Update to the General Amber Force Field for Small Solutes with an Emphasis on Free Energies of Hydration

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · MARCH 2014

Impact Factor: 3.3 · DOI: 10.1021/jp4111234 · Source: PubMed

2 AUTHORS:

Joakim P M Jämbeck
Stockholm University
12 PUBLICATIONS **253** CITATIONS
SEE PROFILE

Alexander P Lyubartsev
Stockholm University
155 PUBLICATIONS **4,760** CITATIONS
SEE PROFILE

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

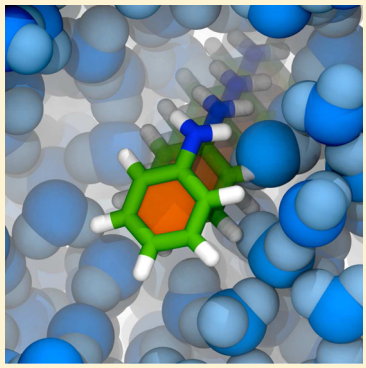
pubs.acs.org/JPCB

Update to the General Amber Force Field for Small Solutes with an Emphasis on Free Energies of Hydration

Joakim P. M. Jämbeck and Alexander P. Lyubartsev*

Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, Stockholm SE-10691, Sweden

ABSTRACT: An approach to a straightforward reparametrization of the general AMBER force field (GAFF) for small organic solutes and druglike compounds is presented. The parametrization is based on specific pair interactions between the solvent and the solute, namely, the interactions between the constituting atoms of the solute and the oxygen of water were tuned in order to reproduce experimental hydration free energies for small model compounds. The key of the parametrization was to abandon the Lorentz−Berthelot combination rules for the van der Waals interactions. These parameters were then used for larger solutes in order to validate the newly derived pair interactions. In total close to 600 hydration free energies are computed, ranging from simple alkanes to multifunctional drug compounds, and compared to experimental data. The results show that the proposed parameters work well in describing the interactions between the solute and the solvent and that the agreement in absolute numbers is good. This modified version of GAFF is a good candidate for computing and predicting hydration free energies on a large scale, which has been a long-sought goal of computational chemists and can be used in rational drug design.



INTRODUCTION

The rapid progress in available computational power for the molecular modeling society has made it possible to critically validate the vast number of modeling techniques available. One of the most central concepts of physical chemistry and subsequently in general chemistry and biology is the free energy, and accurate calculations of this property has been a holey grail for computational chemists for several decades. Hydration free energies often serve as an ultimate test for the computational methods.[1−9] In the field of rational drug design the free energy calculations are of great interest since the solvation free energies determine such fundamental properties of drug-related molecules as solubility and partitioning and thus especially important in preliminary screening of perspective drug candidates. In addition, the free energy computations are used to predict ligand binding affinities,[9,10] as there is an energetic and entropic penalty for a ligand and its target protein related to the desolvation that has to occur during the formation of the complex. Knowledge of free energies is also important in description of enzymatic reactions.[11] Furthermore, free energies determine the partitioning of solutes between solvents of different dielectric nature such as water/*n*-octanol[12−16] or water/lipid bilayers,[17−25] which is inherently related to the distribution and transport of the molecules in living organisms.

One common way of predicting free energies is using molecular dynamics (MD) simulations, where the equations of motion of the atomistic systems are solved using Newtonian mechanics. If the system of interest is sampled using MD techniques (Monte Carlo sampling techniques can also be used), the molecular interactions are often described by an empirical potential energy function, usually referred to as a force field (FF). Popular FFs for biomolecular simulations are AMBER,[26,27] CHARMM,[28,29] GROMOS[30] and OPLS-AA.[31,32] These FFs have often been tuned for amino and nucleic acids to be used in simulations of large biomolecular systems, but lately more general FFs have been developed in order to be used for smaller, druglike compounds such as the the general AMBER FF (GAFF)[33] and CHARMM general FF.[34] Naturally the predictive character of the simulation depends on the quality of the FF, and therefore free energy calculations have remained as an ultimate test for these empirical potentials. Few FFs have been parametrized against free energies, although there are exceptions, e.g., GROMOS.[30] Traditionally it has been the large computational cost of conducting these calculations that has been the obstacle, but because the computational resources available are constantly growing, these issues can begin to be addressed.[35] Nowadays a number of established and validated free energy computation techniques are available which are also implemented in software of general use, so that one can always choose the technique most suitable for the problem at hand.[36] Of course necessity of the proper sampling during the simulations should still be mentioned as a lack of thereof can have a serious impact on the precision of the computations.[19,25,37−40]

In the present investigation we present a modification of GAFF for aqueous solutes. The modification does not require any reparametrization of the original GAFF parameters as only van der Waals interactions between specific pairs of the oxygen of the water molecules and the constituting atoms of the solute are modified by introducing corrections to the Lorentz−Berthelot combination rules. A number of small model

Received: November 12, 2013
Revised: March 17, 2014
Published: March 19, 2014

dx.doi.org/10.1021/jp4111234 | *J. Phys. Chem. B* 2014, 118, 3793−3804

compounds are first chosen for which these interactions are tuned until the agreement between the simulations and experiments is satisfactory. The newly derived parameters are then validated by computing the hydration free energy for a large number of small organic solutes and druglike compounds. The validation shows that the suggested FF parameters are well transferable for other molecules and can be used for accurate predictive calculations of hydration free energies without performing an extensive reparametrization of all of the GAFF parameters. This is especially important in, e.g., rational drug design where large sets of molecules are investigated and the GAFF parameters can be assigned automatically using the Antechamber package.[41]

METHODS AND MODELS

Potential Form and Parametrization Procedure. Because only the Lennard-Jones (LJ) interactions between the solutes and water are subject to the parametrization, the original potential for GAFF is kept

$$U = \sum_{\text{bonds}} k_b(l - l_0)^2 + \sum_{\text{angles}} k_a(\theta - \theta_0)^2 + \sum_{\text{torsion}} \frac{1}{2}k_\phi(1 + \cos(n\phi - \delta)) + \sum_i \sum_{j>i} 4\epsilon_{ij}\left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right) + \sum_i \sum_{j>i} \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}} \quad (1)$$

where all of the intramolecular parameters (three first sums) are taken from the original GAFF[33] and modifications are made to the first double sum (see subsequent text). Free energy calculations are often used as the ultimate benchmark for these types of empirical functions,[7,42−44] and since the free energy of solvation is hardly affected by the intramolecular potentials for small to medium sized compounds, there are several benefits of only refitting specific nonbonded interactions. First, there is no need to reparametrize the bonded interactions which can take a considerable amount of time, and second, if the nonbonded terms are refitted in order to bring, e.g., bulk properties of organic liquids closer to the experimental values,[45−47] the interactions between the solute and water would be unaffected. Therefore this parametrization scheme makes it easy to build upon it further while maintaining the proper behavior in aqueous solution.

The parametrization of the specific LJ interaction pairs between GAFF and the oxygen of water follows the procedure used by Baker et al.[48] and Nerenberg et al.:[35] the optimization is based on target free energies of model compounds. In order to drastically reduce the computational resources required for simultaneous fitting of more than 50 interaction pairs, the fitting was done in a sequential manner meaning that at each step of the fitting procedure a single chemical entity of interest was parametrized to a small model compound having this entity. The final LJ interactions obtained from this fitting were then used for the next model compound; e.g., first parameters for alkanes were derived that were then used as fixed parameters while parametrizing the alcohol−water interactions of methanol. In Table 1 the model compounds used are listed in the sequential order they were parametrized in. For atom types not listed in Table 1, the original GAFF parameters were used.

Table 1. Model Compounds and Their Constituting Atom Types Used in the Parametrization Scheme

model compd	atom type(s)	GAFF original	description
propane	CT, HC	c3, h3	aliphatic carbon and hydrogen
toluene	CA, HA	ca, ha	aromatic carbon and hydrogen
ethyl methyl sulfide	S, H1	s, h1	sulfide and hydrogen of a carbon bound to an electronegative species
methanol	OH	oh	hydroxyl oxygen
acetone	C, O	c, o	sp^2 carbonyl carbon and oxygen
methylamine	N3	n3	sp^3 nitrogen in a primary amines
trimethylamine	NR3[a]	n3	sp^3 nitrogen in a tertiary amine
dimethyl ether	OS	os	ether oxygen
acetalaldehyde	H4	h4	hydrogen bound to a non-sp^3 carbon bound to an electronegative specie
acetamide	N	n	sp^2 nitrogen in a primary amide group
methylacetamide	NX2, OX2[a]	n, o	sp^2 nitrogen and oxygen in a secondary amide
dimethylacetamide	NX3, OX3[a]	n, o	sp^2 nitrogen and oxygen in a tertiary amide
methyl ester	OX, CX[a]	os, c	ester oxygen and sp^2 carbon in ester
ethylene	C2	c2	sp^2 carbon
methyl chloride	Cl	cl	chloride
methyl flouride	F	f	fluorine
bromomethane	Br	br	bromine
nitroethane	NO, O2N[a]	no, o	nitrogen and oxygen of a nitrogroup
pyridine	NB	nb	aromatic nitrogen

[a]Atom type not present in the original GAFF. The best possible GAFF prototype (used also in bonded interactions) is given in the third column (GAFF original).

Since GAFF uses the standard Lorentz−Berthelot combination rules for the van der Waals (vdW) interactions

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}; \quad \epsilon = \sqrt{\epsilon_i \epsilon_j} \quad (2)$$

the parametrization scheme is relatively straightforward; optimization constants can be introduced between the solute atom X and the oxygen atom of water O_w

$$\sigma_{ij}^{X-O_w} = a_{ij}\frac{\sigma_i + \sigma_j}{2}; \quad \epsilon_{ij}^{X-O_w} = b_{ij}\sqrt{\epsilon_i \epsilon_j} \quad (3)$$

Optimization factors a_{ij} and b_{ij} modify Lorents−Berthelot combination rules for specific solute−water oxygen interactions introducing in fact new Lennard-Jones parameters for such cross-interactions. These factors are then varied until the desirable free energy of hydration is obtained. Initially a series of simulations are performed with a and b ranging from 0.7 to 1.3; in total 36 simulations were performed. These simulations are then analyzed with respect to ΔG_{hyd} and refined in order to bring closer agreement between the simulations and experiments. Because performing such a large number of relatively computer demanding free energy calculations would take a considerable amount of time, the alternative thermodynamic cycle is described in Figure 1, adapted from Baker et al.[48] The

dx.doi.org/10.1021/jp4111234 | J. Phys. Chem. B 2014, 118, 3793−3804

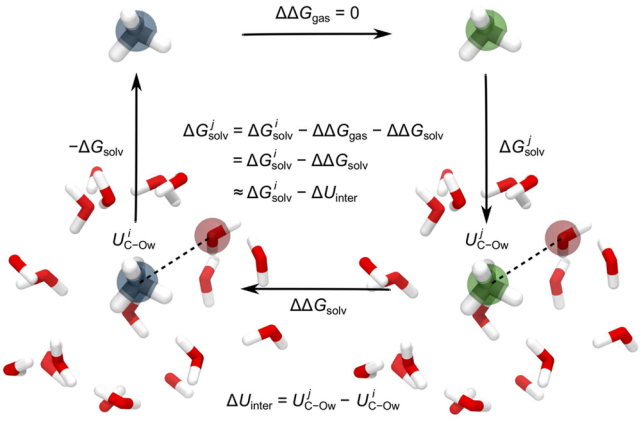


Figure 1. Thermodynamic cycle used during the parametrization procedure. Since only the solute−water interactions are modified between each step, $\Delta\Delta G_{gas} = 0$ and $\Delta\Delta G_{solv}$ is calculated from the difference in the solute−water interactions. The latter is then verified by a final simulation where the free energy of hydration is computed.

main idea is to substitute a relatively expensive full free energy calculation with trial parameters (rightmost vertical path) by calculation of free energy change due to a small change of parameters (the lowermost horizontal path) which can be done by inexpensive perturbation calculation. An initial free energy of hydration is calculated with the original GAFF parameters (leftmost vertical path). After this a series of plain equilibrium simulations are performed with explicit LJ pairs between the model compound and the oxygen of water in the *NPT* ensemble. From these simulations the interaction energies between the solute and water are calculated and used, within the free energy perturbation expression, to determine the free energy difference between the original simulations and the simulation with the specific pair interactions (lowermost horizontal path). Since the parametrization scheme only involves specific pair interactions with water, the free energy difference between the parameters in the gaseous phase during the iterative procedure is zero (uppermost horizontal path). The cycle is then closed by calculating the new free energy of hydration by following the rightmost vertical path. This cycle is performed until satisfactory agreement has been obtained and the only parameters allowed to change between the iterations are *a* and *b*. As a final check a proper free energy calculation is performed to ensure that the parameters are indeed correct. Once the optimal parameters have been obtained for a model compound, the cycle is repeated with another model compound that may or may not use some of the previously derived parameters. This procedure is also possible to use with solvents other than water, and it is likely that one could obtained a highly balanced FF if a large number of specific interactions are fitted on the basis of free energies. Since this is outside the scope of our work, we limit our work to water but point out that this could be a future development.

The water model of choice for our parametrization is TIP3P,[49] and at first glance it might appear as if the parameters derived are highly dependent on the water model used. However, as shown further on, the parameters are relatively robust to a change in water model. The reason for choosing the TIP3P water model instead of, e.g., the four-site models such as TIP4P,[49] TIP4P-Ew,[50] or TIP4P-2005[51] can be explained by several factors. One is the computational efficiency, which is higher for a three-site model than for a four-site model. The accuracy of molecular modeling techniques such as MD or MC is generally limited by two factors: FF quality and sampling.[37] The former issue is addressed in the present work, but the latter is not. The lack of proper sampling is difficult to measure but can have severe consequences on the conclusions drawn from simulations. We therefore prefer to stay with the faster three-site model in order for the FF to be able to be used in simulations covering large systems under long times, such as the receptor binding to membrane proteins and so on, where even biased sampling techniques can be difficult to apply due to slow relaxation times.

Expanded Ensemble Method. The initial free energy calculations during the parametrization procedure were performed using the Bennett acceptance ratio (BAR)[52] which is been well documented in terms of accuracy and efficiency.[4,53] For the validation of the current FF the expanded ensemble (EE) method[54] was used instead. EEMD offers a way to calculate solvation free energies in an efficient manner since they can be obtained from a single simulation, as opposed to, e.g., BAR, which makes the method desirable to use in study with high throughput demands such as this one. As with most insertion methods the solute is gradually inserted/deleted into/from the solvent. A coupling parameter λ is used, where $\lambda = 0$ corresponds to a system where the solute is fully interacting with the solvent and $\lambda = 1$ corresponds to a system where the interactions between the solute and solvent are fully decoupled. The reaction path transferring the system from state $\lambda = 0$ to $\lambda = 1$ should be defined before the simulation and $\lambda \in [0, 1]$. As the simulation proceeds, attempts are made to change the coupling parameter (or subensemble) from λ_i to λ_j with the probability p

$$p(i \rightarrow j) = \min\left(1, \exp\left(-\frac{U(\lambda_j) - U(\lambda_i) - (\eta_j - \eta_i)}{k_B T}\right)\right) \quad (4)$$

where U is the interaction potential between the solvent and the solute, k_B is the Boltzmann constant, and T is the absolute temperature. η are so-called balancing factors which are used to make the distribution over the set of subensembles as close as possible to a uniform one. These balancing factors are updated during the equilibration stage of the simulations according to the Wang−Landau algorithm.[55,56] Namely, after visiting a subensemble, $\{\eta_i\}$ is updated automatically by subtracting a factor $\Delta\eta_i$ from the balancing factor; the initial value of this factor can be taken as $1k_BT$. As an effect of this, the probability of visiting this subensemble is decreased and eventually a uniform distribution between all states will be obtained. During the simulation $\Delta\eta$ is systematically decreased, and once the balancing factors are tuned to give rise to a uniform distribution (this is usually done by setting a criterion for the increment that goes below a predefined value), the balancing factors are fixed and the production run is initiated. From the probabilities of each subensemble the free energy of fully inserting/deleting the solute can be determined according to

$$\Delta G = -k_B T \ln \frac{\rho_M}{\rho_0} + \eta_M - \eta_0 \quad (5)$$

Simulation Details. All simulations were performed in the isothermal−isobaric (*NPT*) ensemble with a temperature of 300 K and a pressure of 1 atm. The temperature was kept constant using the Nosé−Hoover[57,58] thermostat with a

dx.doi.org/10.1021/jp4111234 | J. Phys. Chem. B 2014, 118, 3793−3804

coupling constant of 5 ps, and the pressure was kept constant using the Martyna–Tuckerman–Tobias–Klein barostat[59,60] with a coupling constant of 10 ps. Long-range electrostatic interactions were treated with a particle-mesh Ewald scheme[61,62] with a real space cutoff of 1.0 nm, with a Fourier spacing of 0.12 nm, and a fourth-order interpolation to the Ewald mesh. vdW interactions were treated with a Lennard-Jones potential with a cutoff of 1.0 nm with the standard correction added to the potential of pressure added, which should eliminate the cutoff dependency of the LJ potential. All covalent bonds to hydrogens were constrained using the SHAKE algorithm.[63] A time step of 2 fs was used and the equations of motions were integrated using the velocity Verlet integrator.[64] This setup has been proven to provide the proper thermodynamic ensemble.[65] A soft-core potential[66] was used in order to avoid singularities around the end points of $\{\lambda_i\}$

$$U_{sc} = (1 - \lambda)U_A(r_A) + \lambda U_B(r_B) \quad (6)$$

where

$$r_A = (\alpha\sigma_A{}^6\lambda^p + r^6)^{1/6}; \quad r_B = (\alpha\sigma_A{}^6(1 - \lambda)^p + r^6)^{1/6} \quad (7)$$

U_A and U_B correspond to the regular van der Waals (U_{vdW}) or Coulomb (U_{Coul}) interaction potentials from eq 1 for the thermodynamic states A and B, respectively. In the present study $\alpha = 0.5$ and $p = 1$.

During the free energy calculations the electrostatic and vdW interactions were decoupled separately by employing two sets of coupling parameters, $\{\lambda_{elec}\}$ and $\{\lambda_{vdw}\}$. First the electrostatic interactions are turned off, i.e., $\lambda_{vdw} = 0$ while $\lambda_{elec} \geq 0$, and subsequently the vdW interactions are turned off. For the vdW interactions λ_{vdW} was increased in increments of 0.025, and for the electrostatic interactions λ_{elec} was increased in increments of 0.1; i.e., 49 subensembles were used. Moves in the ensemble space generated by the EE algorithm were attempted with a period of 0.5 ps, and the initial balancing factors were set to $1k_BT$. When $\Delta\eta_i$ reached a value below 0.005 k_BT, the balancing factors were assumed to be converged and the production simulation was initiated. The Metropolis–Gibbs algorithm[67] was employed to perform the moves between the different subensembles.

Simulation boxes containing 768 TIP3P[49] water molecules and one solute were prepared for each studied compound. After the solute had been introduced to the solvent, the energy was minimized using the steepest descent method for 10,000 steps; after this a 250 ps simulation was performed in the canonical ensemble (*NVT*) with the Berendsen thermostat[68] with $\tau = 0.5$. The coordinates obtained from this short simulation were then used for a 500 ps long simulation in the *NPT* ensemble where the temperature and pressure were kept constant according to Berendsen et al.[68] followed by a 2 ns simulation using the coupling algorithms described earlier. The final coordinates were then used for the EEMD simulations which, after optimization of balancing factors, lasted for 100 ns of production simulations. Statistical uncertainty was evaluated by block averaging over 10 ns trajectory fragments and was found to be around 0.1 kJ/mol for most structures (see Table 2) and always within 0.2 kJ/mol. All MD simulations were performed using Gromacs (v. 4.6).[69,70]

Partial atomic charges for the solutes were computed by fitting the electrostatic potential (ESP) obtained at a HF/6-31(d) level of theory. First the geometries of the solutes were optimized using density functional theory (DFT) with the B3LYP exchange-correlation functional[71−74] with the 6-311G-(d,p) basis set. The ESPs were sampled at 10 concentric layers around each atom of the solute with roughly 2500 points/atom. The ESPs were generated using the Gaussian09 software package,[75] and the ESP was fitted with the RESP scheme using the RED software package.[76]

RESULTS AND DISCUSSION

Solvation Free Energies for Model Compounds and Final Parameters. The results of the initial parametrization for the model compounds are presented in Table 2. The agreement between the calculated values and the experimental values is good, as expected. Generally the number of iterations of the thermodynamic cycle illustrated in Figure 1 was between five and 10 depending on the solute. For the hydrophilic ones a larger number of iterations was necessary because changing, e.g., σ^{X-O_W} has a large impact on ΔG_{hyd} due to the electrostatics. Once the difference between the simulations and experiments was decreased to 3%, a proper free energy calculation was performed in order to verify the newly obtained parameters. All values reported in Table 2 are taken from these final simulations. In Table 3 the final parameters for Lennard-Jones cross-interactions are presented together with the original GAFF parameters, and as can be seen, it is somewhat difficult to make out any trend in terms of if the σ or ϵ values should be increased or decreased compared to the Lorentz–Berthelot rules. For most polar species a better agreement between the simulations and experiments was reached if σ was made smaller compared to the original parameters. For ϵ the trend is not as evident, but as already mentioned, the free energy of hydration is highly sensitive to the electrostatics, so this is expected.

Table 2. Free Energies of Hydration for the Model Compounds after Optomization of the Lennard-Jones Cross-Interaction Parameters Compared to the Experimental Values Compiled by Shivakumar et al.[77] (All Values in kilojoules per mole)

model compd	computed	experimental
propane	8.15 ± 0.08	8.2
toluene	−3.61 ± 0.12	−3.7
ethyl methyl sulfide	−6.33 ± 0.09	−6.3
methanol	−21.53 ± 0.11	−21.4
acetone	−16.38 ± 0.10	−16.1
methylamine	−18.83 ± 0.12	−19.1
dimethyl ether	−11.64 ± 0.09	−11.1
acetalaldehyde	−14.84 ± 0.13	−14.7
acetamide	−40.04 ± 0.16	−40.7
methylacetamide	−41.67 ± 0.15	−41.9
dimethylacetamide	−35.00 ± 0.12	−35.6
methyl ester	−13.64 ± 0.09	−13.9
ethylene	5.20 ± 0.05	5.3
nitroethane	−15.51 ± 0.07	−15.5
methyl chloride	−2.56 ± 0.04	−2.3
methyl flouride	−0.95 ± 0.05	−0.9
bromomethane	−3.35 ± 0.05	−3.4
pyridine	−19.14 ± 0.10	−19.7
methyl ester	−13.64 ± 0.09	−13.9

Validation Set. In order to validate the transferability of the newly derived parameters, we calculated the free energy of hydration for the test set comprising 239 molecules assembled by Shivakumar et al.,[43] which has been used to verify previous

dx.doi.org/10.1021/jp4111234 | J. Phys. Chem. B 2014, 118, 3793−3804

Table 3. Final Lennard-Jones Parameters for the Explicit Atomic Pairs between the Given Atom Type and the Oxygen Atom of Water

	modified GAFF		original GAFF	
type	σ (nm)	ϵ (kJ mol^{-1})	σ (nm)	ϵ (kJ mol^{-1})
CT	0.290000	0.600000	0.327514	0.539716
HC	0.260000	0.270000	0.290007	0.204459
CA	0.340000	0.454000	0.327514	0.478526
HA	0.280000	0.190000	0.287513	0.199849
S	0.290000	1.030000	0.335710	0.815880
H1	0.260000	0.291400	0.281098	0.204459
OH	0.309500	0.772400	0.310854	0.748478
C	0.327500	0.470500	0.327514	0.478526
O	0.310500	0.747700	0.305527	0.747766
N3	0.324000	0.570000	0.320031	0.672792
NR3[a]	0.260000	0.960000	0.320031	0.672792[a]
OS	0.280000	0.660000	0.307531	0.672792
H4	0.260000	0.291400	0.283058	0.199849
N	0.319000	0.690800	0.307531	0.672792
NX2[a]	0.285000	0.730800	0.307531	0.672792
OX2[a]	0.303500	0.650500	0.305527	0.747766
NX3[a]	0.290000	0.790800	0.307531	0.672792
OX3[a]	0.295000	0.650500	0.305527	0.747766
OX[a]	0.345000	0.700000	0.305527	0.747766
CX[a]	0.355400	0.400000	0.327514	0.478526
C2	0.355000	0.584000	0.327514	0.478526
Cl	0.320000	0.880000	0.377583	0.516008
F	0.310000	0.400000	0.313438	0.403015
Br	0.350000	1.159100	0.337492	1.057501
NO	0.330000	0.620800	0.320031	0.672792
O2N[a]	0.341900	0.650000	0.305527	0.747766
NB	0.305000	0.650800	0.320031	0.672792

[a]Atom type not present in the original GAFF; the best possible match (see Table 1) is used for the two rightmost columns.

versions of the OPLS FF[43] and the next generation OPLS FF.[44] In the present investigation a number of the model compounds used are also present in this data set, and in order to not bias the results these entries were ignored. This validation data set consists of several different chemical groups and categories, such as saturated, unsaturated, and cyclic hydrocarbons; aromatic and conjugated systems; and polar functional groups (alcohols, amides, ketones, carboxylic acids, ethers, esters, amines, nitriles, and halogens, etc.), making it highly appropriate for validating the new parameters.

In Figure 2 the correlation between the simulation and experimental data is compared. The overall agreement is high, and there is no clear trend indicating, e.g., a too hydrophobic or hydrophilic FF as the mean error is only 0.29 kJ mol^{-1}. The mean unsigned error (MUE) is 2.17 kJ/mol which is more that twice lower than the MUE reported for the original GAFF parameters set of 4.8 kJ/mol.[43] The smallest deviations are observed for the hydrophobic solutes whereas the spread in values is larger for the solutes of a more hydrophilic nature. In order to identify further weaknesses/strengths of the our parameter set, the validation data set is categorized into different chemical groups such as alkanes, alkenes, and alcohols, etc.

In Table 4 the group based analysis is presented. In accordance to Figure 2, the errors for the more hydrophobic entries (alkanes, alkenes, alkynes, and branched alkanes) are small and for all groups except alkynes the mean errors are

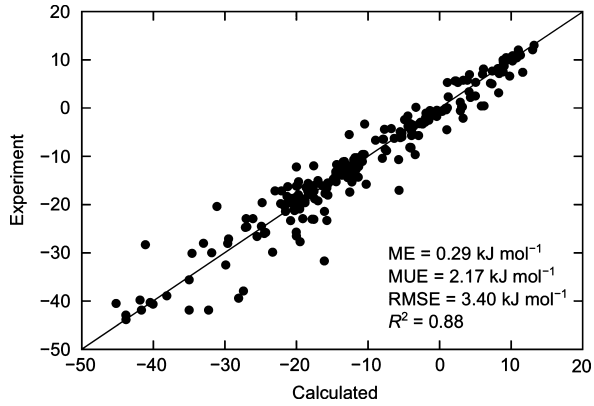


Figure 2. Correlation between simulations and experiments for the validation data set and the corresponding statistics: mean error (ME), mean unsigned error (MUE), root-mean-square error (RMSE), and correlation coefficient (R^2).

Table 4. Errors with Respect to Experimental Data for the Validation Set (kJ mol^{-1})

type	mean error	mean unsigned error	root-mean-square error
alkanes	−0.05	0.65	0.77
alkenes	−1.90	2.39	2.66
alkynes	1.10	1.11	1.15
alcohols	−0.80	2.63	3.37
aldehydes	0.26	0.90	1.29
aliphatic amines	2.15	3.96	4.31
amides	3.42	5.04	4.53
arenes	−0.48	0.80	0.85
aromatic amines	−0.34	2.67	3.93
ketones	0.48	1.11	1.23
bifunctional	−2.40	3.97	5.80
branched alkanes	−0.92	0.87	0.90
carboxylic	0.03	1.61	1.66
nitro	0.50	0.46	0.95
disulfides	0.39	1.10	1.17
esters	−0.94	1.57	2.05
ethers	−3.97	4.02	4.79
halogens (bromo)	1.81	1.81	2.66
halogens (chloro)	0.77	2.32	2.74
halogen (flouro)	0.27	2.24	1.32
halogen (multiple)	3.12	3.46	3.93
nitriles	−4.55	2.16	4.58
cycloalkanes	2.75	2.75	3.07
sulfides	0.67	2.51	2.56
total	0.29	2.17	3.40

negative, indicating a slightly too hydrophilic FF. However, the errors are small for these entries, perhaps with the exception of the alkenes, which has a RMSE which is higher than for the other compounds of this category. A plausible explanation for this could be the model compound used for sp^2 carbons, namely, ethylene. The alkenes in the validation set are all longer, meaning that the electron distribution for the nonterminal sp^2 carbon will differ compared to the terminal one. This charge re-distribution makes it possible for the alkenes that are nonsymmetric to possess a dipole moment which could potentially explain the negative mean error and larger RMSE. The cycloalkanes do not perform as well as their noncyclic counterparts, which indicates that specific atom types might be needed. Following the straightforward parametriza-

dx.doi.org/10.1021/jp4111234 | J. Phys. Chem. B 2014, 118, 3793−3804

tion strategy used here and elsewhere,[35] this should not pose a major problem.

Amines and amides represent a group that have relatively pronounced tendency to show errors. This was also found by Shivakumar et al.[44] when comparing different FFs, but the reason for this is not clear. The ether group, which is a relatively simple chemical moiety, also has a surprisingly large error compared to experiments and the same goes for nitriles. For the latter, introducing new atom types could bring the simulated values closer to the experimental ones, whereas for the former entities the solution does not appear to be as straightforward. For the bifunctional group the RMSE is the largest in Table 4, which is somewhat anticipated. Previously it has been shown that it is significantly more difficult to predict/reproduce ΔG_{hyd} values for polyfunctional solutes.[5−7] That being said, the parameters presented here still perform to an adequate level and beyond when compared to other fixed-charge FFs.[44]

Most popular FFs with fixed atomic charges use one point charge per atom, which is an obvious simplification of the charge distribution as, e.g., lone pairs are neglected this way. The neglect of these oriented dependent charges could be the reason for the observed difficulties for a number of chemical entities here. Another issue is the lack of electronic polarization[78,79] which can be included either in an explicit manner,[80−84] by adding correction terms,[78,79,85−89] or by implicitly including it.[42,56,90] For classical point charge FFs to address these problems major efforts are required, and this is indeed the future of the field; however, as issues regarding sampling are becoming more and more evident,[8,19,21,25,37,91] we believe that these simplified models will be a necessity for the nearest future. Also, for scanning large databases of compounds the simpler models together with the EE method will be highly regarded and necessary.

SAMPL Blind Tests. The SAMPL blind tests[5,92−94] have proposed a way for computational chemists to put their methodology and various parameters to the test. The compounds that are chosen for these blind tests often include several functional groups, and most of these are either drugs or druglike compounds, making these available data sets an excellent test for the FF parameters derived here. Therefore we set out to test our parameters against these published values.

SAMPL0. In Figure 3a the correlations between the computed and experimental[5] ΔG_{hyd} are shown. This data set is relatively small but contains a number of functional groups with esters being the quantitively largest. The general trend is that the more hydrophilic entries are tending to be slightly too hydrophobic, which also is reflected by the mean error. The largest errors are observed for ethylene glycoldiacetate (9.33 kJ mol^{-1}) and 1,4-dioxane (−7.94 kJ mol^{-1}) followed by trimethylbenzamide (−5.25 kJ mol^{-1}). The large disagreement for the diacetate is somewhat surprising since for the initial validation set the parameters derived here performed relatively well for compounds with similar functional groups. On the basis of previous observations, the discrepancy between simulations and experiments for the benzamide and cyclic ether is expected, and this further verifies the potential difficulties with compounds that possess functional groups with lone pairs. For the halogen substituted benzenes the agreement is, however, excellent.

Despite the differences previously discussed, the agreement with the experimental data is good and the RMSE of 4.18 kJ mol^{-1} is lower than the errors presented by Nicholls et al.[5] who used the original GAFF with various partial atomic charge

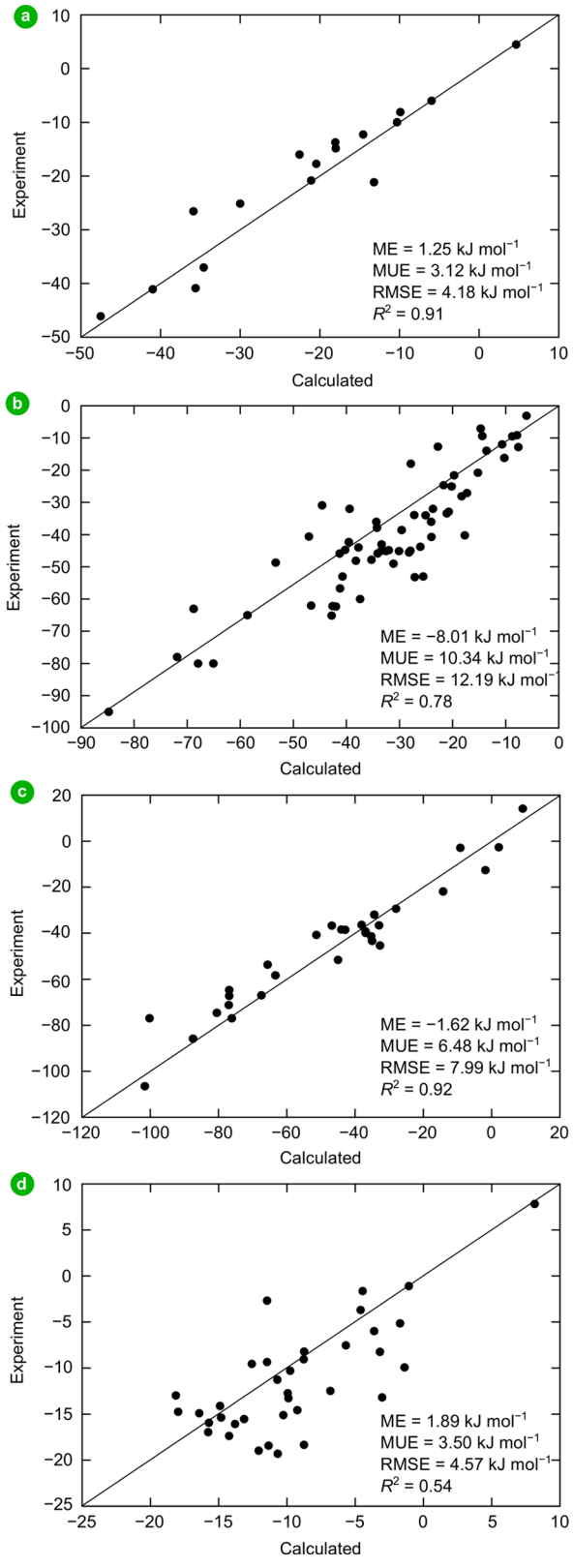


Figure 3. Computed hydration free energies versus experimental values from the (a) SAMPL0, (b) SAMPL1, (c) SAMPL2, and (d) SAMPL3 blind tests.

models and obtained RMSE between 5.5 and 11 kJ/mol for different molecular sets. The major benefit of moving from the previous validation set with molecules of lower molecular weight to a set such as this with larger solutes is that effects that

dx.doi.org/10.1021/jp4111234 | J. Phys. Chem. B 2014, 118, 3793−3804

might have been neglected for the smaller solutes are highlighted. It would be difficult to address these issues following the parametrization scheme used here, and it is likely that the potential energy form instead of the parameters of the FF is unable to properly describe the hydration free energy. One possibility could be the intramolecular parameters, as deficiencies in these should be more pronounced for larger solutes. This was however not investigated as it would be outside the scope of the present investigation. However, there is only a limited amount of comparisons that can be made with this relatively small data set, which should be kept in mind.

SAMPL1. The 63 compounds that make out the SAMPL1 challenge set[92] show a far wider variety in terms of functional groups, and most entries have more than two functional groups. Consequently the results from various researchers showed larger errors than for the previous set. In Figure 3b the results for the modified GAFF are compared to experiments. The results are in worse agreement compared to the previous test, and it is clear that the parameters presented here tend to give solutes that are too hydrophilic. The largest errors are observed for the compounds containing phosphorus, often in combination with sulfur. Naturally this is somewhat expected as no phosphorus compounds were used in the original training set. However, one of the benefits of employing this parametrization scheme is that further development and expansions are relatively straightforward, provided that reliable experimental data are available. For the nitroglycol derivatives the agreement is not excellent albeit the good agreement between the simulations and experiments for the nitro compounds in the initial verification set. Again, one possibility would be to introduce new atom types for these functional groups.

As discussed by Guthrie,[92] for some of the compounds for which it appears to be difficult to predict or reproduce experimental data the problem might be inherent in the experimental data, not in the modeling procedure. Some of the entries in the SAMPL1 blind test have large uncertainties (>8 kJ mol^{-1}) due the lack of experimental data. Also, measuring the vapor pressure or solubility of compounds of strong hydrophilic nature is a demanding experimental task. However, this is only the case for a limited number of entries.

SAMPL2. As for the SAMPL1 compound set, the results for the SAMPL2 set[93] show that the parameters are slightly too hydrophilic (Figure 3c), but not to the same extent. Several entities in this set are difficult to model due to slowly relaxing internal degrees of freedoms, such as the conformational change of the hydrogen of a carboxylic acid group, which goes from a cis to a trans conformation with respect to the carbonyl oxygen when being transferred from gaseous to aqueous phase. This conformational change is associated with a significant activation barrier meaning that the simulations need to be extremely long in order to observe this. If this degree of freedom is not properly sampled, the corresponding solvation free energy would not represent the true free energy associated with the model Hamiltonian or FF.[8,91] It is possible to perform EEMD simulations where the Wang−Landau algorithm is used not only to explore the ensemble or λ-space, but also, e.g., the torsional angle in question.[91] Another solution is to perform the EEMD simulations with one conformation and then add the proper correction to the obtained free energy according to the confine-and-release framework.[8,95] In the present investigation the latter method was used by performing umbrella sampling simulations where the torsional angle in question was scanned with the solute in either gaseous or aqueous phase.

Compounds that were difficult were cyanuric acid, caffeine, 5-triflouromethyluracil, ketoprofen, and trimethylphosphate. These observations are in agreement with the ones reported by Geballe et al.,[93] suggesting again that the (fixed) point charge FFs could have issues with these compounds that have a large fraction of electronegative atoms in close vicinity. The point charges might be able to reproduce the (enhanced) dipole moment of the solute compared to *ab initio* calculations, but the point charges might not do so well for the higher electrostatic moments. Further, the issue of how well these charge models reproduce the ESP at longer distances than the ones sampled during the fitting remains unclear and a potential issue.

The RMSE of 7.99 kJ mol^{-1} remains lower than the MD submissions made to this challenge, for which the most successful had a RMSE of 10.38 kJ mol^{-1}. However, this RMSE is still not as low as the ones obtained with implicit solvent models based on several molecular conformations.[93] The molecular flexibility could also be an issue; all partial atomic charges presented here are taken from one single optimized conformation in the gas phase. It is likely that an average charge distribution that takes into account the molecular flexibility would perform better for some species here.[56]

SAMPL3. In contrast to the previous tests the focus was on chlorinated compounds with a common carbon skeleton in the fourth blind test challenge.[94] It was found that many models fail for compounds with more chlorines, especially for compounds with more than five chlorines.[94] The performance of our modified FF is presented in Figure 3d. The mean error shows that the parameters are slightly too hydrophobic, but not to any large extent. The mean unsigned error and RMSE appear relatively small compared to results obtained from MD simulations,[96] but given the narrow range of experimental data the errors are still not within satisfactory limits. The errors for the compounds with a larger ratio of chlorines remain relatively small though, compared to the averages presented by Geballe and Guthrie.[94]

It is likely that the parametrization scheme presented here does not quite present transferable parameters for compounds with a large number of chlorines and hence the errors are not negligible, despite being relatively small. Another possibility is the polarization effects, which are not treated by classical, fixed-charge FFs. Polarizable models are on the rise within the molecular modeling society,[97−101] and methods where corrections can be made due to the lack of the explicit polarization have been proposed.[56,78,79,87−90,102] For halogens, polarization effects are important[22] and it is possible that abandoning the simple point charges is a necessity in order to bring the agreement between simulations and experiments. It is also possible that a different parametrization philosophy might be more successful for these compounds, but this is not within the scope of the present work.

Impact of Water Model. In order to investigate whether the correction factors a_{ij} and b_{ij} for the solutes and the oxygen of the TIP3P water model are transferable to other water models, the hydration free energies of the amino acid side analogues were computed. The side chain analogues of 14 neutral amino acids were prepared the same as described by Shirts and Pande;[3] GAFF was used for the model in every case with the modified pair interactions introduced here between the solutes and the oxygen of each water model. The partial atomic charges were computed as described in Methods and Models. The amino acid side chain analogues are listed in Table

dx.doi.org/10.1021/jp4111234 | J. Phys. Chem. B 2014, 118, 3793−3804

5. Besides the TIP3P water the following models were tested: SPC,[103] SPC/E,[104] TIP4P,[49] and TIP4P-Ew.[50] The workflow of the simulations was identical to the one described previously.

Table 5. Amino Acids and Their Corresponding Analogues Used in the Present Study

amino acid	analogue
Ala	methane
Val	propane
Ile	*n*-butane
Leu	isobutane
Ser	methanol
Thr	ethanol
Phe	toluene
Tyr	*p*-cresol
Cys	methanethiol
Met	methyl ethyl sulfide
Asn	acetamide
Trp	3-methylindole
His	4-methylimidazole
Hie	4-methylimidazole

The results are presented in the form of a histogram in Figure 4, and it is clear that the three-site models perform more or less equivalently, with a slight favor to TIP3P for obvious reasons. For the four-site models the discrepancy is slightly larger; however, the errors for all of the here reported models are smaller compared to, e.g., the values reported by Shirts and Pande.[3] The largest differences between the three- and four-site models can be seen for the hydrophobic side chains Leu, Ile, and Val. For SPC/E the difference here is also a bit big compared to TIP3P/SPC. The reason for this could in the hydration enthalpy/entropy; Hess and van der Vegt[102] studied these properties and found differences between all water models. For example they found that TIP3P and SPC tend to underestimate the water structuring around hydrophobic solutes whereas TIP4P-Ew showed tendencies toward a too ordered water structuring. These effects are inherent in the water model and do not rely on the solute−solvent interactions to any greater extent. Despite these differences it is clear that the parameters presented here provide good agreement in terms of the actual hydration free energies and that they are relatively transferable between water models. However, we recommend to stay true to the original TIP3P water model as this test is relatively limited in terms of size. It should also be mentioned that these solute−solvent interactions most properly are not ready to be used for large biomolecules such as proteins since the initial hydrophilic/hydrophobic balance of GAFF has been shifted. Nerenberg et al.[35] showed the pitfalls of using a model where this balance has been shifted. Future improvements could be applied to address this point.

Impact of Partial Atomic Charges. As the transferability between different popular water models was more or less established, we investigated the sensitivity of the parameters to the partial atomic charges. Often these parameters have been reported to be the most crucial in terms of computing free energies.[13,42,56,77,106] Besides the original HF/6-31(d) charges, three ESPs were computed for each amino acid side chain analogue using different DFT approaches, namely, at the B3LYP/6-31G(d), B3LYP/cc-pVTZ, and B3LYP/aug-cc-pVTZ levels of theory. This set was then extended by implicitly including the effects of polarization by placing the amino acid side chain in a polarizable continuum, modeled with the IEFPCM model[107,108] with a dielectric constant of 78.4 in order to mimic water. Also the AM1-BCC charges,[109,110] which are parametrized to reproduce the ESP of a HF/6-31G(d) wave function, were used.

In Figure 5 the resulting hydration free energies for each charge model is compared to experimental data.[105] For the hydrophobic solutes the difference is small between the charge models, which goes along one's intuition as the electrostatics should not play a major part in the hydration of these. There are differences between the methods in that they are all smaller than $1k_BT$. For the most hydrophilic side chains (Gln, Asn, Hid, and Hie), the differences are larger and the implicitly polarized charges give the most hydrophilic solutes. For the two larger basis sets, cc-pVTZ and aug-cc-pVTZ the inclusion of implicit polarization increases the errors, whereas, with the smaller Pople set, 6-31G(d), the error is actually decreased once polarization effects are included implicitly. This shows that there exists an error cancellation which can be difficult to probe.

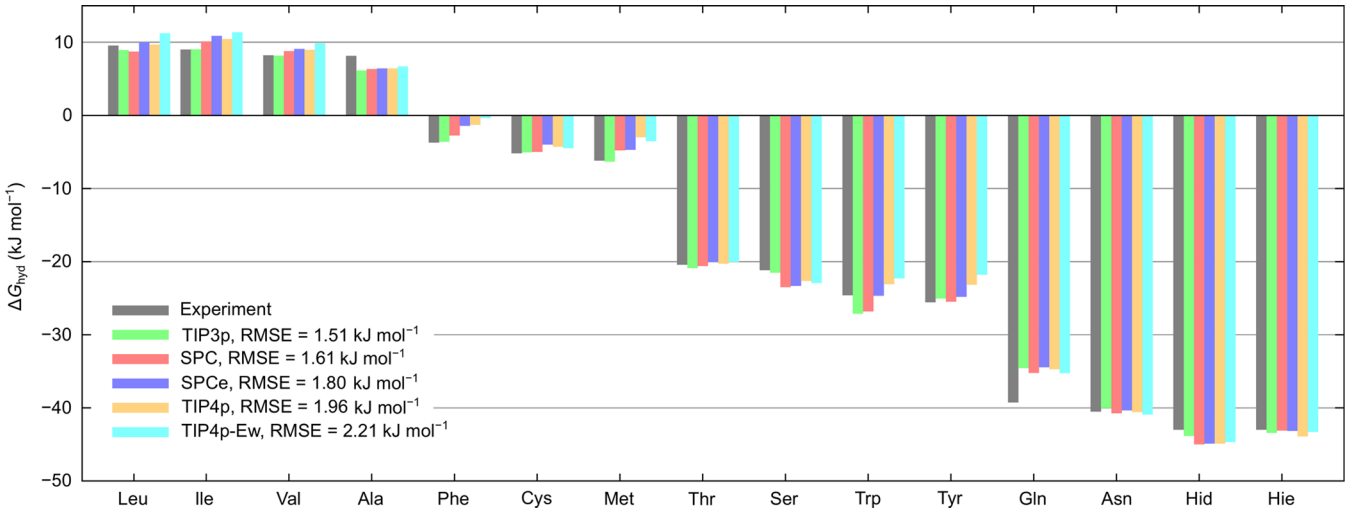


Figure 4. Free energies of hydration for amino acid analogues with the presented force field with different water models. Experimental data are taken from Wolfenden et al.[105]

dx.doi.org/10.1021/jp4111234 | J. Phys. Chem. B 2014, 118, 3793−3804

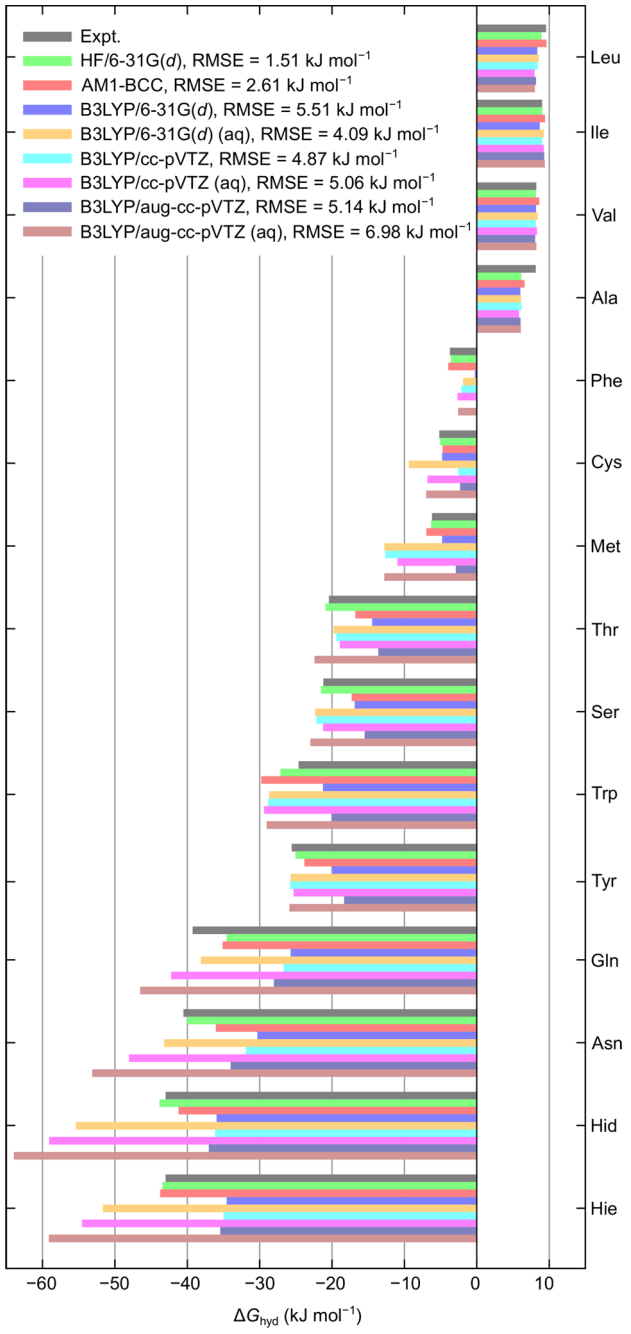


Figure 5. Free energies of hydration for amino acid analogues with the presented force field with different charge models. Atomic charges that were implicitly polarized by applying a self-consistent reaction field with a dieletric constant of 78.4 are denoted by (aq). Experimental data are taken from Wolfenden et al.[105]

Because the parameters presented here have been developed for the original method used to derive charges for GAFF, it is not surprising that the DFT based charges do not perform optimally. The AM1-BCC charges perform better than all of the B3LYP based charges and is not too far off the HF/6-31G(d) charges in performance. With the data presented here it is clear that the charge model is crucial for free energy calculations and that the method that the FF was parametrized against should be used.

CONCLUSIONS

The results of a straightforward reparametrization of the general AMBER FF are presented. A number of simple model compounds were chosen for which the interactions between the constituting atoms of the solute and the solvent were fine-tuned to reproduce experimental hydration free energies. The parameters are transferable and perform well for larger solutes with several functional groups. Because the parametrization only introduces pair specific interactions, no major reparametrization was necessary and the parameters presented here can be used for more than solvation free energy calculations, e.g., to study ligand binding, uptake, and permeability. With the possibility to use a relatively automatized working scheme based on previous work we believe that our parameters will be useful in the development of rational drug design and in other related fields.

AUTHOR INFORMATION

Corresponding Author

*E-mail: alexander.lyubartsev@mmk.su.se. Phone: +46(0) 8161193.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research has been supported by the Swedish Research Council, Grant No. 621-2010-5005. The simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC-003-11-3), at the PDC Center for High Performance Computing. We thank Michael R. Shirts for the assistance with the Expanded Ensemble code in Gromacs.

REFERENCES

(1) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(2) 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*, 5740−5762.

(3) Shirts, M. R.; Pande, V. S. Solvation Free Energies of Amino Acid Side Chain Analogs for Common Molecular Mechanics Water Models. *J. Chem. Phys.* **2005**, *122*, 134508.

(4) Shirts, M. R.; Pande, V. S. Comparison of Efficiency and Bias of Free Energies Computed by Exponential Averaging, the Bennett Acceptance Ratio, and Thermodynamic Integration. *J. Chem. Phys.* **2005**, *122*, 144107.

(5) 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*, 769−778.

(6) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Dill, K. A. Predictions of Hydration Free Energies from All-Atom Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 4533−4537.

(7) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. D.; 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*, 350−358.

(8) Klimovich, P. V.; Mobley, D. L. Predicting Hydration Free Energies using All-atom Molecular Dynamics Simulations and Multiple Starting Conformations. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 307−316.

(9) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical Free Energy Methods for Drug

Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150−160.
(10) Mobley, D. L.; Klimovich, P. V. Perspective: Alchemical Free Energy Calculations for Drug Discovery. *J. Chem. Phys.* **2012**, *137*, 230901.
(11) Warshel, A. Calculations of Enzymic Reactions: Calculations of pKa, Proton Transfer Reactions, and General Acid Catalysis Reactions in Enzymes. *Biochemistry* **1981**, *20*, 3167−3177.
(12) Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and their Uses. *Chem. Rev.* **1971**, *71*, 525−616.
(13) Duffy, E. M.; Jorgensen, W. L. Prediction of Properties from Simulations: Free Energies of Solvation in Hexadecane, Octanol, and Water. *J. Am. Chem. Soc.* **2000**, *122*, 2878−2888.
(14) Tetko, I. V.; Tanchuk, V. Y.; Mark, A. E. Prediction of *n*-Octanol/Water Partition Coefficients from PHYSPROP Database using Artificial Neural Networks and E-state Indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1407−1421.
(15) MacCallum, J. L.; Tieleman, D. P. Calculation of the Water−Cyclohexane Transfer Free Energies of Neutral Amino Acid Side-Chain Analogs using the OPLS All-Atom Force Field. *J. Comput. Chem.* **2003**, *24*, 1930−1935.
(16) Tieleman, D. P.; MacCullum, J. L.; Ash, W.; Kandt, C.; Xu, Z.; Monticelli, T. Membrane protein simulations with a united-atom lipid and all-atom protein model: Lipid-protein interactions, side chain transfer free energies and model proteins. *J. Phys.: Condens. Matter* **2006**, *18*, S1221−S1234.
(17) MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P. Partitioning of Amino Acid Side Chains into Lipid Bilayers: Results from Computer Simulations and Comparison to Experiment. *J. Gen. Physiol.* **2007**, *129*, 371−377.
(18) MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P. Distribution of Amino Acids in a Lipid Bilayer from Computer Simulations. *Biophys. J.*. **2008**, *94*, 3393−3404.
(19) Neale, C.; Bennett, W. F. D.; Tieleman, D. P.; Pomès, R. Statistical Convergence of Equilibrium Properties in Simulations of Molecular Solutes Embedded in Lipid Bilayers. *J. Chem. Theory Comput.* **2011**, *7*, 4175−4188.
(20) Berka, K.; Hendrychová, T.; Anzenbacher, P.; Otyepka, M. Membrane Position of Ibuprofen Agrees with Suggested Access Path Entrance to Cytochrome P450 2C9 Active Site. *J. Phys. Chem. A* **2011**, *115*, 11248−11255.
(21) Paloncýová, M.; Berka, K.; Otyepka, M. Convergence of Free Energy Profile of Coumarin in Lipid Bilayer. *J. Chem. Theory Comput.* **2012**, *8*, 1200−1211.
(22) Vorobyov, I.; Bennett, W. F. D.; Tieleman, D. P.; Allen, T. W.; Noskov, S. The Role of Atomic Polarization in the Thermodynamics of Chloroform Partitioning to Lipid Bilayers. *J. Chem. Theory Comput.* **2012**, *8*, 618−628.
(23) Košinová, P.; Berka, K.; Wykes, M.; Otyepka, M.; Trouillas, P. Positioning of Antioxidant Quercetin and Its Metabolites in Lipid Bilayer Membranes: Implication for Their Lipid-Peroxidation Inhibition. *J. Phys. Chem. B* **2012**, *116*, 1309−1318.
(24) Wennberg, C. L.; van der Spoel, D.; Hub, J. S. Large Influence of Cholesterol on Solute Partitioning into Lipid Membranes. *J. Am. Chem. Soc.* **2012**, *134*, 5351−5361.
(25) Neale, C.; Madill, C.; Rauscher, S.; Pomès, R. Accelerating Convergence in Molecular Dynamics Simulations of Solutes in Lipid Membranes by Conducting a Random Walk along the Bilayer Normal. *J. Chem. Theory Comput.* **2013**, *9*, 3686−3703.
(26) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmayer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(27) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; et al. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(28) MacKerell, A. D.; Bashford, D.; Bellott; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(29) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(30) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.
(31) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(32) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(33) 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.
(34) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.
(35) Nerenberg, P. S.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T. Optimizing Solute-Water van der Waals Interactions To Reproduce Solvation Free Energies. *J. Phys. Chem. B* **2012**, *116*, 4524−4534.
(36) Gumbart, J. C.; Roux, B.; Chipot, C. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *J. Chem. Theory Comput.* **2013**, *9*, 794−802.
(37) Mobley, D. L. Lets Get Honest about Sampling. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 93−95.
(38) Klimovich, P. V.; Mobley, D. L. Predicting Hydration Free Energies using All-atom Molecular Dynamics Simulations and Multiple Starting Conformations. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 307−316.
(39) Paluch, A. S.; Mobley, D. L.; Maginn, E. J. Small Molecule Solvation Free Energy: Enhanced Conformational Sampling Using Expanded Ensemble Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2011**, *7*, 2910−2918.
(40) Jämbeck, J. P. M.; Lyubartsev, A. P. Exploring the Free Energy Landscape of Solutes Embedded in Lipid Bilayers. *J. Phys. Chem. Lett.* **2013**, *4*, 1781−1787.
(41) Wang, J.; Wang, W.; Kollman, P.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.
(42) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent. *J. Phys. Chem. B* **2007**, *111*, 2242−2254.
(43) 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*, 1509−1519.
(44) Shivakumar, D.; Harder, E.; Damm, W.; Friesner, R. A.; Sherman, W. Improving the Prediction of Absolute Solvation Free Energies using the Next Generation OPLS Force Field. *J. Chem. Theory Comput.* **2012**, *8*, 2553−2558.
(45) Wang, J.; Hou, T. Application of Molecular Dynamics Simulations in Molecular Property Prediction. 1. Density and Heat of Vaporization. *J. Chem. Theory Comput.* **2011**, *7*, 2151−2165.
(46) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension,

Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61−74.
(47) Burger, S. K.; Cisneros, G. A. Efficient Optimization of Van der Waals Parameters from Bulk Properties. *J. Comput. Chem.* **2013**, *34*, 2313−2319.
(48) Baker, C. M.; Lopes, P. E. M.; Zhu, X.; Roux, B.; MacKerell, A. D., Jr. Accurate Calculation of Hydration Free Energies using the CHARMM Drude Polarizable Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 1181−1198.
(49) 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*, 926−935.
(50) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.
(51) Abascal, J. L. F.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.
(52) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(53) Pailwal, H.; Shirts, M. R. A Benchmark Test Set for Alchemical Free Energy Transformations and Its Use to Quantify Error in Common Free Energy Methods. *J. Chem. Theory Comput.* **2011**, *7*, 4115−4134.
(54) Lyubartsev, A. P.; Martsinovskii, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. New Approach to Monte Carlo Calculation of the Free Energy: Method of Expanded Ensembles. *J. Chem. Phys.* **1992**, *96*, 1776−1778.
(55) Wang, F.; Landau, D. P. Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053.
(56) Jämbeck, J. P. M.; Mocci, F.; Lyubartsev, A. P.; Laaksonen, A. Partial Atomic Charges and Their Impact on the Free Energy of Solvation. *J. Comput. Chem.* **2013**, *34*, 187−197.
(57) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511−519.
(58) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(59) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit Reversible Integrators for Extended System Dynamics. *Mol. Phys.* **1996**, *87*, 1117−1157.
(60) Tuckerman, M. E.; Alejandre, J.; López-Rendón, R.; Jochim, A. L.; Martyna, G. J. A Liouville-Operator Derived Measure-Preserving Integrator for Molecular Dynamics Simulations in the Isothermal−Isobaric Ensemble. *J. Phys. A: Math. Theor.* **2006**, *39*, 5629−5651.
(61) 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*, 10089−10092.
(62) Essmann, U.; Perera, L.; Berkowitz, T.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(63) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(64) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Applications to Small Water Clusters. *J. Chem. Phys.* **1982**, *76*, 637−649.
(65) Shirts, M. R. Simple Quantitative Tests to Validate Sampling from Thermodynamic Ensembles. *J. Chem. Theory Comput.* **2013**, *9*, 909−926.
(66) Beutler, T. C.; Mark, A. E.; van Scahik, R. C.; Greber, P. R.; van Gunsteren, W. F. Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.* **1994**, *222*, 529−539.
(67) Chodera, J. D.; Shirts, M. R. Replica Exchange and Expanded Ensemble Simulations as Gibbs Sampling: Simple Improvements for Enhanced Mixing. *J. Chem. Phys.* **2011**, *135*, 194110.
(68) 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*, 3684−3690.
(69) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(70) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(71) Becke, A. D. Density functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(72) Lee, C.; Wang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.
(73) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: A critical analysis. *Can. J. Phys.* **1980**, *58*, 1200−1211.
(74) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(75) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennuci, B.; Peterson, G. A.; et al. *Gaussian09*, Revision A.02; Gaussian: Wallingford, CT, USA, 2009.
(76) Dupradeau, F. Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. tools: Adnvances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821−7839.
(77) Shivakumar, D.; Deng, Y.; Roux, B. Computations of Absolute Solvation Free Energies of Small Molecules Using Explicit and Implicit Solvent Model. *J. Chem. Theory Comput.* **2009**, *5*, 919−930.
(78) Swope, W. C.; Horn, H. W.; Rice, J. E. Accounting for Polarization Cost When Using Fixed Charge Force Fields. I. Method for Computing Energy. *J. Phys. Chem. B* **2010**, *114*, 8621−8630.
(79) Swope, W. C.; Horn, H. W.; Rice, J. E. Accounting for Polarization Cost When Using Fixed Charge Force Fields. II. Method and Application for Computing Effect of Polarization Cost on Free Energy of Hydration. *J. Phys. Chem. B* **2010**, *114*, 8631−8645.
(80) Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D., Jr. Polarizable empirical force field for alkanes based on the classical Drude oscillator model. *J. Phys. Chem. B.* **2005**, *109*, 18988−18999.
(81) Soteras, I.; Curuchet, C.; Bidon-Chanal, A.; Dehez, F.; Ángyán, J. G.; Orozco, M.; Chipot, C.; Luque, F. J. Derivation of Distributed Models of Atomic Polarizability for Molecular Simulations. *J. Chem. Theory Comput.* **2007**, *3*, 1901−1913.
(82) Dehez, F.; Ángyán, J. G.; Soteras Gutiérrez, I.; Luque, F. J.; Schulten, K.; Chipot, C. Modeling Induction Phenomena in Intermolecular Interactions with an Ab Initio Force Field. *J. Chem. Theory Comput.* **2007**, *3*, 1914−1926.
(83) Harder, E.; MacKerell, A. D., Jr.; Roux, B. Many-Body Polarization Effects and the Membrane Dipole Potential. *J. Am. Chem. Soc.* **2009**, *131*, 2760−2761.
(84) 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.
(85) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. Progress in *Ab Initio* QM/MM Free-Energy Simulations of Electrostatic Energies in Proteins: Accelerated QM/MM Studies of pK_a, Redox Reactions and Solvation Free Energies. *J. Phys. Chem. B* **2009**, *113*, 1253−1272.
(86) Beierlein, F. R.; Michel, J.; Essex, J. W. A Simple QM/MM Approach for Capturing Polarization Effects in Protein−Ligand Binding Free Energy Calculations. *J. Phys. Chem. B* **2011**, *115*, 4911−4926.

(87) Chipot, C. Rational Determination of Charge Distributions for Free Energy Calculations. *J. Comput. Chem.* **2002**, *24*, 409−415.
(88) Bas, D.; Dorison-Duval, D.; Moreau, S.; Bruneau, P.; Chipot, C. Rational Determination of Transfer Free Energies of Small Drugs across the Water−Oil Interface. *J. Med. Chem.* **2002**, *45*, 151−159.
(89) Jämbeck, J. P. M.; Lyubartsev, A. P. Implicit Inclusion of Atomic Polarization in Modeling of Partitioning between Water and Lipid Bilayers. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4677−4686.
(90) Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A. Derivation of Fixed Partial Charges for Amino Acids Accommodating a Specific Water Model and Implicit Polarization. *J. Phys. Chem. B* **2013**, *117*, 2328−2338.
(91) Paluch, A. S.; Mobley, D. L.; Maginn, E. J. Small Molecule Solvation Free Energy: Enhanced Conformational Sampling Using Expanded Ensemble Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2011**, *7*, 2910−2918.
(92) Guthrie, J. P. A Blind Challenge for Computational Solvation Free Energies: Introduction and Overview. *J. Phys. Chem. B* **2009**, *113*, 4501−4507.
(93) Skillman, A. G.; Geballe, M. T.; Nicholls, A. SAMPL2 challenge: Prediction of solvation energies and tautomer ratios. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 257−258.
(94) Geballe, M. T.; Guthrie, J. P. The SAMPL3 blind prediction challenge: Transfer energy overview. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 489−496.
(95) Mobley, D. L.; Chodera, J. D.; Dill, K. A. Confine-and-Release Method: Obtaining Correct Binding Free Energies in the Presence of Protein Conformational Change. *J. Chem. Theory Comput.* **2007**, *3*, 1231−1235.
(96) Mobley, D. L.; Liu, S.; Cerutti, D. S.; Swope, W. C.; Rice, J. E. Alchemical prediction of hydration free energies for SAMPL. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 551−562.
(97) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933−5947.
(98) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; et al. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.
(99) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143−3161.
(100) Lamoureux, G.; Alexander, D.; MacKerell, J.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *J. Chem. Phys.* **2003**, *119*, 5185−5197.
(101) Yu, W.; Lopes, P. E. M.; Roux, B.; Alexander, D.; MacKerell, J. Six-site polarizable model of water based on the classical Drude oscillator. *J. Chem. Phys.* **2013**, *138*, 034508.
(102) Hess, B.; van der Vegt, N. F. A. Hydration Thermodynamic Properties of Amino Acid Analogues: A Systematic Comparison of Biomolecular Force Fields and Water Models. *J. Phys. Chem. B* **2006**, *110*, 17616−17626.
(103) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel, Dordrecht, The Netherlands, 2008; p 331.
(104) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Chem. Phys.* **1987**, *91*, 6269−6271.
(105) Wolfenden, R.; Andersson, L.; Cullis, P. M.; Southgate, C. C. B. Affinities of Amino Acid Side Chains for Solvent Water. *Biochemistry* **1981**, *20*, 849−855.
(106) Deng, Y.; Roux, B. Hydration of Amino Acid Side Chains: Nonpolar and Electrostatic Contributions Calculated from Staged Molecular Dynamics Free Energy Simulations with Explicit Water Molecules. *J. Phys. Chem. B* **2004**, *108*, 16567−16576.
(107) Tomasi, J.; Mennucci, B.; Cancés. The IEF Version of the PCM Solvation Method: An Overview of a New Method Addressed to Study Molecular Solutes at the QM Ab-Initio Level. *J. Mol. Struct. (THEOCHEM)* **1999**, *464*, 211−226.
(108) Pomelli, C. S.; Tomasi, J.; Barone, V. An Improved Iterative Solution to Solve the Electrostatic Problem in the Polarizable Continuum Model. *Theor. Chem. Acc.* **2001**, *105*, 446−451.
(109) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132−146.
(110) 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*, 1623−1641.