trimethyl orthoacetate	**11.91**	10.06	12.42	13.87	**12.12**
332	trimethyl orthoformate	**10.09**	8.70	10.79	11.63	**10.37**
333	trimethyl phosphate	**11.07**	9.53	11.82	11.83	**11.06**
334	trimethylacetonitrile	**10.00**	9.59	9.59	10.56	**9.92**
335	trimethylene sulfide	**8.64**	7.20	8.86	9.13	**8.40**
336	vinyl acetate	**8.87**	6.67	8.65	12.19	**9.17**
337	2-methylphenol	**13.00**	8.38	14.69	15.97	**13.01**
338	methyloxirane	**6.25**	5.39	5.83	7.16	**6.13**
339	(trifluoromethyl)benzene	**12.25**	8.18	13.36	15.44	**12.33**
340	1,2-dibromobenzene	**16.43**	11.62	19.62	21.57	**17.60**
341	1,2-propanediol	**7.55**	6.97	7.45	8.11	**7.51**
342	1,3-difluorobenzene	**10.33**	6.70	12.10	12.28	**10.36**
343	1,5-cyclooctadiene	**14.16**	11.76	14.55	15.70	**14.00**
344	1-amino-2-propanol	**8.36**	7.51	7.71	8.95	**8.06**
345	1-bromopropane	**9.42**	7.81	8.25	11.73	**9.26**
346	1-chlorobutane	**10.12**	8.41	9.29	11.71	**9.81**

347	1-chloropropane	**8.32**	6.99	7.84	9.03	**7.95**
348	1-iodobutane	**13.33**	10.42	10.97	15.79	**12.39**
349	1-methylpyrrole	**10.24**	7.07	10.83	12.77	**10.22**
350	1-nonyne	**16.62**	14.07	16.20	19.67	**16.65**
351	2,2-dichloropropane	**10.32**	9.44	9.73	10.14	**9.77**
352	2,4-dimethyl-3-pentanol	**14.21**	12.42	13.26	15.82	**13.83**
353	2,4-dimethylphenol	**14.80**	9.75	16.33	19.14	**15.07**
354	2,5-dimethyltetrahydrofuran	**11.40**	9.75	11.21	13.78	**11.58**
355	2-cyanopyridine	**11.70**	7.87	12.53	15.87	**12.09**
356	2-fluoropyridine	**10.61**	6.43	11.04	11.77	**9.75**
357	2-methoxy-1,3-dioxolane	**9.36**	8.41	9.03	10.85	**9.43**
358	2-methyltetrahydrofuran	**9.95**	8.34	9.42	11.43	**9.73**
359	2-propanol	**6.98**	6.19	6.79	7.29	**6.76**
360	3-methylphenol	**13.00**	8.41	14.62	16.13	**13.05**
361	3-methylsulfolane	**12.65**	11.56	12.35	13.72	**12.54**
362	3-pentanol	**10.62**	9.05	9.53	12.65	**10.41**
363	3-phenyl-1-propanol	**16.44**	12.79	16.62	19.98	**16.46**
364	4-heptanone	**13.67**	11.90	13.16	15.48	**13.51**
365	4-methyl-2-pentanone	**11.98**	10.04	11.24	14.37	**11.88**
366	4-picoline	**11.52**	7.82	12.84	14.39	**11.68**
367	N,N-diethylacetamide	**13.28**	11.05	13.96	14.88	**13.30**
368	N,N-diethylcyanamide	**11.71**	9.40	11.92	13.25	**11.53**
369	N,N-diisopropylcyanamide	**15.31**	13.65	14.45	16.71	**14.94**
370	N,N-dimethylcyanamide	**7.94**	6.38	8.27	9.05	**7.90**
371	N,N-dimethylpropionamide	**11.46**	8.94	11.68	14.38	**11.66**
372	N-(tert-butyl)benzylamine	**21.11**	15.99	20.45	27.33	**21.26**
373	N-benzylmethylamine	**15.52**	10.86	15.47	21.09	**15.81**
374	benzoyl bromide	**15.70**	10.12	16.51	21.03	**15.89**
375	benzylamine	**13.60**	9.67	14.18	16.82	**13.56**
376	bis(2-chloroethyl)ether	**12.63**	10.25	11.14	17.96	**13.12**
377	bromoethane	**7.60**	6.42	6.89	8.88	**7.40**
378	chlorobenzene	**12.40**	8.11	13.29	15.91	**12.43**
379	chlorocyclohexane	**13.03**	11.35	12.43	13.32	**12.37**
380	chloroform	**8.53**	6.81	8.77	8.77	**8.12**
381	cis-decalin	**17.40**	13.59	17.09	20.11	**16.93**
382	cyclopentanone	**9.25**	7.66	9.73	10.16	**9.18**
383	bis(2-chloroethyl)ether	**12.60**	10.25	11.14	17.96	**13.12**
384	di-n-butyl sulfide	**18.74**	15.83	17.34	21.14	**18.10**
385	di-n-propyl ether	**12.67**	11.09	11.53	14.66	**12.43**
386	diallylamine	**12.84**	10.79	11.37	17.42	**13.19**
387	dibenzyl ether	**24.50**	17.16	26.06	33.76	**25.66**
388	dichloromethane	**6.52**	5.37	5.94	7.51	**6.27**
389	diethyl ether	**8.98**	7.51	8.03	11.15	**8.90**
390	Diethylene glycol diethyl ether	**17.67**	14.57	15.46	26.05	**18.69**

391	diisopropyl sulfide	**15.19**	12.44	13.71	17.74	**14.63**
392	dodecanenitrile	**22.68**	18.00	19.40	33.84	**23.75**
393	ethyl phenyl ether	**15.02**	9.99	15.19	20.41	**15.20**
394	ethyl salicylate	**17.80**	12.09	18.28	24.21	**18.19**
395	fluorobenzene	**10.33**	6.70	12.03	12.20	**10.31**
396	fluorotrichloromethane	**8.61**	6.82	8.72	8.72	**8.09**
397	glyceryl triacetate	**19.31**	17.18	20.97	24.18	**20.78**
398	hexachloropropene	**17.84**	14.00	17.99	19.47	**17.15**
399	iodobenzene	**15.58**	8.52	13.63	16.80	**12.98**
400	isopropyl nitrate	**9.61**	7.72	8.85	10.59	**9.05**
401	methanol	**3.26**	2.98	3.09	3.70	**3.25**
402	methyl decanoate	**21.73**	17.37	19.68	30.85	**22.63**
403	methyl dodecanoate	**25.40**	20.17	22.82	36.71	**26.57**
404	methyl phenyl ether	**13.11**	8.57	13.72	17.20	**13.16**
405	methyl phenyl sulfide	**15.63**	11.45	15.07	19.44	**15.32**
406	methylcyclohexane	**12.96**	10.34	12.45	14.11	**12.30**
407	n-heptanoic acid	**14.30**	11.66	13.28	18.93	**14.62**
408	n-hexane	**11.94**	9.79	10.62	14.02	**11.47**
409	n-hexanoic acid	**12.50**	10.14	11.48	16.58	**12.73**
410	n-nonane	**17.45**	15.04	15.24	20.49	**16.92**
411	n-pentadecane	**28.55**	23.16	26.95	35.67	**28.59**
412	p-xylene	**14.35**	9.44	14.40	19.11	**14.32**
413	Pentamethylene sulfide	**12.30**	9.90	12.81	12.85	**11.85**
414	piperidine-1-carbonitrile	**12.83**	9.94	12.48	15.03	**12.48**
415	tetrachloroethylene	**12.07**	8.28	13.07	13.09	**11.48**
416	2,2'-Thiobis(ethanol)	**12.50**	10.85	11.11	16.15	**12.70**
417	trans-1,2-dichloroethylene	**8.22**	5.99	7.15	10.48	**7.87**
418	triethyl phosphate	**16.57**	14.61	16.64	18.73	**16.66**
419	triethylene glycol	**14.36**	12.68	13.77	18.84	**15.09**
420	Triethylene glycol dimethyl ether	**18.34**	15.55	16.69	25.44	**19.22**
421	trifluoroacetic acid	**5.40**	4.45	5.90	5.92	**5.42**
422	undecanenitrile	**20.94**	16.63	17.98	30.74	**21.78**

Table S6. Three-body energy calculated by AMOEBA model with two sets of atomic polarizabilities for the 3B-69 dataset. The CCSD(T)/CBS data was taken from literature.[2]

Trimer cluster	CCSD(T)/CBS	AMOEBA (old polar)	AMOEBA (new polar)
01a-water	-1.39	-1.55	-1.53
01b-water	1.08	1.19	1.23
01c-water	-2.42	-2.87	-2.77
02a-formaldehyde	-0.18	-0.16	-0.19
02b-formaldehyde	0.18	0.17	0.20
02c-formaldehyde	-0.09	-0.04	-0.05
03a-methanol_ethyne	-1.31	-1.46	-1.39
03b-methanol_ethyne	0.05	0.03	0.03
03c-methanol_ethyne	0.02	-0.13	-0.13
04a-acetonitrile	0.25	0.17	0.19
04b-acetonitrile	0.34	0.26	0.29
04c-acetonitrile	-0.17	-0.15	-0.21
05a-nitromethane	-0.12	-0.18	-0.20
05b-nitromethane	0.26	0.17	0.18
05c-nitromethane	0.22	0.22	0.24
06a-acetic_acid	0.54	0.30	0.37
06b-acetic_acid	-0.92	-0.72	-0.84
06c-acetic_acid	-0.21	-0.14	-0.16
07a-oxalic_acid	-0.22	-0.16	-0.20
07b-oxalic_acid	-1.20	-0.86	-0.98
07c-oxalic_acid	-0.78	-0.53	-0.61
08a-vinylene_carbonate	-0.31	-0.35	-0.34
08b-vinylene_carbonate	-0.15	-0.23	-0.22
08c-vinylene_carbonate	-0.49	-0.61	-0.60
09a-acetamide	-0.09	-0.09	-0.11
09b-acetamide	0.58	0.80	0.75
09c-acetamide	-0.86	-0.65	-0.70
10a-imidazole	-0.66	-0.82	-0.93
10b-imidazole	0.27	0.21	0.21
10c-imidazole	-1.64	-1.87	-2.18
11a-isoxazole	0.18	-0.08	-0.07
11b-isoxazole	0.19	-0.13	-0.04
11c-isoxazole	-0.13	-0.05	-0.06
12a-pyrazole	-1.28	-1.72	-1.79
12b-pyrazole	0.07	-0.03	-0.02
12c-pyrazole	-0.29	-0.40	-0.45
13a-triazine	-0.01	0.00	0.00
13b-triazine	0.14	0.11	0.05
13c-triazine	0.13	0.02	0.03
14a-cyanoacetamide	1.46	0.03	0.11
14b-cyanoacetamide	0.35	-0.43	-0.48

14c-cyanoacetamide	-1.23	-0.81	-1.03
15a-cyanoguanidine	-0.91	-0.10	-0.50
15b-cyanoguanidine	0.58	0.33	0.42
15c-cyanoguanidine	0.14	0.19	0.19
16a-triazolidinedione	-0.14	-0.07	-0.07
16b-triazolidinedione	0.78	1.05	1.01
16c-triazolidinedione	0.51	0.29	0.28
17a-oxazolidinone	0.54	0.46	0.47
17b-oxazolidinone	-0.21	-0.36	-0.42
17c-oxazolidinone	0.15	0.12	0.13
18a-succinic_anhydride	0.43	0.20	0.23
18b-succinic_anhydride	0.00	0.01	0.00
18c-succinic_anhydride	-0.13	-0.16	-0.19
19a-benzene	0.05	-0.08	-0.06
19b-benzene	0.15	0.08	0.07
19c-benzene	-0.03	-0.08	-0.08
20a-maleic_acid	1.69	1.11	1.27
20b-maleic_acid	-1.45	-0.92	-1.19
20c-maleic_acid	-0.36	-0.19	-0.23
21a-p-benzoquinone	0.33	0.18	0.25
21b-p-benzoquinone	0.00	-0.04	-0.09
21c-p-benzoquinone	0.10	0.04	0.05
22a-uracil	0.00	0.56	0.41
22b-uracil	-0.24	-0.12	-0.13
22c-uracil	0.02	-0.01	0.01
23a-cyclobutylfuran	0.08	-0.05	-0.05
23b-cyclobutylfuran	0.14	0.06	0.06
23c-cyclobutylfuran	0.15	0.03	0.03
RMSE (kcal/mol)	**0**	**0.30**	**0.26**

Table S7. Atomic polarizability parameters (in $Å^3$) for charged molecules. Parameters are derived by fitting polarizability only (*fit-polar*) and fitting electrostatic difference only (*fit-esp*). The final set of parameters are the averages of *fit-polar* and *fit-esp*. The atom types for charged molecules are illustrated in **Figure S1**.

Index	**Atom type**	***fit-polar***	***fit-esp***	***final***
1	C(S-)	2.46896	3.08732	2.77814
2	C+	1.51746	1.54505	1.53125
3	Car-	2.71847	2.20922	2.46385
4	Cn+	0.82718	0.78690	0.80704
5	Co-	2.82278	2.85413	2.83845
6	Coo-	1.86150	2.07823	1.96987
7	Cooo-	2.43886	1.95324	2.19605
8	Hn+	0.25500	0.26853	0.26176
9	Hop	0.36037	0.42908	0.39472
10	NX4+	0.74624	0.80870	0.77747
11	Nar-	1.94962	2.02092	1.98527
12	Nc+	0.87828	0.85571	0.86700
13	O(C)P	1.18271	1.23862	1.21067
14	O(H)P	0.71154	0.87051	0.79102
15	O-	3.28997	3.03702	3.16350
16	Oc-	1.92556	1.93713	1.93135
17	Oc-2	2.79251	3.07775	2.93513
18	Oo-	2.22656	2.02176	2.12416
19	Op-	2.41788	2.28682	2.35235
20	Os-	1.64427	1.51598	1.58013
21	Po4	1.43530	1.42792	1.43161
22	S-	7.36477	5.66354	6.51416
23	So3-	2.50520	2.43678	2.47099

Table S8. Performance of the *final* parameters on the ESP and molecular polarizability sets for charged molecules. RMSE (in kcal/mol/e and Å^3 for ESP and polarizability respectively) and unsigned mean percentage error (UMPE, defined in **Eq.3** in the main text) are reported.

Parameters	**ESP set** (128 probing structures)		**Polarizability set** [a,b] (70 molecules)	
fit-polar	RMSE	0.0563	RMSE (α_{ii})	0.68
	UMPE		UMPE (α_{ii})	4.2%
			RMSE (α_{iso})	0.47
			UMPE (α_{iso})	3.0%
fit-esp	RMSE	0.0448	RMSE (α_{ii})	0.80
	UMPE		UMPE (α_{ii})	4.6%
			RMSE (α_{iso})	0.65
			UMPE (α_{iso})	3.7%
final	RMSE	0.0488	RMSE (α_{ii})	0.71
	UMPE		UMPE (α_{ii})	4.2%
			RMSE (α_{iso})	0.52
			UMPE (α_{iso})	3.1%

a. α_{ii}: statistics based on three eigenvalues, where $i = x, y, z$.
b. α_{iso}: statistics based on the isotropic polarizabilities, i.e., $\frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$.

References

1. Wilkins, D. M.; Grisafi, A.; Yang, Y.; Lao, K. U.; DiStasio, R. A.; Ceriotti, M., Accurate molecular polarizabilities with coupled cluster theory and machine learning. *Proceedings of the National Academy of Sciences* **2019,** *116* 3401-3406.

2. Řezáč, J.; Huang, Y.; Hobza, P.; Beran, G. J. O., Benchmark Calculations of Three-Body Intermolecular Interactions and the Performance of Low-Cost Electronic Structure Methods. *J. Chem. Theory Comput.* **2015,** *11* 3065-3079.