JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

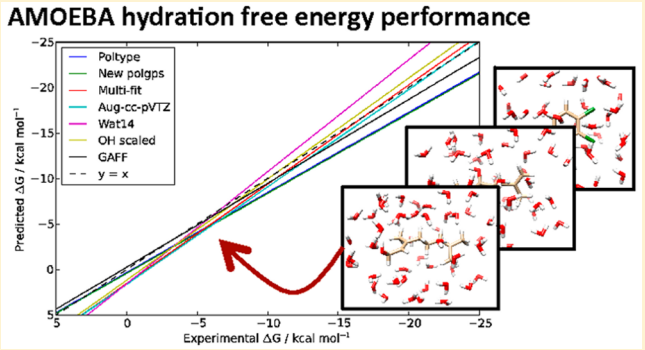
Evaluating Parametrization Protocols for Hydration Free Energy Calculations with the AMOEBA Polarizable Force Field

Richard T. Bradshaw and Jonathan W. Essex*

School of Chemistry, University of Southampton, Highfield Campus, Southampton SO17 1BJ, U.K.

Supporting Information

ABSTRACT: Hydration free energy (HFE) calculations are often used to assess the performance of biomolecular force fields and the quality of assigned parameters. The AMOEBA polarizable force field moves beyond traditional pairwise additive models of electrostatics and may be expected to improve upon predictions of thermodynamic quantities such as HFEs over and above fixed-point-charge models. The recent SAMPL4 challenge evaluated the AMOEBA polarizable force field in this regard but showed substantially worse results than those using the fixed-point-charge GAFF model. Starting with a set of automatically generated AMOEBA parameters for the SAMPL4 data set, we evaluate the cumulative effects of a series of incremental improvements in parametrization protocol, including both solute and solvent model changes. Ultimately, the optimized AMOEBA parameters give a set of results that are not statistically significantly different from those of GAFF in terms of signed and unsigned error metrics. This allows us to propose a number of guidelines for new molecule parameter derivation with AMOEBA, which we expect to have benefits for a range of biomolecular simulation applications such as protein−ligand binding studies.



INTRODUCTION

Hydration free energy (HFE) calculations have long been used as a tool for validating the accuracy of molecular mechanical force fields and comparing performance between different theoretical approaches.[1−4] The widespread use of hydration free energies in force field development can be primarily rationalized with two main arguments: first, the speed of equilibration and sampling for small water-solute systems allows HFEs to be rapidly estimated with high precision and used as fair comparisons between methodologies, and second, accurate HFE predictions have a direct relevance to the correct prediction of other experimental metrics such as protein−ligand binding free energies.[2,5,6]

Despite the utility of HFE data for computational method development, the majority of solutes for which experimental data exist are small, fragment-like compounds that may have little relevance to the larger drug-like molecules of interest in many biomolecular simulations.[7−9] Validating force field transferability with HFE estimates ideally requires a broad evaluation of performance across a range of chemical functionalities and sizes. Since 2008, the community-wide statistical assessment of modeling of proteins and ligands (SAMPL) challenges have included a HFE blind challenge set incorporating rare or previously unpublished experimental data for an ever increasing variety of chemical compounds.[2,10−13] The 2013 challenge, SAMPL4, saw 49 submissions from 19 groups worldwide.[14] Although current and future challenges may move beyond simple free energy calculations toward derived properties such as distribution coefficients,[15] large data sets of experimental and computational HFE results from SAMPL and other sources remain a valuable resource for any researchers' future force field validation attempts.[9]

A significant challenge facing any molecular mechanical force field is in ensuring its transferability to systems beyond those for which it has been explicitly parametrized. There remains a great deal of interest in the development and application of polarizable force fields in this regard.[16−20] By incorporating an explicit representation of polarization into the underlying energy function, polarizable force fields are able to inherently respond to environmental changes dynamically during simulations, in addition to better capturing subtleties in molecular interactions. This response to environment changes should enhance the transferability and accuracy of such force fields and reduce the need for empirical reparametrization compared to traditional pairwise additive models. HFE calculations have repeatedly been used as a challenge in the development of a variety of polarizable force fields, often showing good agreement with experiment for small organic molecules.[21−23]

Owing to their relative novelty, advanced polarizable models have not been widely adopted despite showing considerable improvements over fixed-point-charge models in many areas.[24−30] The AMOEBA polarizable force field has seen

Received: March 16, 2016
Published: June 24, 2016

DOI: 10.1021/acs.jctc.6b00276
J. Chem. Theory Comput. 2016, 12, 3871−3883

extensive development in recent years, including broader parametrization,[31,32] algorithmic improvements,[33−35] and extension to new hardware.[36,37] It has also seen efforts to broaden its applicability to new systems with automated parametrization protocols.[38] However, while AMOEBA performance during the SAMPL4 HFE challenge was fair,[14,39] it was substantially outperformed by the fixed-charge GAFF/AM1-BCC model. Similar AMOEBA performance was seen in the SAMPL4 host−guest binding challenge.[40]

The SAMPL challenges are performed blind, with submitting groups having no a priori knowledge of the "correct" experimental results. As such, no effort is made to optimize the agreement of any particular method with experiment during the period of the challenge, for example, through reparametrization. This likely benefits the more established, mature and broadly tested methodologies, as suggested by the extremely successful performance of the GAFF model.[14] In this study, we wish to re-evaluate AMOEBA performance on the SAMPL4 data set, using the findings of previous HFE studies to carry out a series of incremental improvements to parametrization. This allows us to identify ways of optimizing a set of small molecule AMOEBA parameters within the confines of the existing AMOEBA potential, providing useful guidelines for the setup of, for example, protein−ligand simulations.

AMOEBA Potential. The AMOEBA potential[21,31,41,42] primarily differs from that of GAFF and other pairwise additive models in its treatment of electrostatic interactions. The AMOEBA potential for electrostatic interactions consists of two parts, corresponding to permanent and induced interactions:

$$U_{ele} = U_{ele}^{perm} + U_{ele}^{ind} \quad (1)$$

The first term, U_{ele}^{perm} incorporates fixed atomic multipole (up to quadrupole) interactions with an exclusion list applied to atoms that share a bond or angle and scaling applied to 1−4 and 1−5 interactions. Induction (U_{ele}^{ind}) is included via a Thole damped induced dipole model,[43] whereby at each polarizable site the electric field created by permanent multipoles and induced dipoles induces a dipole, which itself then further polarizes all other sites. Induced interaction scaling between bonded atoms is performed as per permanent interactions, and AMOEBA additionally makes use of "polarization groups" for screening induced interactions.[41] The permanent atomic multipoles of atoms within the same polarization group do not mutually polarize one another, but interactions between induced point dipoles are evaluated between all atoms. Polarization groups are loosely defined as subgroups of atoms with some element of conformational rigidity. Within a group, there is therefore an element of conformational independence in the contributions of group permanent multipoles to induction, meaning that these intragroup contributions may be ignored. Intragroup polarization is therefore taken into account statically in the magnitudes of the permanent atomic multipoles themselves. This often, but not exclusively, means that polarization groups correspond to chemical functional groups.

The derivation of electrostatic parameters for new molecules in the AMOEBA force field is therefore more complex than the analogous process for GAFF, which can either use a charge fitting scheme based on a semiempirical quantum mechanical calculation with bond-charge corrections (AM1-BCC)[44] or a restrained electrostatic potential (RESP) fit.[45] Indeed, the GAFF parametrization process has long been fully automated via the *antechamber* software, used in the SAMPL4 submission.[14,46,47] During their SAMPL4 submission using the AMOEBA force field,[39] Manzoni and Söderhjelm made use of the automated *Poltype* software of Ren and co-workers to generate AMOEBA parameters.[38] However, in the case of anthraquinone molecules problems were identified with dihedral parameters assigned by *Poltype* that led to incorrect, nonplanar, geometries. As a result, we use a revised set of *Poltype*-generated parameters, with corrections to some valence term assignments, as a starting point for a stepwise series of further modifications to small molecule parametrization.

METHODS

Data Set. The SAMPL4 HFE data set consists of 47 neutral organic compounds, hereafter referred to as solutes. The structures of all compounds are shown in Table S2. The full data set may be decomposed into five subsets according to solute type: (1) linear or branched functionalized alkanes and alkenes (n = 10), (2) substituted benzenes and derivatives (n = 13), (3) cycloalkanes/enes and derivatives (n = 10), (4) anthracenes, polyaromatics, and derivatives (n = 8), and (5) polyfunctional compounds and others (n = 6). Solutes assigned to each group can be found in Table S3.

Parametrization Protocols. A summary of a standard protocol for adding small molecule parameters to AMOEBA, as used by the *Poltype* software, is depicted in Figure 1. Briefly,

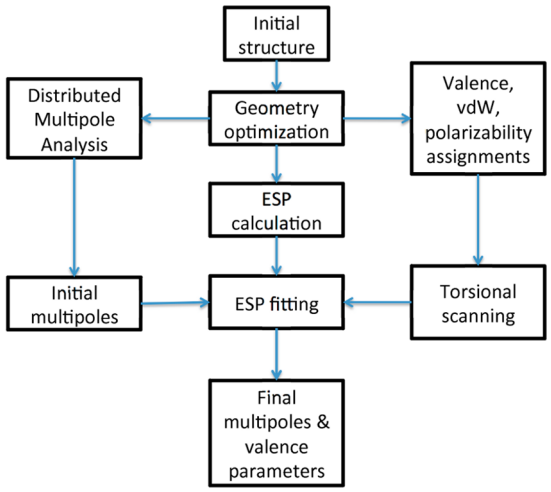


Figure 1. General AMOEBA parametrization protocol. An initial solute conformer is geometry optimized and subject to distributed multipole analysis to extract a set of atomic multipoles. Multipole parameters are refined by fitting to the molecular electrostatic potential of single or multiple solute conformations. Valence, vdW, and polarizability parameters are assigned based on atom type, while other parameters (e.g., polarization groups) may be assigned either automatically or manually by inspection.

multipole parameters are derived from a geometry-optimized QM structure via a distributed multipole analysis (DMA)[48,49] and then further fitted to recreate the molecular electrostatic potential (ESP) calculated with a larger QM basis set. van der Waals (vdW) and valence parameters are assigned based on similarity to existing atom types. Bond and angle geometries are extracted from the geometry-optimized structure used for multipole parameter generation above. Dihedral parameters are first assigned based on atom type similarity within a database of existing torsions, followed by fitting of an up-to 3-term Fourier

DOI: 10.1021/acs.jctc.6b00276
J. Chem. Theory Comput. 2016, 12, 3871−3883

Table 1. Summary of Simulation Protocols for Polarizable and Fixed-Charge Potentials

set	name	modified dihedrals	modified polarization groups	multipole conformational fit	QM ESP basis set	water model	empirical multipole scaling
1	*Poltype*	yes	no	single	6-311++G(2d,2p)	wat03	*Poltype*[a]
2	new polgps	yes	yes	single	6-311++G(2d,2p)	wat03	*Poltype*[a]
3	multifit	yes	yes	multiple	6-311++G(2d,2p)	wat03	none
4	Aug-cc-pvtz	yes	yes	multiple	aug-cc-pVTZ	wat03	none
5	Wat14	yes	yes	multiple	aug-cc-pVTZ	wat14	none
6	OH scaled	yes	yes	multiple	aug-cc-pVTZ	wat14	aliphatic OH
7	GAFF	N/A	N/A	N/A	N/A	TIP3P	N/A

[a]*Poltype* performed automatic scaling of hydroxyl (OH) and amine (NH_n) quadrupole values of functional groups beyond simply the aliphatic OH groups suggested by Shi et al. (see Text S1). The effects of this additional empirical scaling are not directly evaluated here.

series to a QM-derived torsional scanning energy profile if no suitable existing parameters are found.

Modifications to the above general protocol were split into two main aims. First, we investigated the potential for further improvement of parameters generated by *Poltype* with two additions to the standard protocol: (1) Improvement of *Poltype*-assigned dihedral parameters, as noted by Manzoni and Söderhjelm, the chemical space currently covered by *Poltype* does not include some of the functionalities in the SAMPL4 solute set.[39] (2) Improvement of *Poltype*-assigned polarization groups, polarization groups assigned by *Poltype* were adapted to better take account of intramolecular interactions (e.g., H-bonding) for a subset of solutes. Second, as *Poltype* is limited to the derivation of multipoles from a single solute conformation, we carried out four subsequent protocols making use of a manual reparametrization of solute electrostatic parameters: (3) Use of multiple solute conformations during the fitting of multipole parameters to the QM electrostatic potential, up to six diverse, low energy conformations were used for each solute in a simultaneous multiconformational fit of permanent dipole and quadrupole values to molecular ESP. (4) Use of a larger basis set during the fitting of multipole parameters to the QM electrostatic potential—single point solute ESP calculations were performed at the MP2/aug-cc-pVTZ level of theory, as suggested by previous studies.[21,50] (5) Use of an updated AMOEBA water model—solutes were solvated in a box of water using the AMOEBA 2014 water model (wat14),[32] rather than the AMOEBA 2003 model (wat03)[42,51] for condensed phase simulations. (6) Use of empirically scaled multipole parameters for aliphatic alcohol groups—as suggested by Shi and co-workers, quadrupole parameters for hydroxyl groups bound to sp^3 carbons were scaled to 60% of their original value.[50]

A seventh set of simulations using the GAFF fixed-charge force field for solute parameters was also performed for comparison purposes. A summary of the conditions for all seven protocols is given in Table 1. For the sake of brevity, precise details of all parametrization methods are provided in Text S1.

Simulation Protocol. Alchemical hydration free energy calculations were split into three legs according to the thermodynamic cycle of Figure 2. The overall HFE for each solute was then calculated according to eq 2:

$$\Delta G_{hyd} = \Delta G_{discharging,\,vac} - \Delta G_{decoupling,\,sol} - \Delta G_{discharging,\,sol} \quad (2)$$

where $\Delta G_{discharging,vac}$ is the free energy change in switching off electrostatic interactions in vacuum, $\Delta G_{discharging,sol}$ is the equivalent free energy change in solution (including both intramolecular and intermolecular interactions), and $\Delta G_{decoupling,sol}$ is the free energy change in removing vdW interactions between the solute and its environment. A vdW decoupling step is not required in vacuum as there are no interactions between the solute and its environment.

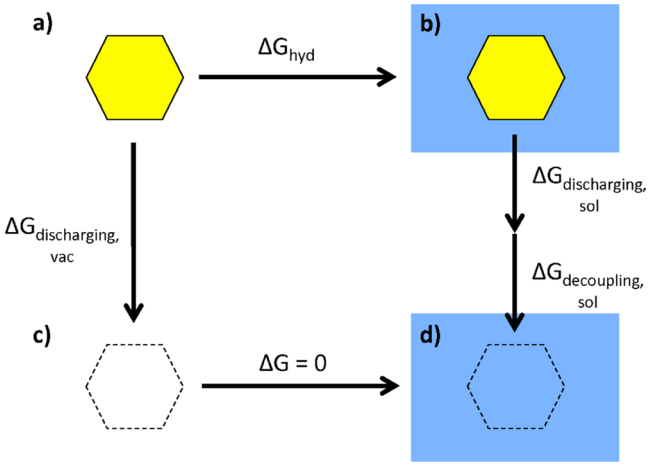


Figure 2. Thermodynamic cycle used in performing alchemical free energy calculations. The cycle has four defined end points: (a) a fully interacting solute in vacuum, (b) a fully interacting solute in water, (c) a discharged and vdW decoupled solute in vacuum, and (d) a discharged and vdW decoupled solute in water. Simulations were separated into three legs: alchemical discharging and vdW decoupling steps in solution, and a discharging step in vacuum. The overall ΔG_{hyd} was calculated as $\Delta G_{discharging,vac} - \Delta G_{decoupling,sol} - \Delta G_{discharging,sol}$.

For simulations with the AMOEBA force field (sets 1−6 in Table 1), vacuum simulations were performed with the *dynamic* program of TINKER 6.3.3[52] and condensed phase simulations with the *pmemd.amoeba* program of Amber 14.[53] For condensed phase simulations, ligands were initially placed in the center of a pre-equilibrated cubic box of 671 AMOEBA water molecules with 30 Å sides. Systems were energy minimized for 1000 steps with a steepest descent algorithm and 1500 steps with a conjugate gradient algorithm. The system was then heated to 300 K over a period of 50 ps in the NVT ensemble, and equilibrated to 1 atm pressure for a further 100 ps in the NPT ensemble. The final structure of the equilibrated system was used as the input for all subsequent simulations at intermediate λ windows. Discharging simulations were performed in 11 steps with equally spaced λ windows between $\lambda = 1.0$ (full atomic multipoles and polarizabilities) and $\lambda = 0.0$ (zero atomic multipoles and polarizabilities). Decoupling of vdW interactions was performed in 11 steps with $\lambda = (1.0, 0.9, 0.8, 0.75, 0.7, 0.65, 0.6, 0.5, 0.4, 0.2, 0.0)$ and using a softcore vdW potential.[50] This nonlinear scaling has previously been

DOI: 10.1021/acs.jctc.6b00276
J. Chem. Theory Comput. 2016, 12, 3871−3883

shown to be appropriate for HFE simulations of this type.[50] Production simulations were then performed at each λ window for a total of 2 ns in the NPT ensemble. The first 200 ps of each trajectory were discarded as equilibration, and the remaining 1.8 ns used for analysis.

A Langevin thermostat[54−56] and Berendsen barostat[57] (excluding constant volume simulations) were used to maintain temperature at 300 K and pressure at 1 atm, respectively, in all simulations. A velocity Verlet integrator with 1 fs time step was used throughout. vdW interactions were subject to a 9 Å cutoff with an analytical long-range correction. Long-range electrostatic interactions were treated with a Particle Mesh Ewald summation with an 8 Å real-space cutoff.[58] During the propagation of dynamics, induced dipole convergence was set to 10^{-2} D/atom. Simulation snapshots were then reanalyzed using a tighter convergence criterion of 10^{-5} D/atom during free energy calculations.[50] All simulations were performed in triplicate. Repeat simulations used different random seeds for the Langevin thermostat to ensure independence of trajectories.[59]

Gas phase discharging simulations used 11 equally spaced λ windows between λ = 1.0 and λ = 0.0. The QM geometry optimized solute structure was used as the input for all windows. Production simulations were then performed for a total of 200 ps using a stochastic integrator with a 0.1 fs time step and temperature maintained at 300 K. All nonbonded interactions were explicitly evaluated through the use of an effectively infinite cutoff. Induced dipoles were converged to 10^{-6} D/atom during both simulations and free energy analyses. Simulations were performed in triplicate, and the first 20 ps of each simulation was discarded as equilibration.

For simulations with the GAFF force field (set 7, Table 1), gas phase simulations were performed with the *sander* program and condensed phase with the *pmemd* program, both of Amber 14. Solvated systems were first prepared by soaking the solute in a box of TIP3P[60] water molecules such that no solute atom was less than 10.2 Å from the edge of the box, resulting in similar box edges to those used in the AMOEBA simulations (30 Å). Systems were energy minimized for 1000 steps with a steepest descent algorithm and 1500 steps with a conjugate gradient algorithm, heated to 300 K over 50 ps under NVT conditions and equilibrated to 1 atm over 100 ps under NPT conditions. The final equilibrated structure of each system was used as input for production simulations at all λ windows. Identical λ spacing was used for GAFF simulations as for AMOEBA simulations. Production simulations were run for 2 ns in the NPT ensemble at each window, and the first 200 ps of each simulation discarded as equilibration.

All simulations again used a Langevin thermostat and Berendsen barostat (excluding constant volume simulations) to maintain temperature at 300 K and pressure at 1 atm. A 1 fs time step was used throughout, and the SHAKE algorithm used to constrain all bonds involving hydrogen.[61] vdW interactions were subject to a 8 Å cutoff with an analytical long-range correction and long-range electrostatic interactions were treated with a Particle Mesh Ewald summation with an 8 Å real-space cutoff.[62] Simulations were performed in triplicate.

Gas phase simulations with GAFF used identical λ spacing to those with AMOEBA, and the geometry optimized output structure from *antechamber* parametrization was used as the input for all production simulations. Production simulations were run for 2 ns at each window using a velocity Verlet integrator with 1 fs time step and Langevin thermostat to maintain temperature at 300 K. All nonbonded interactions were again explicitly evaluated. Simulations were performed in triplicate and the first 200 ps of each trajectory discarded as equilibration.

Statistical Analysis. Free energy differences between adjacent λ windows were calculated with the BAR methodology,[63] using an in-house script for AMOEBA outputs and the *pymbar* package[64] for GAFF outputs. For each solute, the mean and standard deviation in HFE was taken across the three repeat simulations. For each parameter set, the experimental and predicted HFEs were compared for all 47 solutes. The mean signed error (MSE), mean unsigned error (MUE), coefficient of determination (R^2), and Kendall's tau coefficient (τ) were calculated for all parameter sets. 95% confidence intervals (CI) for each metric were estimated via bootstrapping of the underlying data set with replacement. A total of 10 000 resamples were performed for each metric.

As recognized during the SAMPL4 challenge, testing for statistical significance between HFE results with different but related methodologies is nontrivial.[14] Rather than test the distributions of predicted HFEs directly, we instead tested for significance between the distributions of predicted vs experimental values (i.e., the signed or unsigned errors) for the sets being compared. We performed a Student's paired t test for tests between signed errors. Distributions of unsigned errors are by definition non-normal as negative values are transformed into their absolute positive equivalents; the assumption of normality in the Student's t test may therefore make it unsuitable. We therefore additionally performed the nonparametric Wilcoxon signed rank test to assess significance between sets of unsigned errors.

Finally, differences in free energy estimates for individual solutes were also tested for significance. For each solute, a one-way ANOVA of the three predicted HFE values across all AMOEBA parameter sets was evaluated, testing whether at least one of the parameter sets gave significantly different HFE results to the others. Finally, a Tukey honest significant difference (HSD) test was performed to evaluate pairwise differences between all parameter sets for each solute. For all statistical tests above, a difference was deemed to be significant if $p < 0.05$.

RESULTS AND DISCUSSION

Improvement of *Poltype* Parameters. The performance of the AMOEBA force field in the SAMPL4 HFE challenge indicated that improvements were likely possible in the parametrization process. As a reference point for judging the effects of these improvements, parameter set 1 first evaluated solute parameters generated by *Poltype*, with minor changes to valence parameters to remove clear errors for some members of the SAMPL4 data set. In particular, these changes ensured correct geometries of the anthraquinones, aldehydes, ketones in rings, and nitro groups; full details are given in Text S1. Comparison of predicted and experimental HFEs is shown in Figure 3, and overall results in comparison with other parameter sets are shown in Table 2.

The set 1 MUE of 1.85 kcal mol^{-1} compares favorably with the MUE of 2.34 kcal mol^{-1} reported for the same data set and similar parametrization protocol during the SAMPL4 challenge.[39] MSE values, however, remained similar (1.03 kcal mol^{-1} and 1.10 kcal mol^{-1}, respectively) suggesting that while random errors in the results have been reduced in this study, the AMOEBA model still systematically underestimates HFE.

DOI: 10.1021/acs.jctc.6b00276
J. Chem. Theory Comput. 2016, 12, 3871−3883

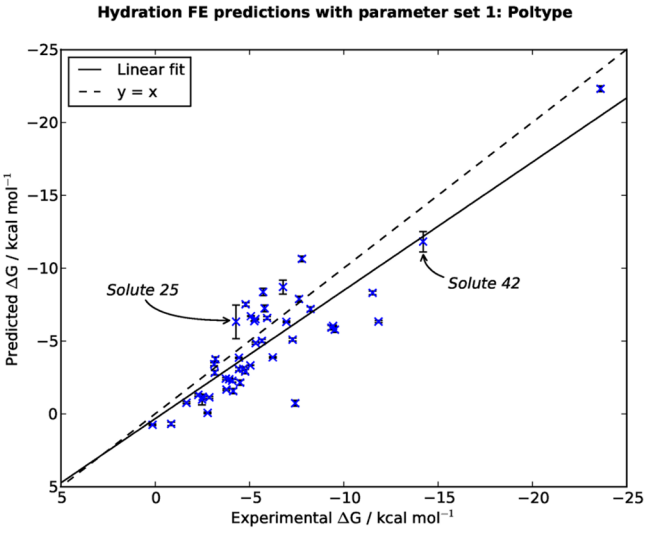


Figure 3. Plot of predicted against experimental HFE for parameter set 1. AMOEBA solute parameters were derived using *Poltype*[38] with modifications outlined in the Supporting Information. Error bars are modeled as standard deviations across three repeat simulations. Solutes 25 and 42 (labeled) had large associated uncertainties due to sampling differences between runs.

Two molecules in Figure 3 (solutes 25 and 42) have clearly larger error bars than others, suggesting sampling differences across the three repeat simulations. In both cases, visual inspection of the solution-phase trajectories identified some conformational variability. Solute 25 (1,3-bis(nitrooxy)butane) had multiple potential conformers resulting from rotations of the two nitrooxy groups and the branched butane chain. Solute 42 (1-(2-hydroxyethylamino)-9,10-anthraquinone) showed differences in the populations of an intramolecular hydrogen bond between the anthraquinone carbonyl and aminoethanol substituent, along with multiple potential conformations of the aliphatic chain of the substituent. However, we do not expect these observed sampling differences to negatively impact the accuracy of HFE estimates; in fact, they serve to highlight the necessity of performing multiple independent simulations when evaluating free energy calculations.

Discrepancies between our results and those submitted during SAMPL4 are likely to be caused by minor differences in the simulation protocol, namely, the slightly longer simulation lengths used here, the additional parameters for aldehydes, ketones, and nitro groups (Manzoni and Söderhjelm also added high barriers to dihedrals in anthraquinone rings), and our use of 15° intervals for torsional scanning rather than the 30° intervals used previously. Nevertheless, the AMOEBA results with parameter set 1 remain broadly comparable with those reported previously and provide a starting point for further improvements to parametrization.

Within the AMOEBA force field, intramolecular polarization is by definition sensitive to the choice of polarization group. Upon inspection of set 1 parameters, it was noted that definition of these groups was overly conservative in the case of substituted benzene solutes. Figure 4 illustrates an example of

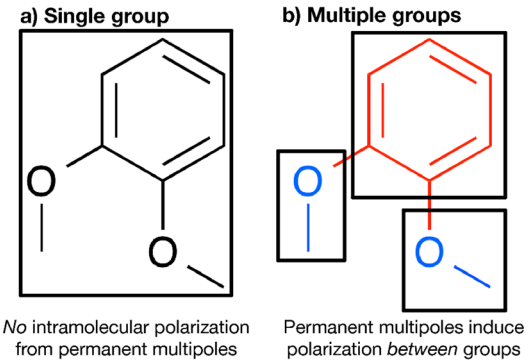


Figure 4. Example of possible polarization group definitions for solute L05 (1,2-dimethoxybenzene). A careful choice of polarization groups (right) allows a more realistic response of induced polarization to changes in molecular conformation, e.g., rotation of the two methoxy substituents.

the polarization groups chosen by *Poltype* and an alternative manually chosen arrangement consistent with the AMOEBA protein force fields. This shows the effect that choice of polarization group can have on intramolecular interactions. The single polarization group assigned in Figure 4a was typical of the groups automatically assigned by *Poltype* to the substituted benzene solutes. Therefore, set 2 parameters evaluated the effects of manually assigning polarization groups for the substituted benzenes (groups detailed in Table S2).

As the majority of solutes were not reparameterized in set 2, the metrics for the agreement of the full data set with experiment do not differ greatly from those of set 1 (1.84 kcal mol^{-1} and 1.13 kcal mol^{-1} for set 2 MUE and MSE, respectively, compared to 1.85 kcal mol^{-1} and 1.03 kcal mol^{-1}, respectively, for set 1). However, results for the subgroup of substituted benzene solutes alone are compared in Figure 5 and Table 3. Despite the wide confidence interval estimates arising from the small data set size (n = 13) used for bootstrapping, the use of manually assigned polarization groups resulted in a small improvement in MUE and a small deterioration in MSE, but neither of these differences are statistically significant on this small subset of 13 solutes (MSE, Student's paired t test, p = 0.34, MUE, Wilcoxon signed rank test, p = 0.75). From Figure 5, it is clear that the effects of new

Table 2. Comparison of Result Metrics for All Parameter Sets[a]

set	name	MSE (kcal mol^{-1})	MUE (kcal mol^{-1})	R^2	Kendall τ
1	*Poltype*	0.484 ≤ **1.032** ≤ 1.616	1.553 ≤ **1.850** ≤ 2.289	0.50 ≤ **0.75** ≤ 0.93	0.43 ≤ **0.62** ≤ 0.73
2	new polgps	0.590 ≤ **1.125** ≤ 1.687	1.538 ≤ **1.838** ≤ 2.284	0.51 ≤ **0.77** ≤ 0.93	0.46 ≤ **0.63** ≤ 0.75
3	multifit	0.375 ≤ **1.007** ≤ 1.605	1.391 ≤ **1.765** ≤ 2.322	0.56 ≤ **0.80** ≤ 0.95	0.49 ≤ **0.67** ≤ 0.77
4	Aug-cc-pvtz	0.634 ≤ **1.218** ≤ 1.761	1.410 ≤ **1.774** ≤ 2.265	0.59 ≤ **0.82** ≤ 0.96	0.51 ≤ **0.69** ≤ 0.78
5	Wat14	−0.489 ≤ **0.226** ≤ 0.828	1.262 ≤ **1.616** ≤ 2.232	0.62 ≤ **0.84** ≤ 0.96	0.46 ≤ **0.65** ≤ 0.75
6	OH scaled	0.000 ≤ **0.530** ≤ 1.071	1.240 ≤ **1.529** ≤ 1.955	0.63 ≤ **0.84** ≤ 0.96	0.47 ≤ **0.66** ≤ 0.76
7	GAFF	−0.184 ≤ **0.202** ≤ 0.674	0.873 ≤ **1.100** ≤ 1.470	0.74 ≤ **0.86** ≤ 0.90	0.58 ≤ **0.73** ≤ 0.84

[a]Upper and lower bounds for each metric are 95% CI estimates.

DOI: 10.1021/acs.jctc.6b00276
J. Chem. Theory Comput. 2016, 12, 3871−3883

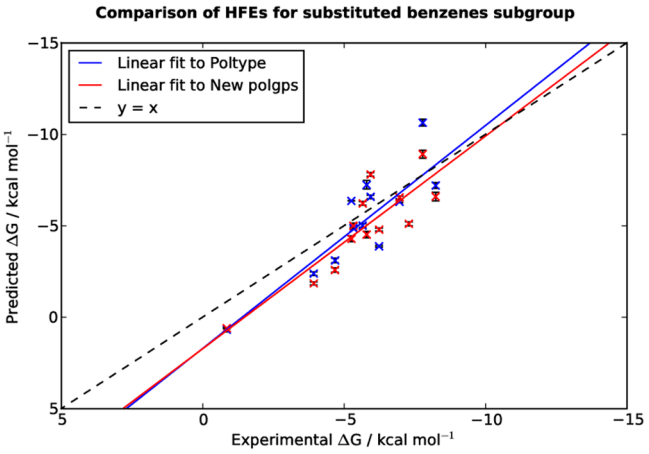


Figure 5. Comparison of parameter sets 1 (blue) and 2 (red) for the subset of substituted benzene solutes. Error bars modeled as standard deviations across three repeat simulations. In set 2, new polarization groups were assigned manually where necessary for each molecule in the subset. Precise polarization groups are detailed in Figure S2.

polarization groups have varying magnitudes for individual solutes across the data set. However, as demonstrated in Figure 4, a poor automatic choice of polarization group may affect the fundamental character of intra- and intermolecular interactions, and we expect such systematic errors would become clearer in larger data sets.

The use of polarization groups in the AMOEBA force field avoids the potential for errors arising from the large direct induction effects between atomic multipoles on bonded atoms, while maintaining accuracy in recreating the molecular electrostatic potential, as detailed by Ren and Ponder.[41] As stated above, polarization groups are loosely defined as groups with limited conformational flexibility, such that the neglected intragroup induction effects should be small and conformation independent.

The automatic definition of polarization groups for any new molecule is highly desirable from the point of view of developing automated parametrization protocols such as *Poltype*. However, the difficulty in creating rules to do so is highlighted by this study. Within *Poltype*, polarization groups are defined by splitting molecules between rotatable bonds, as defined by molecular descriptors in the OpenBabel toolkit.[38,65] However, to avoid creating single-atom groups (for example, the central oxygen in an ether, for which both C−O bonds are defined as rotatable), *Poltype* also defines rotatable bonds that should not be split, using SMARTS strings to define the chemical space matching these desired bonds. These manual definitions do not cover all possible chemical space encountered and may sometimes result in groups that are too broad or too narrow for new molecules, as demonstrated here. We believe this to be an inherent difficulty that would be faced in any attempt to automatically assign polarization groups for the AMOEBA force field. Nevertheless, it should be noted that coverage of chemical space and adjustment of atom typing rules is a problem equally faced by automatic parametrization protocols for pairwise-additive force fields.[47,66]

A further limitation of the *Poltype* parametrization is the single conformation used in fitting of atomic multipoles to QM ESP. Use of multiple conformations in a fitting process is well-known to improve electrostatic parameter assignment (for example fitting of partial charges), with associated improvements in representation of intra- and intermolecular interactions during simulation.[21,67−70] For the remaining parameter sets, we therefore make use of a manual parametrization of atomic multipoles using a multiconformational ESP fit, as well as a manual choice of polarization groups for all solutes.

Improvement of Manual Parametrization. Parameter set 3 made use of a multiconformational fit procedure for solute atomic multipoles and manual choice of solute polarization groups. The protocol for choosing solute conformations for fitting, as well as the assigned polarization groups, can be found in Text S1 and Table S2, respectively. Valence and vdW parameters were transferred from those of set 2.

Overall, the use of a multiconformational fit led to a slight improvement in MUE, R^2, and Kendall τ coefficient over the results of set 2, with no clear difference in the MSE metric. This suggests a reduction in the magnitude of random errors in predictions or alternatively an improvement in the robustness of parameters to differences in sampling. To evaluate this robustness, we investigated how well the assigned AMOEBA solute parameters could recreate the QM ESP of a variety of solute conformations against which they were *not* parametrized. In theory, a multiconformational fit generates atomic multipoles that suitably describe the molecular ESP around multiple solute conformations and so, on average, are more robust to changes in molecular conformation than those generated from a single ESP fit.[21,31,69]

We tested this hypothesis using the subgroup of substituted benzene solutes as an example data set, but excluding solutes 28 and 35, for which only a single conformation was determined to be relevant for parametrization (Text S1). Only the substituted benzene solutes used identical polarization groups in set 2 and set 3. Therefore, this directly evaluated the effects of the multiconformational ESP fit compared to set 2, rather than a combination of ESP fitting and polarization group changes. For each of the substituted benzenes, 50 solute-only structural snapshots were extracted from a single "fully on" (λ = 1.0) trajectory of the molecule in solvent. Each snapshot was geometry optimized and subject to a single point QM ESP calculation using the same level of theory as that used in parametrization (HF/6-31G* and MP2/6-311++G(2d,2p) respectively). AMOEBA and QM ESP for each structure are compared in Figure 6.

The use of a multiconformational fit during parametrization notably improves agreement of AMOEBA and QM ESP for conformations visited during the trajectory. In particular, R^2 of a linear fit to each data set improves from 0.44 (set 2, single fit) to 0.97 (set 3, multiple fit), indicating that the parameters

Table 3. Comparison of Result Metrics between Parameter Sets 1 and 2 for the Subgroup of Substituted Benzene Solutes[a]

set	name	MSE (kcal mol^{-1})	MUE (kcal mol^{-1})	R^2	Kendall τ
1	*Poltype*	−0.501 ≤ **0.463** ≤ 1.149	1.064 ≤ **1.391** ≤ 1.839	0.25 ≤ **0.71** ≤ 0.93	0.08 ≤ **0.56** ≤ 0.86
2	new polgps	0.007 ≤ **0.799** ≤ 1.364	0.989 ≤ **1.347** ≤ 1.664	0.30 ≤ **0.75** ≤ 0.93	0.32 ≤ **0.72** ≤ 0.92

[a]Upper and lower bounds for each metric are 95% CI estimates.

DOI: 10.1021/acs.jctc.6b00276
J. Chem. Theory Comput. 2016, 12, 3871−3883

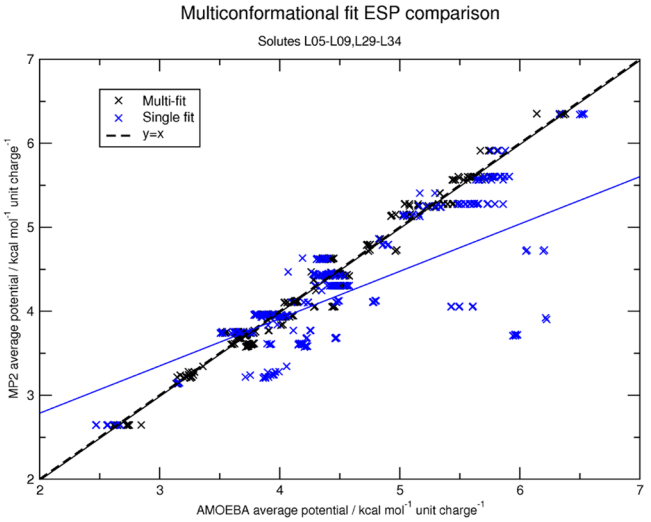


Figure 6. Comparison of mean molecular ESP around substituted benzene solutes calculated with QM and with AMOEBA parameter sets 2 and 3. Fifty solute structures were extracted from each MD simulation at λ = 1.0, geometry optimized, and subject to QM and AMOEBA ESP calculations. Each point corresponds to a single MD structure, but multiple structures may optimize to similar QM geometries, resulting in irregular horizontal clustering of points. Generating multipoles using a multiconformational fit (parameter set 3, black) results in a much more accurate recreation of QM ESP for conformations visited during a simulation (R^2 = 0.97) than a single conformational fit (parameter set 2, blue, and R^2 = 0.44).

generated with a multiconformational fit far better recreate QM ESP for realistic solute conformations visited during trajectories. While the solute structures evaluated were extracted from independent simulations using different parameter sets, the better agreement for parameter set 3 is not simply due to differences in sampling, as snapshots evaluated for both sets exhibit a similar range of RMSD values (Figure S6) when compared to the initial solute structure used in parametrization.

We consider it likely that an improvement in representing molecular ESP would feed through to improved energetics of intermolecular interactions and hence improved free energy of hydration estimates for these ligands. This explains the moderate improvements in MUE, R^2, and Kendall τ observed for set 3 compared to those of set 2.

Parameter set 4 evaluated the effect of fitting solute atomic multipoles to a QM ESP calculated with a larger basis set. In previous AMOEBA HFE studies and the latest AMOEBA force field for proteins, parameters have been fitted to QM ESP calculated using the Dunning-type aug-cc-pVTZ basis set rather than the Pople 6-311++G(2d,2p) set used for parameter sets 1−3.[31,50] AMOEBA parametrization protocols have recommended the use of aug-cc-pVTZ for molecules where this is computationally feasible (<50 atoms),[21] although a detailed understanding of the effect of the basis set has been limited by the size of previous studies.

Results obtained with parameter set 4 are shown in Table 2. In general, set 4 metrics showed little difference to that of set 3, suggesting that any improvements arising from the use of a larger basis set are by no means generalizable across the whole SAMPL4 data set. The largest difference between sets 3 and 4 is in MSE (1.01 and 1.22 kcal mol^{-1}, respectively). This small increase in signed error, corresponding to slightly more positive HFE estimates, is consistent with a previous evaluation of basis set fitting, but effects beyond signed error are clearly minor.[50]

At this stage in parameter optimization, solute electrostatic parameters were considered optimized within the confines of the standard AMOEBA parametrization protocol outlined in Figure 1. Other parameters involved in intermolecular interactions, such as vdW radii or atomic polarizabilities, are defined in AMOEBA by atom type. Reoptimizing and validating parameters for specific atom types are tasks requiring extensive computational effort and comparison with both gas phase QM and bulk phase experimental data, and would not normally be attempted on a compound-by-compound basis.

Nevertheless, the lack of significant improvement in results when changing the basis set used for parametrization may suggest a plateau in the quality of solute electrostatic parameters when considered in isolation. Simultaneous reoptimization of all nonbonded parameters may be necessary to further improve agreement with experiment.[71] The vdW parameters used here have been parametrized based on dimer interaction energies calculated at the MP2/aug-cc-pVTZ level and bulk phase simulations of simple organic molecules[16,21] but could feasibly require some reoptimization for the specific molecular interactions under study here. The use of automated parameter fitting methodologies such as ForceBalance[72] facilitates this type of simultaneous parameter refinement, but as stated above, it remains a challenging and expensive task on a solute-by-solute basis.

Therefore, parameter set 5 moved beyond reparametrization of the solute and instead used an alternative solvent model for HFE simulations. The recent AMOEBA wat14 model was developed making use of the ForceBalance methodology to simultaneously fit multiple parameters to a variety of experimental and QM target data.[72,73] During parametrization, it was found to better recreate a number of experimental water properties, including thermodynamic properties such as enthalpy of vaporization, than the previously used AMOEBA wat03 model.[32] Simulations with parameter set 5 were therefore performed identically to those of set 4, except using the wat14 model for solvent parameters. Solute parameters were taken directly from set 4.

Results with the wat14 model, summarized in Table 2, showed a small reduction in τ, small improvements in MUE and R^2, and a large improvement in MSE over those calculated with the wat03 model. This suggests the wat03 model systematically underestimates HFE for this data set, while wat14 goes some way to overcoming this. The underestimation of HFE with wat03 observed here goes beyond that observed in SAMPL2, where a MSE of 0.446 kcal mol^{-1} was reported across 30 small molecules[16] but is consistent with other previous trends of positive signed errors in HFE estimates.[9,74] In addition, the more negative HFE seen with wat14 is consistent with results observed for monovalent ion HFE in the original wat14 paper.[32] The simulations performed here are the first broader study of wat14 HFE and suggest that its improved performance in recreating water properties also leads to improved representation of intermolecular interactions and accompanying experimental metrics such as HFE.

A further systematic error suggested by previous AMOEBA HFE studies has been the overestimation of interaction energies for aliphatic hydroxyl groups. Atomic quadrupoles in the 2003 AMOEBA water model were scaled to 0.73 of their original value in order to better recreate interaction dimer structures and energetics,[42] and a scaling value of 0.6 for all

DOI: 10.1021/acs.jctc.6b00276
J. Chem. Theory Comput. 2016, 12, 3871−3883

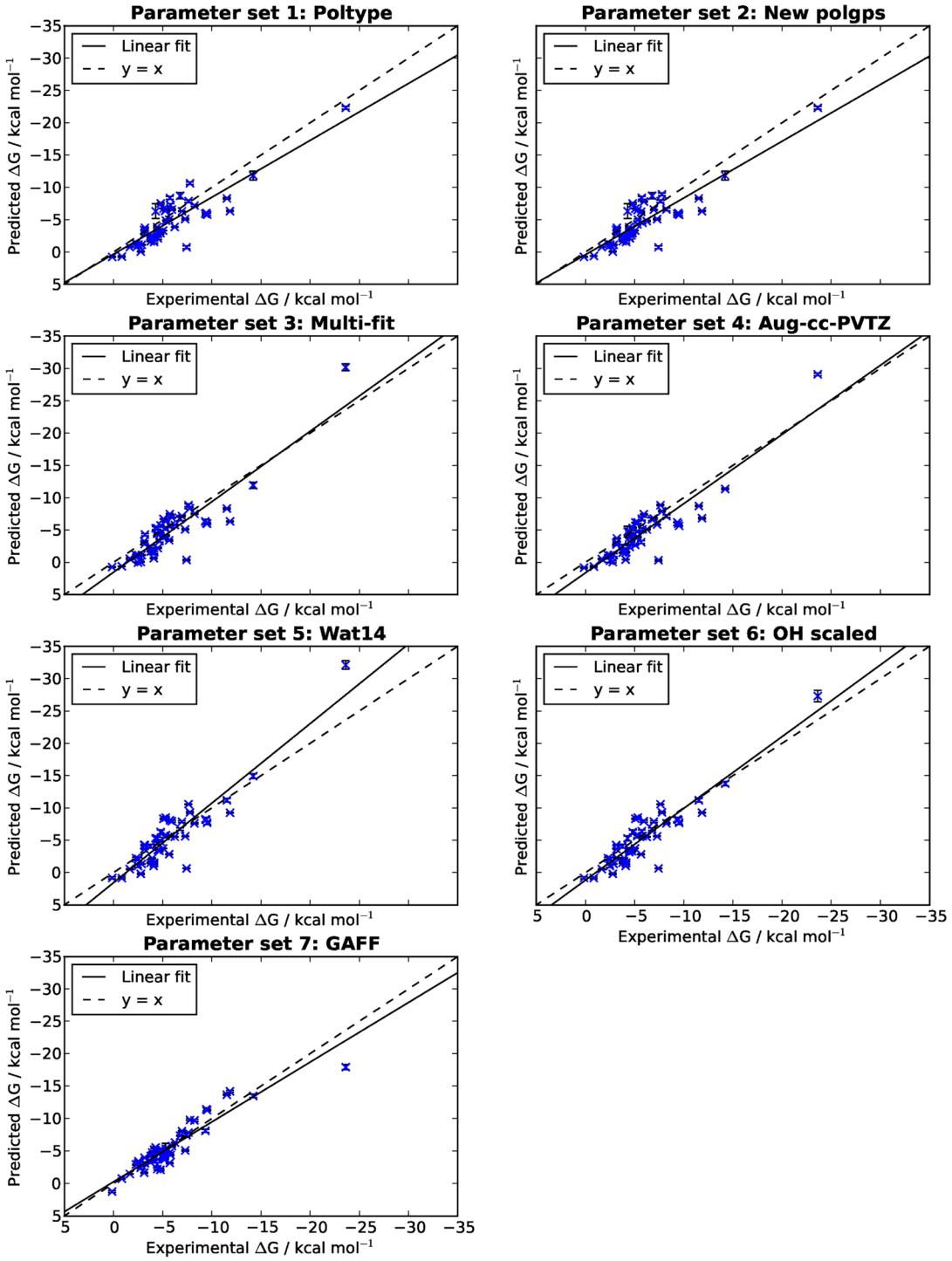


Figure 7. Comparison of predicted and experimental ΔG_{hyd} for all parameter sets and accompanying linear regressions.

aliphatic hydroxyl groups was suggested in the HFE study of Shi and co-workers, albeit only tested on two aliphatic alcohols.[50] Empirical scaling of polarizability parameters has also been suggested for other polarizable force fields to account for the difference in molecular polarizability between the condensed phase and gas phase.[75,76] Parameter set 6 investigated the effect of scaling quadrupoles of hydroxyl atoms in aliphatic alcohol groups by 0.6, in accordance with the empirical recommendations of Shi et al. This affected 7 of the 47 solutes, as detailed in Text S1. All other parameters used were identical to those of set 5.

Overall results from parameter set 6 (Table 2) showed a further small decrease in MUE and small increase in MSE over set 5. These changes are consistent with the reduction of an overestimation of HFE for aliphatic alcohol groups; by making alcohol HFE less negative, the mean signed error of the data set becomes more positive, while the mean unsigned error is reduced in magnitude. This reinforces the validity of the hydroxyl group scaling first suggested by Shi et al. and perhaps reveals the potential for further improvements in AMOEBA HFE prediction through empirical scaling of parameters for additional functional groups, similar to that proposed in

DOI: 10.1021/acs.jctc.6b00276
J. Chem. Theory Comput. 2016, 12, 3871−3883

Table 4. p Values for Statistical Tests between Error Distributions for Different Parameter Sets[a]

		comparison set						
reference set	metric	1	2	3	4	5	6	7
1	MSE	N/A	0.3275	0.9101	0.3671	**0.0050**	**0.0255**	**0.0386**
1	MUE	N/A	0.7537	0.2854	0.4982	0.0884	0.1054	**0.0001**
7	MSE	**0.0386**	**0.0198**	0.0879	**0.0238**	0.9595	0.3956	N/A
7	MUE	**0.0001**	**0.0003**	**0.0040**	**0.0056**	0.0904	0.2236	N/A

[a]Significant differences are highlighted in bold. Tests between MSE distributions use paired Student's t-tests, and those between MUE distributions used Wilcoxon signed-rank tests.

alternative force fields.[75,76] However, as with other reparametrizations of specific atom types discussed earlier, the validation of such empirical scaling factors would require both extensive comparison with experimental properties and data, and a data set large enough to ensure statistical confidence in the findings. As demonstrated when comparing results for the substituted benzene subgroup above, if observed differences are subtle, ensuring significance will require a larger data set than the SAMPL4 compounds tested here.

Parameter set 6 included the cumulative effects of all modifications tested here and can therefore be considered an optimum set of AMOEBA parameters for the SAMPL4 solute set, within the confines of the protocol of Figure 1. Comparison to a pairwise-additive, a fixed-point-charge force field was carried out with parameter set 7, which made use of the GAFF force field for solute valence parameters, AM1-BCC solute charges, and the TIP3P water model. During the SAMPL4 challenge, GAFF/AM1-BCC was among the best performing submissions.[14] Results with parameter set 7 are almost identical to those published during the SAMPL4 challenge and exhibited the lowest errors and highest correlation of the parametrization protocols tested here (Table 2).

Statistical Comparison. Comparison of results between all parameter sets is shown in Figure 7. Over 40 submissions of HFE predictions, carried out with various methodologies, were made to the SAMPL4 challenge. Results of a large number of the submissions were found to be statistically indistinguishable, that is, despite apparent differences in performance, error distributions no longer showed statistically significant differences.[14] We performed similar comparisons in this study to answer two main questions. First, do the cumulative effects of AMOEBA parametrization changes from set 1 to set 6 result in a statistically significant improvement in overall performance? Second, is the observed difference between the GAFF fixed-point-charge force field (set 7) and the optimum AMOEBA parameters (set 6) statistically significant?

Significance tests were performed as outlined in the Methods, and the resulting p values are detailed in Table 4. MSE and MUE results for each parameter set were compared to those of the AMOEBA *Poltype* parameters (set 1) and fixed-charge GAFF parameters (set 7) individually, with significant differences defined as those with $p < 0.05$.

The majority of changes between parameter sets 1−6 resulted in minor improvements to mean errors, and therefore, many results sets remained statistically indistinguishable from one another. Cumulative improvements in AMOEBA parameters only resulted in a significant difference in mean signed error for sets 5 and 6 (Table 4, row 1), both of which made use of the latest wat14 AMOEBA water model. This lends further weight to the conclusion that the wat14 model is a significant improvement in modeling of water interactions with AMOEBA, at least for thermodynamic applications such as the HFE calculations tested here. While differences in unsigned error remained nonsignificant for all AMOEBA sets (Table 4, row 2), there was a clear reduction in p value as cumulative improvements were made. Additional HFE studies with a larger data set may therefore clarify the significance (or otherwise) of observed improvements in MUE.

Comparisons of AMOEBA and GAFF results (Table 4, rows 3−4) initially showed significant differences between the methodologies, but as AMOEBA results improved, the two parameter sets became indistinguishable for both MSE and MUE metrics. In this sense, our results are consistent with the findings of the SAMPL4 study, namely, that as parametrization protocols or prediction methods improve, results begin to show a consensus among high-performing methods. Our cumulative improvements to AMOEBA parametrization have resulted in a set of predictions that on average are statistically comparable with those of the GAFF fixed-point-charge force field.

A lack of significant difference between parameter sets for the data set as a whole may hide significant changes in predictions for individual solutes. We therefore performed a one-way ANOVA test for each solute across HFE results with the AMOEBA parameter sets 1−6. For each molecule, this evaluated whether at least one of the parameter sets was different from any other. Resulting p values are available in Table S4. Almost every solute showed a significant difference in at least one of the parameter sets, with only solutes 16, 25, 35, 38, 45, and 47 having statistically indistinguishable predictions across all six parameter sets. These were predominantly among the more hydrophobic, conformationally rigid molecules in the data set, so it is perhaps unsurprising that parametrization changes focusing on treatment of polarization and of conformational changes did not result in significant differences in HFE for these molecules. Solute 25, 1,3-bis(nitrooxy)butane, was not among the most hydrophobic compounds but instead showed one of the largest variances between simulation repeats, suggesting that sampling variations during simulations led to the statistical equivalence of results for this molecule.

Results for individual solutes were further compared pairwise across parameter sets by means of a Tukey HSD test. This is a *posthoc* analysis, carried out following an ANOVA performed across all 7 parameter sets, which more robustly compares means of multiple samples than pairwise Student's t tests.[77] A summary of these comparisons is available in Figure S5, while results for all 47 solutes are available as an online data set.[78] Despite cumulative parameter improvements, the majority of solutes (37/47) show significant differences between GAFF results and the final AMOEBA parameter set 6. We note, however, that a simple statistically significant difference between solutes provides no information on which parameter set shows a better agreement with experiment, nor on the performance of a parameter set as a whole; indeed, as shown above, the overlap of error distributions between AMOEBA

DOI: 10.1021/acs.jctc.6b00276
J. Chem. Theory Comput. 2016, 12, 3871−3883

and GAFF means that results cannot be considered different as a whole. Inconsistencies between the two analyses, as observed here, serve to reiterate the importance of using large and diverse data sets when evaluating methodologies.

CONCLUSIONS

Parametrization changes, particularly for an advanced, nonadditive potential energy function such as AMOEBA, may have subtle and interlinked effects on the agreement of computed thermodynamic properties with experiment. Starting from a set of automatically assigned parameters, we have investigated a number of options for modifying parametrization, all of which are compatible with the wider AMOEBA force field. Overall, these options have shown small but sequential improvements in agreement of AMOEBA hydration free energy predictions with experiment, which have long been used as an evaluation tool for computational methodologies and force field parameters.

The cumulative effects of each change set allow us to draw a number of conclusions regarding parametrization protocols for adding molecules ad hoc to the AMOEBA force field, for example, in protein−ligand binding studies. First, while automated parametrization software such as *Poltype* can quickly assign a valid set of initial parameters, manual intervention should be used to refine these and ensure their suitability for the desired system. Here, parameter sets 1 and 2, which, respectively, cumulatively improve solute dihedral and polarization group assignment, show lower MSE and MUE than the original SAMPL4 submission of Manzoni and Söderhjelm, despite similar simulation methods.[39] The assignment of polarization group in particular is nontrivial to automate. We would expect other parameters that may be assigned automatically using molecular symmetry, for example, reference frames for atomic multipoles, to also occasionally require manual improvement and careful assignment.[69] However, no specific misassignments were identified in this regard for the molecules tested here.

Second, where relevant, multipole parameters should be refined using a multiconformational fitting procedure in the TINKER POTENTIAL program. Parameter set 3 improved error and correlation metrics over the single fit procedures. Choice of which relevant conformations to include may clearly affect parameters for individual solutes, and the standardized protocol we use here (Text S1) may be suboptimal, particularly for the larger, more conformationally diverse molecules. Nevertheless, multiconformational parametrization can exhibit clear improvements and is in accordance with the philosophy of the broader AMOEBA protein force field.

Third, in calculations of the QM molecular ESP around conformations used for multipole fitting, basis set size does not seem to have a large effect on the agreement of HFE with experiment. Although a previous AMOEBA HFE study recommended the use of aug-cc-pVTZ or larger, the effect size was small (0.23 kcal mol^{-1} improvement in MUE) and the data set limited (n = 7).[50] Our results, from parameter set 4, are in general agreement with those published, in that any effects are small, but the larger data set used here suggests that benefits of a basis set beyond 6-311++G(2d,2p) will likely be molecule-dependent and that further investigation with a far larger data set would be required to identify statistically significant differences for such a small effect size.

Fourth, the recent AMOEBA 2014 water model used in parameter set 5 does improve HFE estimates for small molecules, particularly via a large, statistically significant reduction in MSE. To our knowledge, this is the first broad study of the effects of the 2014 water model on small molecule solvation and suggests the elimination of a systematic underestimation of HFE by the previous AMOEBA water model.

Fifth, the empirical scaling of quadrupole parameters on aliphatic hydroxyl groups (parameter set 6) also slightly improves agreement of the overall data set with experiment and validates the previous proposals in this regard from HFE studies and in the derivation of the original AMOEBA water model.[42,50] It is possible that further empirical modifications to specific functional group parameters may continue to improve agreement. However, robust validation of these empirical adjustments would require the investigation of a much larger data set of small molecules, with multiple well-defined members of each functional group class. Although such studies have become commonplace for fixed-charge force fields, the AMOEBA force field has been restricted to smaller data sets and blind challenges such as the SAMPL set evaluated here because of its computational cost.

Finally, the cumulative effect of the parametrization improvements identified herein has resulted in a set of AMOEBA HFE results whose error distributions are not statistically distinguishable from those of GAFF. While our aim in this study has not been to compare a polarizable and nonpolarizable force field directly, it is encouraging that our findings are consistent with those of the SAMPL4 challenge, i.e., that a large number of methodologies converged upon a set of statistically indistinguishable HFE results.[14] In addition, the continuing development of the AMOEBA potential, particularly the incorporation of experimental and theoretical properties as explicit targets for systematic parameter fitting,[32,73,79−81] may result in further improvements in basic thermodynamic metrics such as HFE. Examining the performance of polarizable force fields for a wide variety of applications remains a challenging prospect. Further systematic studies, such as that presented here, will aid the maturation of polarizable force fields and encourage the expansion of their uptake, in line with the long development process of pairwise additive force fields in recent decades.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00276.

Full details of AMOEBA parametrization choices and methods for each parameter set, all SAMPL4 solute structures and details of manually defined solute polarization groups, comparison of solute RMSD between parameter sets 2 and 3, definitions of solute subgroups, statistical p values for the ANOVA test between solute HFE predictions with different methods, and a summary of pairwise significant differences between solute results across parameter sets are available (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: J.W.Essex@soton.ac.uk.

Funding

R.T.B. and J.W.E. gratefully acknowledge support from EPSRC, grant number EP/K039156/1.

Notes

The authors declare no competing financial interest.

Underlying research data, parametrization and statistical analysis scripts and results (including all calculated hydration free energies) are available without restriction and have been uploaded to the Zenodo data repository.[78,82]

■ ACKNOWLEDGMENTS

Calculations in this work made use of the Iridis3 and Iridis4 supercomputers at the University of Southampton. We thank all members of the EPSRC-NSF funded SI^2 consortium, particularly Jay Ponder, Teresa Head-Gordon, Martin Head-Gordon, David Case, Jason Swails, Chris-Kriton Skylaris, Mark Tuckerman, Paul Nerenberg, Lorna Smith, and Ilian Todorov, for helpful discussions throughout.

■ ABBREVIATIONS

AM1-BCC, AM1-bond charge correction; AMOEBA, atomic multipole optimized energetics for biomolecular applications; ANOVA, analysis of variance; CI, confidence interval; DMA, distributed multipole analysis; ESP, electrostatic potential; GAFF, general Amber force field; HFE, hydration free energy; HSD, honest significant difference; MD, molecular dynamics; MSE, mean signed error; MUE, mean unsigned error; QM, quantum mechanics; RESP, restrained electrostatic potential; RMSD, root mean square deviation; SAMPL, statistical assessment for modeling of proteins and ligands; SMARTS, Smiles arbitrary target specification; vdW, van der Waals

■ REFERENCES

(1) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely Precise Free Energy Calculations of Amino Acid Side Chain Analogs: Comparison of Common Molecular Mechanics Force Fields for Proteins. *J. Chem. Phys.* **2003**, *119* (11), 5740.

(2) Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. Predicting Small-Molecule Solvation Free Energies: An Informal Blind Test for Computational Chemistry. *J. Med. Chem.* **2008**, *51* (4), 769−779.

(3) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies Using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, *6* (5), 1509−1519.

(4) Martins, S. A.; Sousa, S. F.; Ramos, M. J.; Fernandes, P. A. Prediction of Solvation Free Energies with Thermodynamic Integration Using the General Amber Force Field. *J. Chem. Theory Comput.* **2014**, *10* (8), 3570−3577.

(5) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, *10* (7), 2632−2647.

(6) Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303* (5665), 1813−1818.

(7) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. SM6: A Density Functional Theory Continuum Solvation Model for Calculating Aqueous Solvation Free Energies of Neutrals, Ions, and Solute−Water Clusters. *J. Chem. Theory Comput.* **2005**, *1* (6), 1133−1152.

(8) Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges. *J. Chem. Theory Comput.* **2007**, *3* (6), 2011−2033.

(9) Mobley, D. L.; Guthrie, J. P. FreeSolv: A Database of Experimental and Calculated Hydration Free Energies, with Input Files. *J. Comput.-Aided Mol. Des.* **2014**, *28* (7), 711−720.

(10) Guthrie, J. P. A Blind Challenge for Computational Solvation Free Energies: Introduction and Overview. *J. Phys. Chem. B* **2009**, *113* (14), 4501−4507.

(11) Geballe, M. T.; Skillman, A. G.; Nicholls, A.; Guthrie, J. P.; Taylor, P. J. The SAMPL2 Blind Prediction Challenge: Introduction and Overview. *J. Comput.-Aided Mol. Des.* **2010**, *24* (4), 259−279.

(12) Geballe, M. T.; Guthrie, J. P. The SAMPL3 Blind Prediction Challenge: Transfer Energy Overview. *J. Comput.-Aided Mol. Des.* **2012**, *26* (5), 489−496.

(13) Guthrie, J. P. SAMPL4, a Blind Challenge for Computational Solvation Free Energies: The Compounds Considered. *J. Comput.-Aided Mol. Des.* **2014**, *28* (3), 151−168.

(14) Mobley, D.; Wymer, K.; Lim, N.; Guthrie, J. P. Blind Prediction of Solvation Free Energies from the SAMPL4 Challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28* (3), 135−150.

(15) SAMPL5 https://drugdesigndata.org/about/sampl5 (accessed Dec 16, 2015).

(16) Ponder, J. W.; Wu, C. J.; Ren, P. Y.; 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), 2549−2564.

(17) Demerdash, O.; Yap, E.-H.; Head-Gordon, T. Advanced Potential Energy Surfaces for Condensed Phase Simulation. *Annu. Rev. Phys. Chem.* **2014**, *65* (1), 149−174.

(18) Shi, Y.; Ren, P.; Schnieders, M.; Piquemal, J.-P. Polarizable Force Fields for Biomolecular Modeling. In *Reviews in Computational Chemistry*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2015; Vol. *28*, pp 51−86.

(19) Huang, J.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Recent Advances in Polarizable Force Fields for Macromolecules: Microsecond Simulations of Proteins Using the Classical Drude Oscillator Model. *J. Phys. Chem. Lett.* **2014**, *5* (18), 3144−3150.

(20) Baker, C. M. Polarizable Force Fields for Molecular Dynamics Simulations of Biomolecules. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, *5* (2), 241−254.

(21) Ren, P. Y.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7* (10), 3143−3161.

(22) Baker, C. M.; Lopes, P. E. M.; Zhu, X.; Roux, B.; Mackerell, A. D. Accurate Calculation of Hydration Free Energies Using Pair-Specific Lennard-Jones Parameters in the CHARMM Drude Polarizable Force Field. *J. Chem. Theory Comput.* **2010**, *6* (4), 1181−1198.

(23) Zhong, Y.; Patel, S. Nonadditive Empirical Force Fields for Short-Chain Linear Alcohols: Methanol to Butanol. Hydration Free Energetics and Kirkwood-Buff Analysis Using Charge Equilibration Models. *J. Phys. Chem. B* **2010**, *114* (34), 11076−11092.

(24) 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), 1314−1324.

(25) Fried, S. D.; Wang, L.-P.; Boxer, S. G.; Ren, P.; Pande, V. S. Calculations of the Electric Fields in Liquid Solutions. *J. Phys. Chem. B* **2013**, *117* (50), 16236−16248.

(26) Kuster, D. J.; Liu, C.; Fang, Z.; Ponder, J. W.; Marshall, G. R. High-Resolution Crystal Structures of Protein Helices Reconciled with Three-Centered Hydrogen Bonds and Multipole Electrostatics. *PLoS One* **2015**, *10* (4), e0123146.

(27) El Hage, K.; Piquemal, J.-P.; Hobaika, Z.; Maroun, R. G.; Gresh, N. Substituent-Modulated Affinities of Halobenzene Derivatives to the HIV-1 Integrase Recognition Site. Analyses of the Interaction Energies by Parallel Quantum Chemical and Polarizable Molecular Mechanics. *J. Phys. Chem. A* **2014**, *118* (41), 9772−9782.

(28) MacDermaid, C. M.; Kaminski, G. A. Electrostatic Polarization Is Crucial for Reproducing pKa Shifts of Carboxylic Residues in Turkey Ovomucoid Third Domain. *J. Phys. Chem. B* **2007**, *111* (30), 9036−9044.

(29) Lemkul, J. A.; Savelyev, A.; MacKerell, A. D. Induced Polarization Influences the Fundamental Forces in DNA Base Flipping. *J. Phys. Chem. Lett.* **2014**, *5* (12), 2077−2083.

(30) Huang, J.; MacKerell, A. D. Induction of Peptide Bond Dipoles Drives Cooperative Helix Formation in the (AAQAA)3 Peptide. *Biophys. J.* **2014**, *107* (4), 991−997.

DOI: 10.1021/acs.jctc.6b00276
J. Chem. Theory Comput. 2016, 12, 3871−3883

(31) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046−4063.
(32) Laury, M. L.; Wang, L.-P.; Pande, V. S.; Head-Gordon, T. L.; Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B* **2015**, *119* (29), 9423−9437.
(33) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardère, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. A General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015**, *11* (6), 2609−2618.
(34) Lagardère, L.; Lipparini, F.; Polack, E.; Stamm, B.; Schnieders, M.; Cances, E.; Ren, P.; Maday, Y.; Piquemal, J.-P. Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: II.Towards Massively Parallel Computations Using Smooth Particle Mesh Ewald. *J. Chem. Theory Comput.* **2015**, *11* (6), 2589−2599.
(35) Albaugh, A.; Demerdash, O.; Head-Gordon, T. An Efficient and Stable Hybrid Extended Lagrangian/self-Consistent Field Scheme for Solving Classical Mutual Induction. *J. Chem. Phys.* **2015**, *143* (17), 174104.
(36) Lindert, S.; Bucher, D.; Eastman, P.; Pande, V.; McCammon, J. A. Accelerated Molecular Dynamics Simulations with the AMOEBA Polarizable Force Field on Graphics Processing Units. *J. Chem. Theory Comput.* **2013**, *9* (11), 4684−4691.
(37) Peng, X.; Zhang, Y.; Chu, H.; Li, G. Free Energy Simulations with the AMOEBA Polarizable Force Field and Metadynamics on GPU Platform. *J. Comput. Chem.* **2016**, *37* (6), 614−622.
(38) Wu, J. C.; Chattree, G.; Ren, P. Automation of AMOEBA Polarizable Force Field Parameterization for Small Molecules. *Theor. Chem. Acc.* **2012**, *131* (3), 1138.
(39) Manzoni, F.; Söderhjelm, P. Prediction of Hydration Free Energies for the SAMPL4 Data Set with the AMOEBA Polarizable Force Field. *J. Comput.-Aided Mol. Des.* **2014**, *28* (3), 235−244.
(40) Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K. The SAMPL4 Host−guest Blind Prediction Challenge: An Overview. *J. Comput.-Aided Mol. Des.* **2014**, *28* (4), 305−317.
(41) Ren, P. Y.; Ponder, J. W. Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanics Calculations. *J. Comput. Chem.* **2002**, *23* (16), 1497−1506.
(42) Ren, P. Y.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107* (24), 5933−5947.
(43) Thole, B. T. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chem. Phys.* **1981**, *59* (3), 341−350.
(44) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23* (16), 1623−1641.
(45) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; 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.
(46) 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* (9), 1157−1174.
(47) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25* (2), 247−260.
(48) Stone, A. J. Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution. *Chem. Phys. Lett.* **1981**, *83* (2), 233−239.
(49) Stone, A. J.; Alderton, M. Distributed Multipole Analysis: Methods and Applications. *Mol. Phys.* **1985**, *56* (5), 1047−1064.
(50) Shi, Y.; Wu, C. J.; Ponder, J. W.; Ren, P. Y. Multipole Electrostatics in Hydration Free Energy Calculations. *J. Comput. Chem.* **2011**, *32* (5), 967−977.
(51) Ren, P. Y.; Ponder, J. W. Temperature and Pressure Dependence of the AMOEBA Water Model. *J. Phys. Chem. B* **2004**, *108* (35), 13427−13437.
(52) Ponder, J. *TINKER: Software Tools for Molecular Design*; Washington University School of Medicine: St. Louis, MO, 2001.
(53) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E. I.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P. A.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M., Jr.; Paesani, F.; Roe, D. R.; Roitberg, A. E.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J. M.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *AMBER 14*; University of California: San Francisco, CA, 2014.
(54) Pastor, R. W.; Brooks, B. R.; Szabo, A. An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Mol. Phys.* **1988**, *65*, 1409−1419.
(55) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N′-methylamide. *Biopolymers* **1992**, *32* (5), 523−535.
(56) Izaguirre, J. A.; Catarello, D. P.; Wozniak, J. M.; Skeel, R. D. Langevin Stabilization of Molecular Dynamics. *J. Chem. Phys.* **2001**, *114*, 2090−2098.
(57) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684.
(58) Sagui, C.; Pedersen, L. G.; Darden, T. A. Towards an Accurate Representation of Electrostatics in Classical Force Fields: Efficient Implementation of Multipolar Interactions in Biomolecular Simulations. *J. Chem. Phys.* **2004**, *120* (1), 73−87.
(59) Cerutti, D. S.; Duke, R.; Freddolino, P. L.; Fan, H.; Lybrand, T. P. A Vulnerability in Popular Molecular Dynamics Packages Concerning Langevin and Andersen Dynamics. *J. Chem. Theory Comput.* **2008**, *4* (10), 1669−1680.
(60) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926.
(61) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(62) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089.
(63) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22* (2), 245−268.
(64) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129* (12), 124105.
(65) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3* (1), 33.
(66) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52* (12), 3144−3154.
(67) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* **1995**, *16* (11), 1357−1377.
(68) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821−7839.
(69) Kramer, C.; Gedeck, P.; Meuwly, M. Atomic Multipoles: Electrostatic Potential Fit, Local Reference Axis Systems, and Conformational Dependence. *J. Comput. Chem.* **2012**, *33* (20), 1673−1688.

DOI: 10.1021/acs.jctc.6b00276
J. Chem. Theory Comput. 2016, 12, 3871−3883

(70) Reynolds, C. A.; Essex, J. W.; Richards, W. G. Atomic Charges for Variable Molecular Conformations. *J. Am. Chem. Soc.* **1992**, *114* (23), 9075−9079.
(71) Kramer, C.; Gedeck, P.; Meuwly, M. Multipole-Based Force Fields from Ab Initio Interaction Energies and the Need for Jointly Refitting All Intermolecular Parameters. *J. Chem. Theory Comput.* **2013**, *9* (3), 1499−1511.
(72) Wang, L.-P.; Chen, J.; Van Voorhis, T. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *J. Chem. Theory Comput.* **2013**, *9* (1), 452−460.
(73) 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), 1885−1891.
(74) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.* **2009**, *5* (2), 350−358.
(75) Baker, C. M.; Mackerell, A. D. Polarizability Rescaling and Atom-Based Thole Scaling in the CHARMM Drude Polarizable Force Field for Ethers. *J. Mol. Model.* **2010**, *16* (3), 567−576.
(76) Lopes, P. E. M.; Lamoureux, G.; Mackerell, A. D. Polarizable Empirical Force Field for Nitrogen-Containing Heteroaromatic Compounds Based on the Classical Drude Oscillator. *J. Comput. Chem.* **2009**, *30* (12), 1821−1838.
(77) Tukey, J. W. Comparing Individual Means in the Analysis of Variance. *Biometrics* **1949**, *5* (2), 99.
(78) Bradshaw, R. T.; Essex, J. W. Supplementary underlying data for "Evaluating parameterization protocols for hydration free energy calculations with the AMOEBA polarizable force field" http://dx.doi.org/10.5281/zenodo.54959.
(79) Wang, L.-P.; Chen, J.; Van Voorhis, T. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *J. Chem. Theory Comput.* **2013**, *9* (1), 452−460.
(80) 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), 9956−9972.
(81) 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.
(82) Bradshaw, R. T.; Essex, J. W. Underlying data for "Evaluating parameterization protocols for hydration free energy calculations with the AMOEBA polarizable force field" http://dx.doi.org/10.5281/zenodo.35586.

DOI: 10.1021/acs.jctc.6b00276
J. Chem. Theory Comput. 2016, 12, 3871−3883