

Original scientific paper

GAFF/IPolQ-Mod+LJ-Fit: Optimized force field parameters for solvation free energy predictions

Andreas Mecklenfeld^{1,2} and Gabriele Raabe^{1,2*}

¹Institut für Thermodynamik, Technische Universität Braunschweig, Hans-Sommer Strasse 5, 38106 Braunschweig, Germany

²Center of Pharmaceutical Engineering, Technische Universität Braunschweig, Franz-Liszt-Strasse 35a, 38106 Braunschweig, Germany

*Corresponding Author: E-mail: g.raabe@tu-bs.de; Tel.: +49 (531) 391 2628; Fax: +49 (531) 391 7814

Received: April 29, 2020; Revised: June 18, 2020; Published: June 28, 2020

Abstract

Rational drug design featuring explicit solubility considerations can greatly benefit from molecular dynamics simulations, as they allow for the prediction of the Gibbs free energy of solvation and thus relative solubilities. In our previous work (A. Mecklenfeld, G. Raabe. J. Chem. Theory Comput. **13** no. 12 (2017) 6266–6274), we have compared solvation free energy results obtained with the General Amber Force Field (GAFF) and its default restrained electrostatic potential (RESP) partial charges to those obtained by modified implicitly polarized charges (IPolQ-Mod) for an implicit representation of impactful polarization effects. In this work, we have adapted Lennard-Jones parameters for GAFF atom types in combination with IPolQ-Mod to further improve the accuracies of solvation free energy and liquid density predictions. We thereby focus on prominent atom types in common drugs. For the refitting, 357 respectively 384 systems were considered for free energies and densities and validation was performed for 142 free energies and 100 densities of binary mixtures. By the in-depth comparison of simulation results for default GAFF, GAFF with IPolQ-Mod and our new set of parameters, which we label GAFF/IPolQ-Mod+LJ-Fit, we can clearly highlight the improvements of our new model for the description of both relative solubilities and fluid phase behaviour.

©2020 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords

Molecular dynamics simulations; force field optimization; solvation free energies; relative solubilities

Introduction

Solubility is a crucial thermophysical property in the pharmaceutical industry [1] and its assessment is therefore of utmost importance. Molecular simulations can be interpreted as experiments on the computer and are capable to complement the computer-aided drug design. They allow for the accurate prediction of thermophysical properties to determine suitable solvents for an active pharmaceutical ingredient in an early stage of the drug development process. According to Liu *et al.* [2], the relative solubilities can be determined by Eq. (1):

$$\ln \frac{c_S^A}{c_S^B} = -\beta(\Delta G_{\text{solv}}^A - \Delta G_{\text{solv}}^B) \quad (1)$$

in which c_s is the molar concentration of solute S in solvent A or B, β is the inverse temperature and ΔG_{solv} is the Gibbs free energy of solvation of the solute. The Gibbs free energy of solvation corresponds to the change in the Gibbs free energy for transferring a single solute from a vacuum into a condensed phase [3]. In the vacuum phase, no intermolecular interactions are considered between solute and solvent, whereas they are fully existent in the condensed phase. In order for the Gibbs free energy to converge, the change of state from vacuum to condensed phase is subdivided into a series of intermediate states, each representing an inherent molecular simulation. These intermediates create a linking chain of shared configurational space overlap and are characterized by scaled solute-solvent interactions. As the intermediates often represent non-physical states, they are referred to as alchemical pathway and the scaling factor is the alchemical variable λ . In a previous study [4] we proposed our “pathfinder” method to define a set of λ -states with reduced computational effort but high statistical certainty. However, the quality of solubility predictions is greatly affected by the molecular models that describe the intermolecular interactions.

Several studies compared different molecular models, the so called force fields, for different sets of test systems [5–15]. Widely used molecular models such as the General Amber Force Field (GAFF) [16], the Optimized Potentials for Liquid Simulations model (OPLS-AA) [17] or the CHARMM General Force Field (CGenFF) [18] are applied for the description of drug-like molecules and are categorized as “Class 1” force fields. This class of models is characterized by the usage of fixed partial charges located on the atom’s center of mass for the evaluation of electrostatics by Coulomb’s law. Class 1 force fields represent a moderate computational effort, which makes them applicable for systems with high time constants or simulation techniques such as free energy calculations, which require increased effort due to the multitude of intermediate λ -states. A significant disadvantage is the lacking ability to represent polarization effects. This means that physical meaningful phenomena, like for the transition of a solute from a vacuum into a condensed phase, cannot be described adequately. Although polarizable models [19–30] could be used, these are considered computational expensive [31], while not necessarily more accurate than Class 1 force fields due to the more demanding parametrization [29]. For an at least implicit representation of polarization effects for solvation free energy calculations, Cerutti *et al.* [32] developed the IPolQ („implicitly polarized charges“) method, which was later modified by Muddana *et al.* [11] and referred to as IPolQ-Mod. In an extensive study [6], we compared solvation free energy results obtained with GAFF and its default two-stage RESP (“restrained electrostatic potential”) [33] partial charge scheme, and GAFF with IPolQ-Mod partial charges. We concluded a general compatibility of the GAFF model and the IPolQ-Mod method, though we highlighted shortcomings for specific compound classes due to the disturbed self-consistency of the molecular model. As a consequence, we recommended the refitting of the atom type specific parameters ϵ and σ of the Lennard-Jones (LJ) potential:

$$U_{\text{LJ}} = 4 \cdot \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (2)$$

with r_{ij} being the distance between atoms i and j . In this study we present our methodology to optimize a large number of Lennard-Jones parameters based on GAFF for an improved representation of densities and particularly solvation free energies. We thereby consider a large number and diversity of systems and optimized atom types. Our aim is to provide model parameters to restore the self-consistency of the GAFF model in combination with IPolQ-Mod charges, as stressed in our previous work [6]. In detailed analyses we compare our newly developed model parameters with results from the standard GAFF model with both standard RESP and IPolQ-Mod partial charges for different data sets. We conclude this manuscript with a summary of our findings.

Methodology

Targeted functional groups, atom types and data sets

The choice of atom types for the re-parametrization was based on the analysis of substance groups occurring in potential active pharmaceutical ingredients or relevant solvents [34]. This includes alkanes, alkenes, alkynes, cycloalkanes, arenes, azoles, azines, amides, nitriles, aldehydes, ketones, alcohols, phenols, amines, ethers as well as haloalkanes, i.e. compounds with bonded fluorine, chlorine, bromine and iodine atoms. GAFF includes a variety of atom types that feature identical Lennard-Jones, but different bonded parameters. As the chemical environments for these types are comparable, no distinction was made in the re-parametrization. Furthermore, only non-hydrogen atom types were considered. Table 1 summarizes the GAFF atom types included in the refitting of the LJ-parameters.

Table 1. Standard GAFF atom types targeted at the parameter optimization.

Atom type	Description [16]
br	any bromine
c	sp ² carbon in C=O
c1 (cg)	sp ¹ carbon (in conjugated ring systems)
c2	sp ² carbon, aliphatic
c3	sp ³ carbon
ca (cc / cd / ce)	sp ² carbon, aromatic / conjugated
cl	any chlorine
f	any fluorine
i	any iodine
n	sp ² nitrogen in amides
n1	sp ¹ nitrogen
na	sp ² nitrogen with 3 subst.
nb (n2)	aromatic nitrogen / sp ² nitrogen with 2 subst.
nh (n3)	amine nitrogen / sp ³ nitrogen with 3 subst.
o	sp ² oxygen in C=O
oh	sp ³ oxygen in hydroxyl groups
os	sp ³ oxygen in ethers and esters

Besides solvation free energies, we also chose liquid densities ρ of pure compounds in a broad temperature range as target quantities in order to allow for accurate solubility predictions as well as for precise descriptions of liquid bulk phases. The data set considered in the refitting process consists of 357 solvation free energy systems and 384 densities for small model compounds. For validation, three additional data sets were used. Set I and II cover 100 solvation free energy systems for small model compounds respectively 42 solvation free energy systems with solutes haloperidol, phenacetin, temazepam and trimethoprim. For the latter, relative solubilities were calculated according to Eq. (1) and compared to experimental relative solubility data. In validation data set III, 100 liquid densities of binary mixtures were considered. None of the validation systems were included in the refitting process. The various data sets are summarized in Table 2.

Table 2. Data sets used for fitting respectively validating new Lennard-Jones parameters.

Data set	Content
Refitting	357 ΔG_{solv} -systems: 112 solute / 37 solvent compounds 384 ρ -systems: 78 compounds, $\Delta T = (183.15 \dots 478.15)$ K
Validation I	100 ΔG_{solv} -systems: 59 solute / 34 solvent compounds
Validation II	42 ΔG_{solv} -systems: 4 solute / 23 solvent compounds
Validation III	100 ρ -systems of binary mixtures: 72 compounds, $\Delta T = (183.15 \dots 383.15)$ K

All data sets except for validation II include water as solvent compound. New parameters for solutes dissolved in water were derived and tested both for the TIP3P [35], as well as for the TIP4P/2005 [36] water models. While TIP4P/2005 is considered to be a multi-purpose model for the description of liquid water, default GAFF parameters were derived using TIP3P. As the Lennard-Jones parameters for the water models themselves were unaltered, adaptable interaction parameters ξ_{ij} and ζ_{ij} were used in the Lorentz-Berthelot combining rules, i.e.

$$\epsilon_{ij} = \left(1 + \xi_{ij}\right) \cdot \sqrt{\epsilon_{ii} \cdot \epsilon_{jj}} \quad , \quad (3)$$

and

$$\sigma_{ij} = \left(1 + \zeta_{ij}\right) \frac{\sigma_{ii} + \sigma_{jj}}{2} \quad . \quad (4)$$

Index *i* refers to the atom type of interest, while index *j* represents the oxygen atom type of the corresponding water model. Hydrogen atom types for both water models do not participate in Lennard-Jones interactions. For all but the water interactions, parameters ξ_{ij} and ζ_{ij} were set to zero.

Our main aim in the composition of the data sets was to consider both, diverse compound pairs and a broad temperature range for the density calculations. All simulation results are given as numerical values in the Supporting Information.

Molecular models

The adaptation of Lennard-Jones parameters is based on GAFF Version 1.8. Molecule topologies were generated using Antechamber [37] from AmberTools18 [38], followed by the transfer into the GROMACS [39–45] format using ACPYPE [46]. For comparison, we performed all simulations with default GAFF parameters using RESP [33] partial charges (GAFF/RESP), while the new Lennard-Jones parameters were specifically derived for the IPolQ-Mod method (GAFF/IPolQ-Mod+LJ-Fit). For the fitting as well as the validation sets I and III, we also considered IPolQ-Mod charges with default GAFF parameters (GAFF/IPolQ-Mod) for the sake of comparison. Partial charges were calculated using *ab initio* simulations in Gaussian09 [47]. Therefore, an energy optimization was performed for each system at the HF/6-31G* [48–58] level of theory prior to the partial charge calculation. For GAFF/RESP, the electrostatic potential (ESP) was calculated at HF/6-31G*, and partial charges were fitted according to the two-stage restrained electrostatic potential (RESP) [33] to match the ESP. However, this approach cannot be applied to compounds with bonded iodine, as this element is not included in the 6-31G* basis set. Although GAFF should be compatible with partial charges derived by the semi-empirically AM1-BCC [59,60] method, Muddana *et al.* [11] highlighted that results for ΔG_{solv} obtained by using RESP and AM1-BCC partial charges may differ by several kJ/mol. Therefore, i.e. for the sake of consistency, no AM1-BCC charges were applied, and systems including iodine were dismissed from simulations using GAFF/RESP.

For deriving the IPolQ-Mod partial charges, the MP2/aug-cc-pVDZ [61–70] level of theory was used for all ESP calculations as initially proposed by Muddana *et al.* [11]. For the solutes in ΔG_{solv} simulations, two sets of ESP's were calculated. One ESP represents the condensed phase by applying a polarizable continuum model [71] of the solvent (Gaussian keyword: SCRF=PCM), while no continuum model is applied for the representation of the vacuum phase. Partial charges from both ESP calculations were derived by RESP fitting. The charge sets were then averaged for an implicit representation of polarization effects caused by the transition of the solute from the vacuum into the solvent phase. As free energy simulations describe the behavior of the solute in infinite dilution, charges for the solvent compounds were derived in solvent phase only.

For the density calculation of a binary mixture with compounds A and B, the partial charge for an atom i in compound A q_i^A is calculated according to Eq. (5):

$$q_i^A = x_A \cdot q_i^{A,A} + (1 - x_A) \cdot q_i^{A,B}, \quad (5)$$

whereas x_A is the mole fraction of A in the binary mixture. Expressions $q_i^{A,A}$ and $q_i^{A,B}$ are the partial charges for atom i in compound A considering the continuum models for compounds A and B respectively. For compounds including the element iodine, a pseudopotential aug-cc-pVDZ basis set [72,73] from the EMSL Basis Set Library [74,75] was employed.

Objective approach and optimization algorithms

The adaptation of Lennard-Jones parameters has been widely discussed [13,15,76,77]. In order to reduce the complexity of the optimization task, pair interactions can be adapted sequentially. As the issue of high computational efforts remains, we have studied the applicability of a thermodynamic cycle approach to obtain accurate free energy results but with significant decreased expenses [78]. This approach is also employed here, when possible.

In our optimization, we target both solvation free energies and liquid densities. While the focus of our work is the accurate prediction of ΔG_{solv} respectively relative solubilities, the ability to describe liquid bulk phases is essential for the transfer of our new model parameters to further applications. Eq. (6) displays the chosen objective function Z^i for optimization step i .

$$Z^i = \frac{1}{w_{\Delta G_{\text{solv}}} + w_{\rho}} \left(w_{\Delta G_{\text{solv}}} \frac{\text{RMSD}_{\Delta G_{\text{solv}}}^{(i)}}{\text{RMSD}_{\Delta G_{\text{solv}}}^{(\text{ref})}} + w_{\rho} \frac{\text{RMSD}_{\rho}^{(i)}}{\text{RMSD}_{\rho}^{(\text{ref})}} \right) \quad (6)$$

The root-mean-square deviations (RMSD) describe the divergence between simulation and experimental data. Given the two target properties ΔG_{solv} and ρ , we normalize the RMSD for ΔG_{solv} and ρ with regards to reference results (ref) obtained from simulations with unaltered model parameters. RMSD ratios are weighted by $w_{\Delta G_{\text{solv}}}$ and w_{ρ} , individually defined to obtain accurate results for both ΔG_{solv} and ρ .

Lennard-Jones parameters ϵ_{ii} and σ_{ii} respectively ξ_{ij} and ζ_{ij} are not explicitly associated with target properties ΔG_{solv} and ρ , as ΔG_{solv} depends on both enthalpic interactions as well as entropic effects between solute and solvent molecules. This requires the application of a 2-dimensional optimization algorithm. A further issue is the occurrence of statistical noise [63]. Optimization algorithms typically evaluate the change of the target function inflicted by small parameter changes. If the change in simulation results and the statistical noise are in the same order of magnitude, the optimization is likely to fail. This especially concerns derivation-based algorithms, as the usually applied difference quotients require particular small parameter changes. As a consequence, Faller *et al.* [79] proposed the usage of the derivative free and robust Downhill Simplex by Nelder and Mead [80], which was used for several subsequent force field optimizations [81–85]. To prevent local optima, optimization cycles have been repeated with differently orientated initial simpli. The optimization procedure including the thermodynamic cycle approach has been implemented into Python for a fully automated workflow.

Simulation details

All molecular simulations were performed using GROMACS 2016.1 or 2018.1. For all simulations, the stochastic dynamics integrator [86] with a time step of $\delta t = 0.5$ fs was employed, which additionally controlled the temperature using an inverse friction constant of $\tau_T = 2.5$ ps. The pressure was set to $p = 1$ atm and adjusted by the Berendsen barostat [87] during equilibration respectively the Parrinello-Rahman

barostat [88] during production phases, each with a time constant of $\tau_p = 5$ ps. The particle mesh Ewald scheme with interpolation order 4, a real space cutoff radius $r_{\text{coulomb}} = 1.2$ nm and a Fourier spacing of 0.12 nm was employed for the calculation of electrostatic interactions. The cutoff radius for Lennard-Jones interactions was set to $r_{\text{vdW}} = 1.2$ nm, and long-range corrections were applied for energy and pressure calculations. System sizes exceed the recommendations proposed by Parameswaran and Mobley [89], and initial configurations were created using PACKMOL [90]. The rigid water models TIP3P and TIP4P/2005 were constrained using the settle algorithm [91]. For all solvation free energy calculations, the 1-1-48 softcore-potential with $\alpha = 0.003$ was used [92,93], while enthalpy differences for the calculation of ΔG_{solv} were written out every 100 steps. For the evaluation of ΔG_{solv} we used the Multistate Bennett Acceptance Ratio (MBAR) method [94] as implemented in the “alchemical_analysis.py” tool [95]. The evaluations for density calculations were performed using the GROMACS “gmx energy” utility, and their uncertainties refer to the standard error of the mean using 5 blocks of equal length.

Further simulation details presented in the following are specific to the target property, and depend on whether the simulation was performed as part of the parameter refitting or validation. For the non-refitting simulations, density calculations were performed from production runs of 4e6 simulation steps after a short energy minimization and an equilibration run of 2e6 steps. Solvation free energy calculations were performed following the workflow of our “pathfinder” method [4]. Given an initial number of λ -states, an energy minimization and an equilibration of 4e5 steps was performed for each state. Following that, the number and distribution of λ -states was adjusted in 3 to 5 trial simulations of 4e5 steps each. The aim of the trial simulations is to obtain equal partial free energy uncertainties for an overall improvement of statistical certainty [96,97], while simultaneously ensuring sufficient configurational space overlap. This is followed by 5 productions runs of 4e6 simulations steps, which were evaluated using the block averaging technique.

For density calculations during the refitting, 4e5 equilibration steps, respectively 3.6e6 production steps were performed. To reduce the equilibration period, initial configurations were used from completed runs of similar model parameters. The number and distribution of λ -states was not adjusted during the fitting but adopted from corresponding GAFF/IPolQ-Mod simulations, although the existence of sufficient configurational space overlap was carefully monitored. Equivalent to the density calculations, initial configurations were taken from previous optimization steps. By this, equilibration could be reduced to 4e5 steps, while for production 4.6e6 to 5.6e6 steps were used depending on the dynamics of the systems. All ΔG_{solv} calculations in the refitting procedure were evaluated using bootstrapping analysis [98]. The simulation protocols were carefully tested to ensure reproducibility of the results.

Procedure of the atom type adaptation

Due to the complexity of the optimization problem, atom types were not adjusted simultaneously, but in a subsequent manner. That means that for a current atom type optimization, only compounds were included that featured already optimized atom types. By this, the number of atom types considered, and thus the complexity of compounds increased during the optimization process. Atom types that occur in compounds with a large volume of reference data were prioritized in the order of their adaptation. For the following atom type adjustments, the number and diversity of the considered components thus quickly increased.

In the following, the procedure of the adaptation will be discussed, referring to atom types presented in Table 1. Figure 1 shows the order of the atom type adaptation.

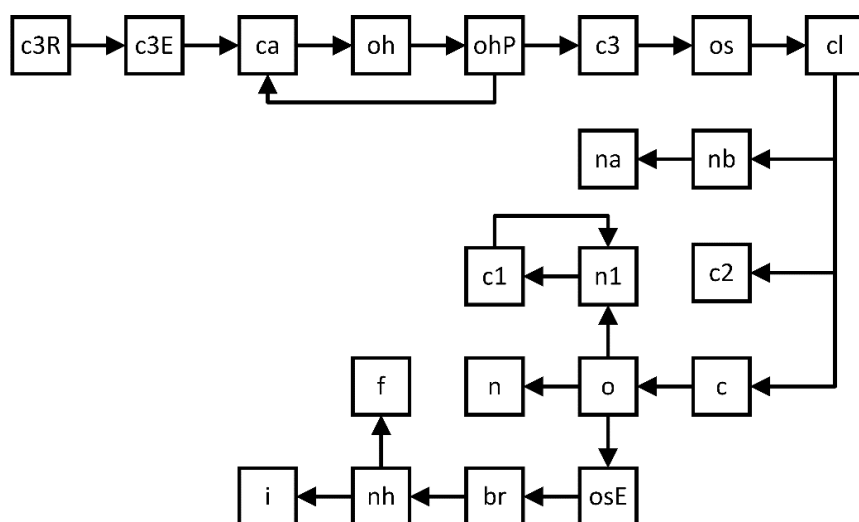


Figure 1. Illustration of the order of the atom type adaptation with newly defined atom types.

At first, we optimized the sp^3 hybridized carbon in ring structures (c3R), followed by the sp^3 hybridized carbon in acyclic hydrocarbons. The aromatic sp^2 hybridized carbon (ca) was adapted in connection with the subsequent oxygen type in alcohols and phenols. Due to the poor performance of ΔG_{solv} predictions for both alcohols and phenols, we decided to differentiate between an oxygen type in alcohols (oh) and phenols (ohP). Because of difficulties in representing the solvent n-octanol, a sp^3 hybridized carbon (c3) was introduced as additional atom type. That is, we differentiate between the sp^3 hybridized carbons of a CH_2 -group within an alkyl chain (c3), the carbons in a CH_3 -group at the end of an alkyl chain (c3E), and the c3R carbon type in ring structures described in the beginning.

After the adaptation of the oxygen type in ethers (os) and the chlorine atom type (cl), atom types for nitrogen in arenes (nb) as well as sp^2 hybridized carbon in alkenes (c2) and $c=O$ structures (c) were optimized independently. After the refitting of type nb, the sp^2 nitrogen in heteroaromatics (na) was modified. The adaptation of types c and o was performed successively for the description of $c=O$ in aldehydes, ketones and esters, the latter featuring a newly defined atom type osE. After that, atom types n in amides and n1 in nitriles were targeted. For nitrile compounds, both characteristic atom types n1 and c1 were adapted successively again. Atom type bromine (br) was adjusted considering the optimized ester parameters, followed by nitrogen in amines (nh). The halogen atom types fluorine (f) and iodine (i) were reparametrized at the end. For each of these atom types, the adjustable interaction parameters ξ_{ij} and ζ_{ij} for the description of interactions between any atom type i and the oxygen atom type in TIP3P respectively TIP4P/2005 were optimized analogously.

Results and discussion

Evaluation of the refitting

Overall, 21 atom types respectively 42 pairs of adaptable coefficients were adjusted in the refitting process. The newly derived parameters are given in the Supporting Information. In Figure 2, simulation results of the free energy of solvation $\Delta G_{solv,simulation}$ for the refitting data set are given for GAFF/RESP, GAFF/IPolQ-Mod and GAFF/IPolQ-Mod+LJ-Fit with respect to experimental data $\Delta G_{solv,experiment}$ [92,93,99,100]. The diagram highlights the broad deviations of the results especially for GAFF/IPolQ-Mod and to a minor degree also for GAFF/RESP, while results for GAFF/IPolQ-Mod+LJ-Fit show significant better agreement with the experimental data.

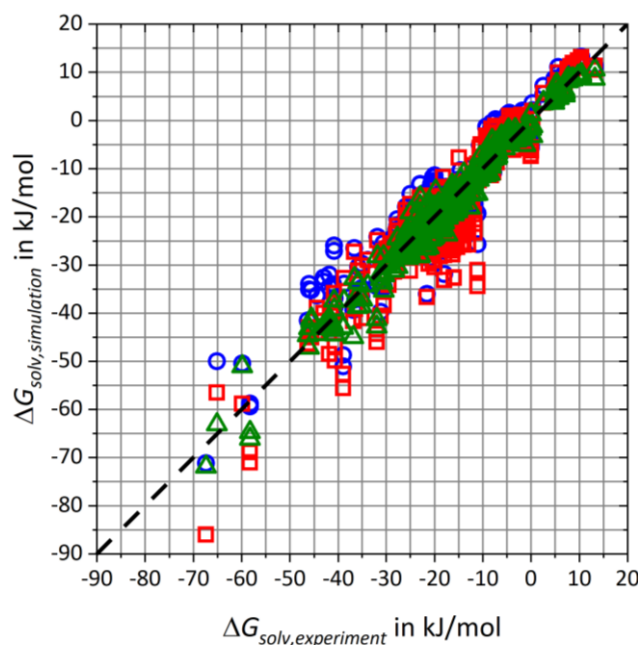


Figure 2. Comparison of solvation free energies $\Delta G_{\text{solv,simulation}}$ vs. experimental data $\Delta G_{\text{solv,experiment}}$ [92,93,99,100] from the refitting data set. The results are represented by blue circles for GAFF/RESP, red squares for GAFF/IPolQ-Mod and green triangles for GAFF/IPolQ-Mod+LJ-Fit.

RMSD and MAE deviations as well as the linear regression fits are summarized in Table 3. The table highlights that the RMSD value of GAFF/IPolQ-Mod+LJ-Fit has been decreased by approximately 2 kJ/mol respectively 3 kJ/mol compared to GAFF/RESP and GAFF/IPolQ-Mod. Furthermore, both the slope m of the linear function and the Pearson coefficient R are closest to 1.

Table 3. Summary of the evaluation of free energy results from the refitting data set. Besides the root-mean-square deviations (RMSD) and the mean absolute errors (MAE), the slopes m of the linear fitting curves with corresponding Pearson correlation coefficients R are given. For a better comparison with GAFF/RESP, additional values for GAFF/IPolQ-Mod and GAFF/IPolQ-Mod+LJ-Fit are stated in brackets that exclude iodine compounds.

	GAFF/RESP	GAFF/IPolQ-Mod	GAFF/IPolQ-Mod+LJ-Fit
RMSD in kJ/mol	4.64	5.39 (5.49)	2.39 (2.43)
MAE in kJ/mol	3.52	3.93 (4.00)	1.77 (1.79)
Slope m	0.9511	1.0452 (1.0397)	1.0083 (1.0085)
Pearson R	0.9438	0.9350 (0.9340)	0.9850 (0.9848)

In Figure 3, RMSD-values are given for solvation free energy results clustered by substance groups of solutes respectively solvents.

The arrangement by solutes in diagram a) demonstrates that GAFF/IPolQ-Mod+LJ-Fit yields better results than GAFF/RESP for all groups except for azoles and indoles, as well as aldehydes and ketones. However, the performance for azoles and indoles is almost equal, while the RMSD value for aldehydes and ketones lies well below its total RMSD value, indicated by the horizontal dash-dotted line. The comparison with GAFF/IPolQ-Mod further shows better performance for all groups but bromine compounds, while the corresponding RMSD is again smaller than the overall RMSD. Regarding the deviations by solvent in diagram b), GAFF/IPolQ-Mod+LJ-Fit describes phenols rather poorly compared to GAFF/RESP and GAFF/IPolQ-Mod, while this is distinctively reversed for the grouping by solutes. This is analogous to iodine compounds and GAFF/IPolQ-Mod. For all the other groups, GAFF/IPolQ-Mod+LJ-Fit yields smaller RMSD deviations.

That is, the performance of GAFF/IPolQ-Mod+LJ-Fit is much more homogenous for both solute and solvent groups compared to GAFF/RESP and GAFF/IPolQ-Mod. While these exhibit individual weaknesses, for example in the description of azines & diazoles, esters, amides and nitriles, the level of accuracy for GAFF/IPolQ-Mod+LJ-Fit remains nearly constant for the ΔG_{solv} refitting data set.

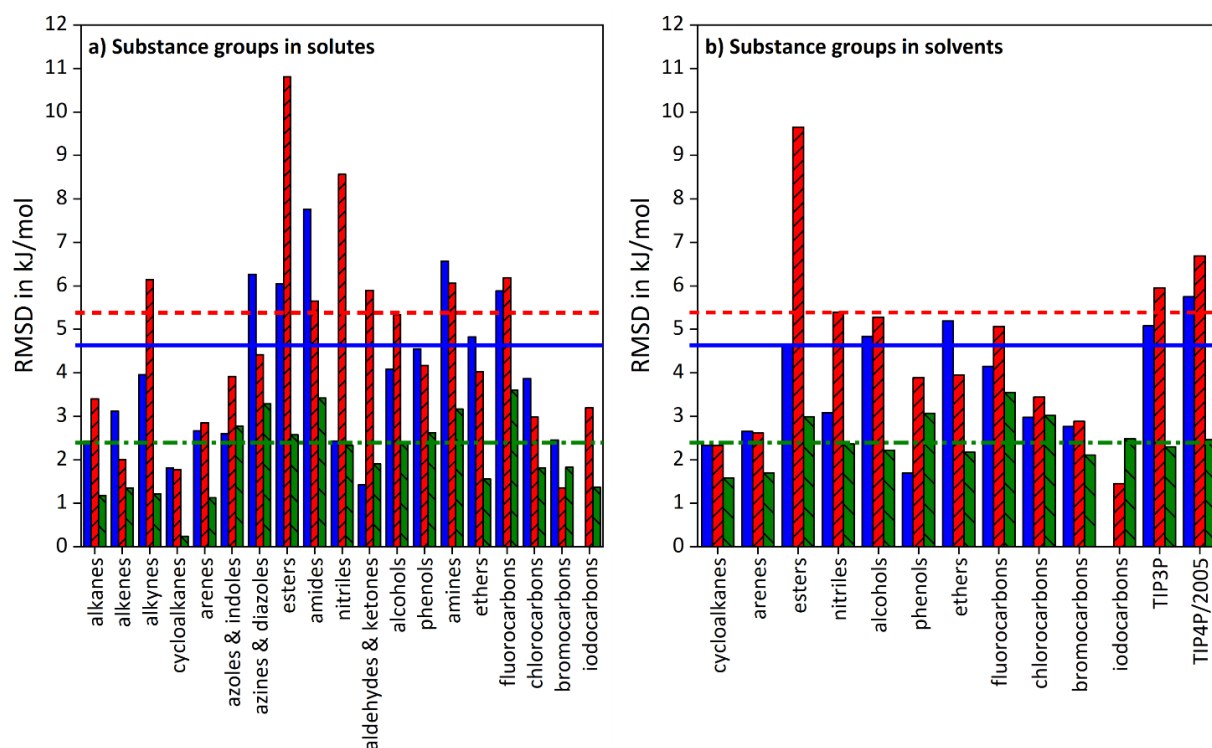


Figure 3. Comparison of RMSD-values for the solvation free energy results from the refitting data set obtained with the various model parameters. In diagram a), RMSD values are given aggregated by substance groups in solutes, while in diagram b), RMSD values are given by substance groups in solvents. Blue full bars refer to GAFF/RESP, red bars with rising pattern to GAFF/IPolQ-Mod and green bars with sloping pattern to GAFF/IPolQ-Mod+LJ-Fit. The bold horizontal lines indicate the overall RMSD values for the data set, whereas the continuous lines represents GAFF/RESP, the dashed lines GAFF/IPolQ-Mod and the dash-dotted lines GAFF/IPolQ-Mod+LJ-Fit.

Some compounds of the refitting data set were also subject of a previous work [101], in which we have compared the performance of different force fields regarding the reproduction of hydration free energies. In the SI we provide the new results of GAFF/IPolQ-Mod+LJ-Fit for this small set of test systems considered in [101] compared to the previous results from GAFF/RESP and GAFF/IPolQ-Mod as well as CGenFF and OPLS-AA.

Simulation results of the density $\rho_{\text{simulation}}$ from the refitting data sets are given in Figure 4 vs. experimental data $\rho_{\text{experiment}}$ [102–131].

Table 4 summarizes the root-mean-square deviations, mean absolute errors (MAE) as well as the slopes m and the Pearson correlation coefficients R of the linear regression curves. GAFF/IPolQ-Mod+LJ-Fit demonstrates slightly improved agreement with experimental data compared to GAFF/RESP, while the results for GAFF/IPolQ-Mod are significantly worse.

In Figure 5, RMSD values for the three sets of model parameters are presented by groups of substances respectively temperature intervals. As for GAFF/IPolQ-Mod, large deviations occur for amines, amides, azoles, nitriles and alcohols, respectively, for temperature ranges $-10\text{ °C} < \theta \leq 20\text{ °C}$ as well as for $\theta > 80\text{ °C}$, while GAFF/RESP demonstrates poor performance for amines and temperatures within the range of $-10\text{ °C} < \theta \leq 0\text{ °C}$. GAFF/IPolQ-Mod+LJ-Fit shows lacking accuracies for amides, as well as for fluorocarbons, bromocarbons and iodocarbons. However, the impact of the temperature is less pronounced.

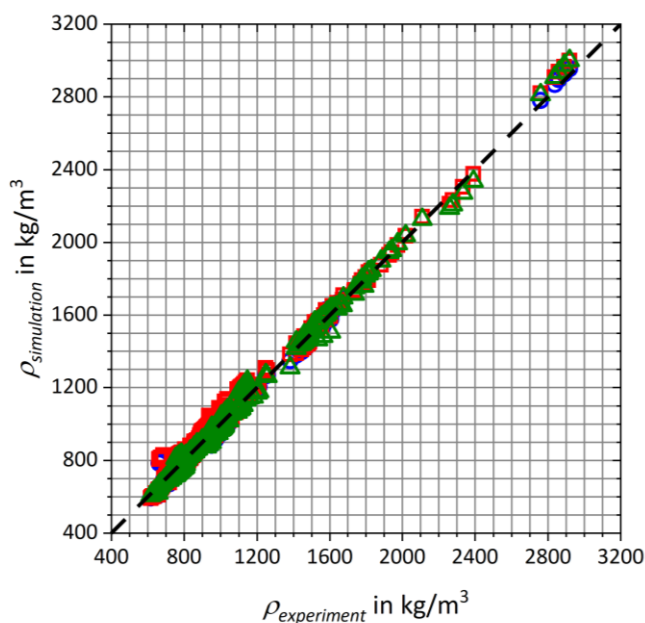


Figure 4. Comparison of simulated densities $\rho_{\text{simulation}}$ vs. experimental reference data $\rho_{\text{experiment}}$ [102–131] from the refitting data set. The results are represented by blue circles for GAFF/RESP, red squares for GAFF/IPolQ-Mod and green triangles for GAFF/IPolQ-Mod+LJ-Fit.

Table 4. Summary of the evaluation of density results from the refitting data set. Besides the root-mean-square deviations (RMSD) and the mean absolute errors (MAE), the slopes m of the linear fitting curves with corresponding Pearson correlation coefficients R are given. For a better comparison with GAFF/RESP, additional values for GAFF/IPolQ-Mod and GAFF/IPolQ-Mod+LJ-Fit are stated in brackets that exclude iodine compounds.

	GAFF/RESP	GAFF/IPolQ-Mod	GAFF/IPolQ-Mod+LJ-Fit
RMSD in kg/m^3	29.80	42.55 (43.27)	26.15 (25.66)
MAE in kg/m^3	22.51	33.78 (34.56)	19.20 (18.80)
Slope m	0.9970	0.9938 (1.0036)	1.0168 (1.0266)
Pearson R	0.9959	0.9956 (0.9941)	0.9979 (0.9974)

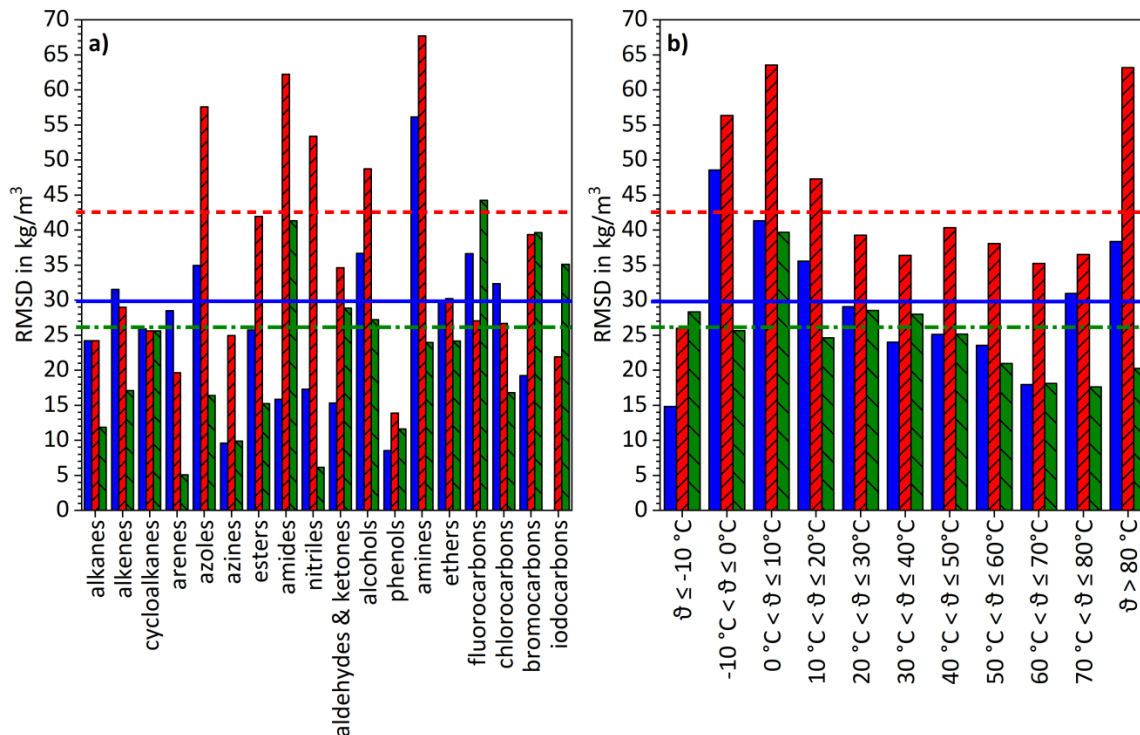


Figure 5. Comparison of RMSD-values for the density results from the refitting data set obtained with the various model parameters. In diagram a), RMSD values are given aggregated by substance groups, while in diagram b), RMSD values are clustered by temperature intervals. Blue full bars refer to GAFF/RESP, red bars with rising pattern to GAFF/IPolQ-Mod and green bars with sloping pattern to GAFF/IPolQ-Mod+LJ-Fit. The bold horizontal lines indicate the overall RMSD values for the data set, whereas the continuous lines represents GAFF/RESP, the dashed lines GAFF/IPolQ-Mod and the dash-dotted lines GAFF/IPolQ-Mod+LJ-Fit.

Evaluation of the Validation I Data Set

Figure 6 shows the simulation results of the solvation free energies from the validation I data set over experimental data [99].

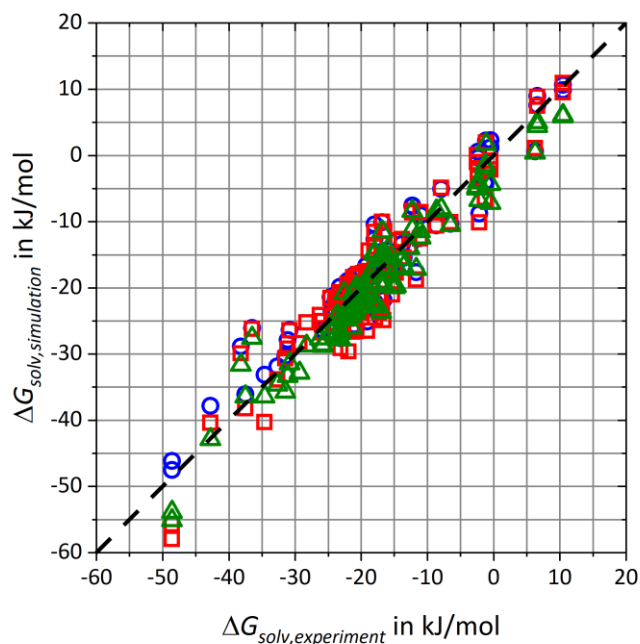


Figure 6. Comparison of simulated solvation free energies $\Delta G_{\text{solv,simulation}}$ vs. experimental reference data $\Delta G_{\text{solv,experiment}}$ [99] from the validation I data set. The results are represented by blue circles for GAFF/RESP, red squares for GAFF/IPolQ-Mod and green triangles for GAFF/IPolQ-Mod+LJ-Fit.

The RMSD and MAE deviations as well as the slopes of the regression curves and the Pearson correlation coefficients are summarized in Table 5. The results shown in Figure 6 and Table 5 illustrate that although GAFF/IPolQ-Mod+LJ-Fit allows for the most accurate predictions of the solvation free energy, the RMSD and MAE deviations for the validation I data set are much higher than for the refitting data set shown in Table 3. In contrast to this, the deviations for GAFF/RESP and GAFF/IPolQ-Mod have decreased significantly, so that GAFF/RESP and GAFF/IPolQ-Mod+LJ-Fit show very similar accuracies in the prediction of ΔG_{solv} .

Table 5. Summary of the evaluation of solvation free energy results from the validation I data set. Besides the root-mean-square deviations (RMSD) and the mean absolute errors (MAE), the slopes m of the linear fitting curves with corresponding Pearson correlation coefficients R are given. For a better comparison with GAFF/RESP, additional values for GAFF/IPolQ-Mod and GAFF/IPolQ-Mod+LJ-Fit are stated in brackets that exclude iodine compounds.

	GAFF/RESP	GAFF/IPolQ-Mod	GAFF/IPolQ-Mod+LJ-Fit
RMSD in kJ/mol	3.30	3.96 (4.07)	3.14 (3.09)
MAE in kJ/mol	2.56	3.14 (3.22)	2.55 (2.50)
Slope m	0.9192	0.9801 (0.9805)	0.9623 (0.9623)
Pearson R	0.9578	0.9418 (0.9412)	0.9701 (0.9706)

Figure 7 demonstrates the RMSD deviations for substance groups in solute and solvent molecules.

Remarkable is the poor description of alkanenes, amides and fluorocarbons as solutes with the new parameters. However, the discussed groups of substances as solvents exhibit better-than-average RMSD values. This suggests that the corresponding substance groups or the atom types associated with them do not generally fail to reflect interactions. This raises the question whether identical atom types are justified for solute and solvent if the partial charges of the molecules are determined differently and the polarization effects are only approximated with IPolQ-Mod. Based on this consideration, there is a risk of underfitting model parameters if no distinction is made between the atom type adaptation in the solute or solvent. Consequently, individual pair potentials might be ideal, though this clearly leads the concept of a general force field ad absurdum.

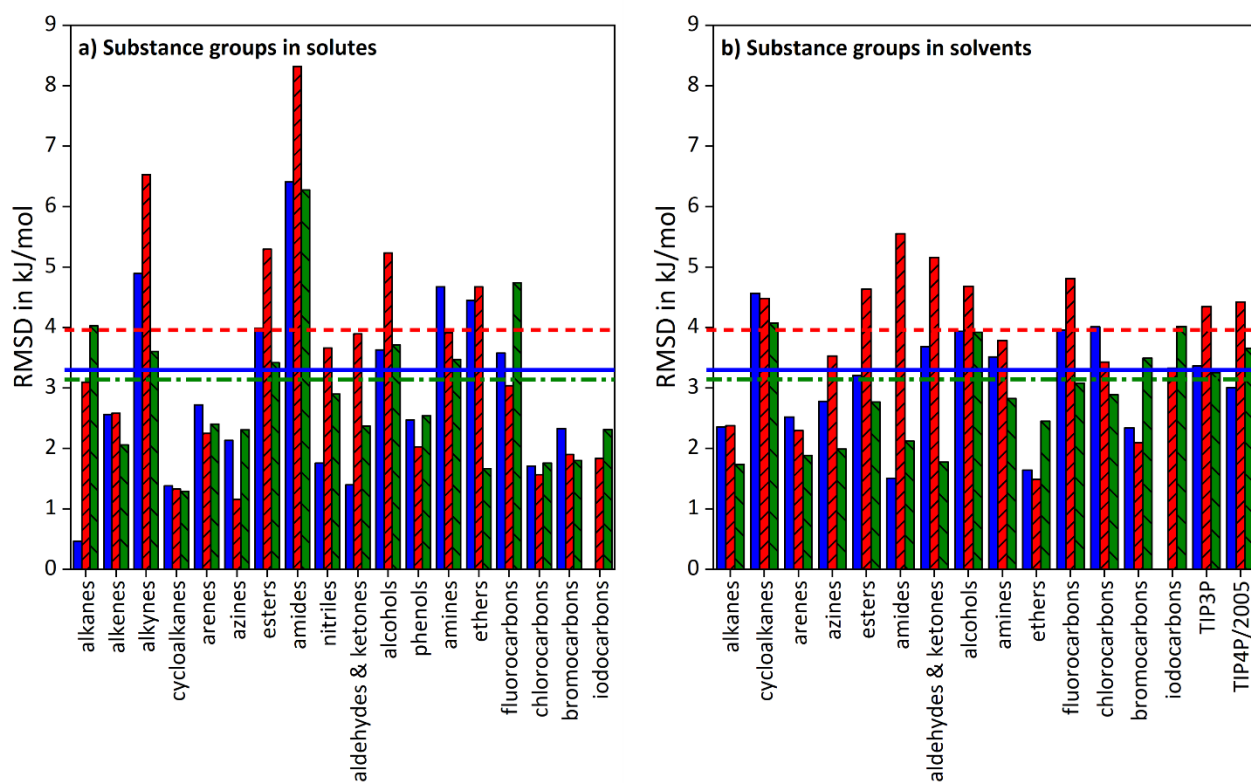


Figure 7. Comparison of RMSD-values for the solvation free energy results from the validation I data set obtained with the various model parameters. In diagram a), RMSD values are given aggregated by substance groups in solutes, while in diagram b), RMSD values are given by substance groups in solvents. Blue full bars refer to GAFF/RESP, red bars with rising pattern to GAFF/IPolQ-Mod and green bars with sloping pattern to GAFF/IPolQ-Mod+LJ-Fit. The bold horizontal lines indicate the overall RMSD values for the data set, whereas the continuous lines represent GAFF/RESP, the dashed lines GAFF/IPolQ-Mod and the dash-dotted lines GAFF/IPolQ-Mod+LJ-Fit.

However, not only do the RMSD values differ significantly between refitting- and validation I data set for GAFF/IPolQ-Mod+LJ-Fit, but also between GAFF/RESP and GAFF/IPolQ-Mod. The differences in the RMSD values between refitting and validation I data sets are $\Delta\text{RMSD}_{\text{GAFF/RESP}} = 1.34$ kJ/mol, $\Delta\text{RMSD}_{\text{GAFF/IPolQ-Mod}} = 1.43$ kJ/mol and $\Delta\text{RMSD}_{\text{GAFF/IPolQ-Mod+LJ-Fit}} = -0.75$ kJ/mol. As the systems from both data sets are of similar complexity and as identical simulation protocols were used leading to comparable statistical accuracies, these discrepancies cannot be attributed to systematic errors in the calculation of ΔG_{solv} . A further analysis (see SI) highlights that the accuracies obtained from the validation I data set do not represent the predictive quality of neither of the models.

Evaluation of the validation II data set

As a consequence, the validation II data set was introduced for the prediction of relative solubilities for the more complex solute substances haloperidol, phenacetin, temazepam and trimethoprim. Based on the calculation of 42 solute/solvent pairs, 237 individual relative solubilities can be determined according to Eq. (1) using combinatorics, which are plotted against experimental data [132] in Figure 8.

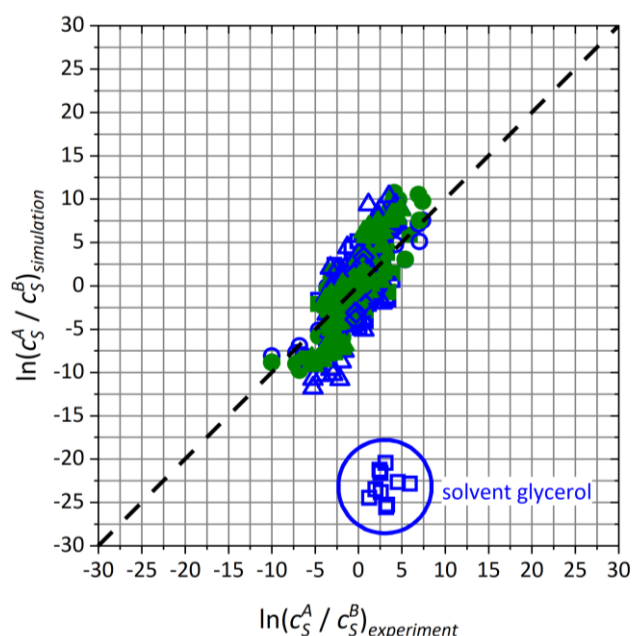


Figure 8. Relative solubilities obtained from simulated solvation free energies of the validation II data set over experimental data [132]. Blue hollow symbols refer to GAFF/RESP, while green full symbols represent GAFF/IPolQ-Mod+LJ-Fit. Solutes haloperidol, phenacetin, temazepam and trimethoprim are displayed by squares, circles, triangles and diamond shapes respectively.

Due to the consistently poor performance of GAFF/IPolQ-Mod, only GAFF/RESP and GAFF/IPolQ-Mod+LJ-Fit models are compared for the validation II data set. Figure 8 illustrates the poor description of relative solubilities for solvate haloperidol and solvent glycerol with GAFF/RESP. However, other relative solubilities also show significant deviations from the experimental reference data, especially for GAFF/RESP. Table 6 summarizes RMSD and MAE deviations as well as slopes m and coefficients R of the regression lines. Even when the haloperidol/glycerol systems are excluded, GAFF/IPolQ-Mod+LJ-Fit has a significantly better overall accuracy than GAFF/RESP.

Table 6. Summary of the evaluation of solvation free energy results from the validation II data set. Besides the root-mean-square deviations (RMSD) and the mean absolute errors (MAE), the slopes m of the linear fitting curves with corresponding Pearson correlation coefficients R are given. In order to determine the effect of the outliers for GAFF/RESP, the results with haloperidol as solute and glycerol as solvent were not considered for the values in brackets. RMSD and MAE deviations for the individual solute compounds are given below the overall values.

	GAFF/RESP	GAFF/IPolQ-Mod+LJ-Fit
RMSD	6.14 (2.98)	2.28 (2.26)
haloperidol	11.53 (2.91)	2.03 (1.89)
phenacetin	1.92	2.97
temazepam	3.27	2.24
trimethoprim	2.09	0.51
MAE	3.41 (2.40)	1.74 (1.72)
haloperidol	6.78 (2.45)	1.67 (1.55)
phenacetin	1.55	2.49
temazepam	2.66	1.66
trimethoprim	1.80	0.45
Slope m	0.5550 (1.1032)	1.2512 (1.2311)
Pearson R	0.2358 (0.6859)	0.8286 (0.8183)

Depending on whether the haloperidol/glycerol outliers are taken into account, the reduction of the RMSD value from the validation II data set for GAFF/IPolQ-Mod+LJ-Fit compared to GAFF/RESP is between 24 - 63 %, which corresponds to a mean value of approximately 44 %. This value agrees very well with the decrease in the RMSD value of about 42 % for a hypothetical basis population from the refitting and validation I data set as discussed in the SI. In contrast, the validation I data set shows a decrease of only 5 %. We therefore conclude that the validation II data set gives a more reasonable estimate for the potential of our newly developed model parameters aiming at improved predictions of solvation free energies and relative solubilities.

Evaluation of the validation III data set

For validation III, 100 liquid densities of binary mixtures were calculated. Simulation results over experimental reference data [102,118,120,133–138] are given in Figure 9.

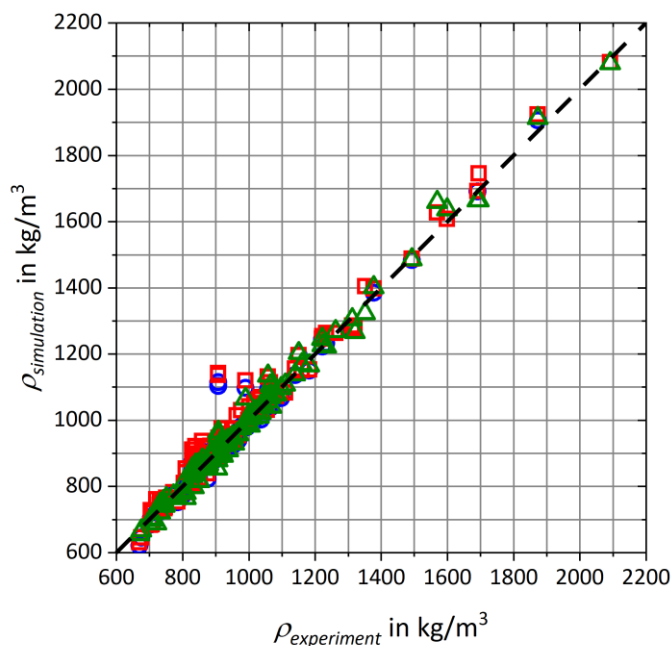


Figure 9. Comparison of simulated densities $\rho_{\text{simulation}}$ vs. experimental reference data $\rho_{\text{experiment}}$ [102,118,120,133–138] for binary mixtures from the validation III data set. The results are represented by blue circles for GAFF/RESP, red squares for GAFF/IPolQ-Mod and green triangles for GAFF/IPolQ-Mod+LJ-Fit.

The evaluation of the predictive performance by set of model parameters is summarized in Table 7.

Table 7. Summary of the evaluation of density results for binary mixtures from the validation III data set. Besides the root-mean-square deviations (RMSD) and the mean absolute errors (MAE), the slopes m of the linear fitting curves with corresponding Pearson correlation coefficients R are given. For a better comparison with GAFF/RESP, additional values for GAFF/IPolQ-Mod and GAFF/IPolQ-Mod+LJ-Fit are stated in brackets that exclude iodine compounds.

	GAFF/RESP	GAFF/IPolQ-Mod	GAFF/IPolQ-Mod+LJ-Fit
RMSD in kg/m^3	39.06	47.85 (49.08)	24.93 (22.89)
MAE in kg/m^3	23.47	31.71 (32.04)	17.84 (16.60)
Slope m	1.002	0.9977 (1.0041)	1.0132 (1.0193)
Pearson R	0.9816	0.9856 (0.9766)	0.9954 (0.9939)

The comparison of results in Table 7 demonstrates high correlations R and regression slopes m close to 1 for all parameter sets. However, the RMSD value for GAFF/IPolQ-Mod+LJ-Fit decreased by approximately 41 % and 48 % compared to GAFF/RESP and GAFF/IPolQ-Mod, respectively, and is in close agreement to the accuracies presented in Table 4. RMSD values aggregated by substance groups respectively temperature intervals are displayed in Figure 10.

For GAFF/RESP and GAFF/IPolQ-Mod, the left figure demonstrates shortcomings for amines as well as for both the TIP3P and TIP4P/2005 water models. Although GAFF/IPolQ-Mod+LJ-Fit indicates deviations for fluorine and iodine compounds, these outliers are still significantly smaller than those of the other parameter sets. Regarding the reproduction of experimental data over a broad temperature range, GAFF/RESP and GAFF/IPolQ-Mod show extreme faults for systems within the temperature range of $70^\circ\text{C} < \vartheta \leq 80^\circ\text{C}$, while the temperature impact on GAFF/IPolQ-Mod+LJ-Fit results is comparably small.

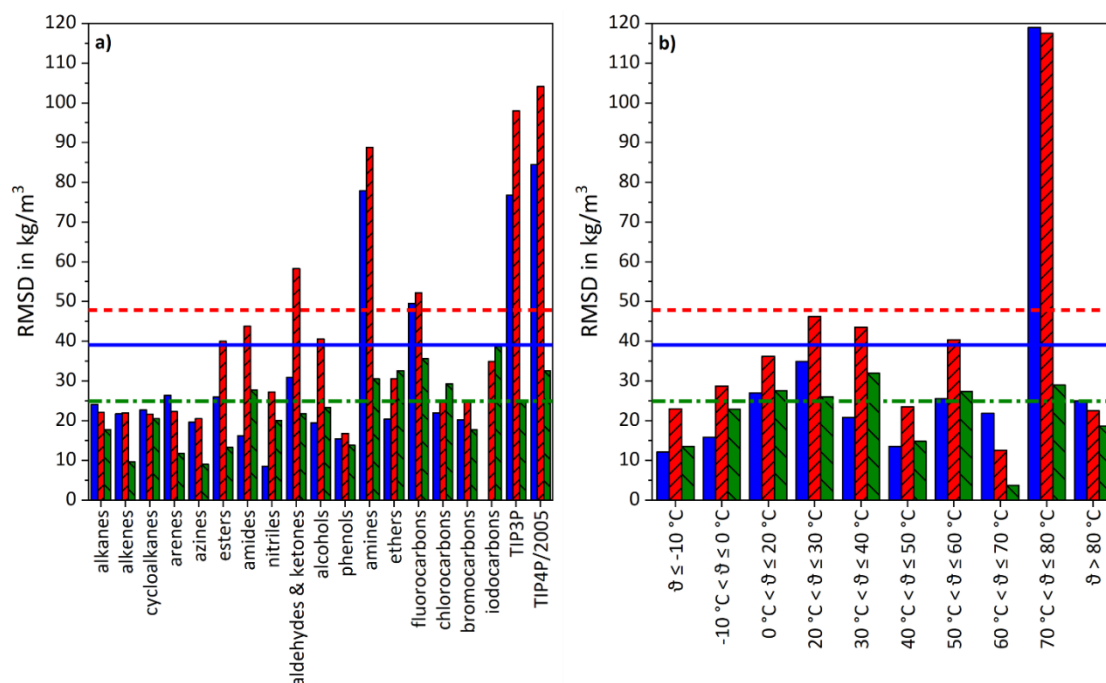


Figure 10. Comparison of RMSD-values for the density results from the validation III data set obtained with the various model parameters. In diagram a), RMSD values are given aggregated by substance groups, while in diagram b), RMSD values are clustered by temperature intervals. Blue full bars refer to GAFF/RESP, red bars with rising pattern to GAFF/IPolQ-Mod and green bars with sloping pattern to GAFF/IPolQ-Mod+LJ-Fit. The bold horizontal lines indicate the overall RMSD values for the data set, whereas the continuous lines represent GAFF/RESP, the dashed lines GAFF/IPolQ-Mod and the dash-dotted lines GAFF/IPolQ-Mod+LJ-Fit.

Conclusions

Molecular simulations offer great potential for a better understanding of complex processes such as solubility as they sample systems on the molecular level. However, accurate simulations require accurate molecular models. As polarization is considered to be an impacting factor, though linked to high computational effort, there is need for an implicit representation of polarization effects, for example using the IPolQ-Mod method. In this work, we have optimized GAFF atom types (GAFF/IPolQ-Mod+LJ-Fit) for a variety of substance groups considering IPolQ-Mod partial charges to improve the description of solvation free energies, respectively relative solubilities, as well as liquid densities. The evaluation of our refitting data set highlights significant improvements in the description of solvation free energies for our new parameters compared to default GAFF (GAFF/RESP), but especially for GAFF with IPolQ-Mod charges but not-optimized parameters (GAFF/IPolQ-Mod). The improvement regarding the prediction of liquid densities for pure compounds is minor compared to default GAFF.

Regarding the validation, the description of densities for binary mixtures is significantly better with our new parameters. However, the accuracies of the free energy predictions for default RESP and our optimized parameters are almost identical. By in-depth analyses, partly given in the SI, we conclude that the free energy validation data does not represent the overall performance for neither GAFF/RESP nor GAFF/IPolQ-Mod. We therefore deduce that the quality of our new model parameters is misrepresented as well. As a consequence, we further compared GAFF/RESP and GAFF/IPolQ-Mod+LJ-Fit for the description of relative solubilities of four drug-like structures in a multitude of solvents, resulting in a total of 237 individual relative solubilities. The improvement in root-mean square deviations between GAFF/IPolQ-Mod+LJ and GAFF/RESP of around 44 % is in much better agreement with the reduction of RMSD for the total of both previous free energy data sets. By this, our newly derived parameters for GAFF in combination with IPolQ-Mod apparently allow for a significant improvement in the prediction of relative solubilities.

Acknowledgements: Andreas Mecklenfeld acknowledges funding by a Georg-Christoph-Lichtenberg Fellowship by the Federal State of Lower Saxony, Germany. Molecular dynamics simulations were performed with resources provided by the North-German Supercomputing Alliance (HLRN). We greatly appreciate the support.

Conflict of interest: The authors declare no competing financial interest.

References

- [1] Y. Gong, D. J. Grant, H. G. Brittain. *Solvent Systems and Their Selection in Pharmaceuticals and Biopharmaceutics*. P. Augustijns, and M. E. Brewster, Eds., Springer, New York, NY, USA 2007 p.1.
- [2] S. Liu, S. Cao, K. Hoang, K. L. Young, A. S. Paluch, D. L. Mobley. Using MD Simulations To Calculate How Solvents Modulate Solubility. *J Chem Theory Comput.* **12** no. 4 (2016) 1930–1941.
- [3] *Free Energy Calculations*. C. Chipot, and A. Pohorille, Eds., Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg, Germany 2007.
- [4] A. Mecklenfeld, G. Raabe. Efficient solvation free energy simulations. *Mol. Phys.* **115** 9-12 (2017) 1322–1334.
- [5] J. P. M. Jambeck, F. Mocci, A. P. Lyubartsev, A. Laaksonen. Partial atomic charges and their impact on the free energy of solvation. *J. Comput. Chem.* **34** no. 3 (2013) 187–197.
- [6] A. Mecklenfeld, G. Raabe. Comparison of RESP and IPolQ-Mod Partial Charges for Solvation Free Energy Calculations of Various Solute/Solvent Pairs. *J. Chem. Theory Comput.* **13** no. 12 (2017) 6266–6274.
- [7] A. Nicholls, D. L. Mobley, J. P. Guthrie, J. D. Chodera, C. I. Bayly, M. D. Cooper, V. S. Pande. Predicting small-molecule solvation free energies. *J. Med.Chem.* **51** no. 4 (2008) 769–779.
- [8] J. P. Guthrie. A blind challenge for computational solvation free energies. *J. Phys. Chem. B* **113** no. 14 (2009) 4501–4507.
- [9] M. T. Geballe, A. G. Skillman, A. Nicholls, J. P. Guthrie, P. J. Taylor. The SAMPL2 blind prediction challenge: introduction and overview. *J. Comput. Aided Mol. Des.* **24** no. 4 259–279.
- [10] D. L. Mobley, K. L. Wymer, N. M. Lim, J. P. Guthrie. Blind prediction of solvation free energies from the SAMPL4 challenge. *J Comput Aided Mol Des* **28** no. 3 (2014) 135–150.
- [11] H. S. Muddana, N. V. Sapra, A. T. Fenley, M. K. Gilson. The SAMPL4 hydration challenge: evaluation of partial charge sets with explicit-water molecular dynamics simulations. *J. Comput. Aided Mol. Des.* **28** no. 3 (2014) 277–287.
- [12] R. C. Rizzo, T. Aynechi, D. A. Case, I. D. Kuntz. Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. *J. Chem. Theory Comput.* **2** no. 1 (2006) 128–139.
- [13] C. J. Fennell, K. L. Wymer, D. L. Mobley. A fixed-charge model for alcohol polarization in the condensed phase, and its role in small molecule hydration. *J. Phys. Chem. B* **118** no. 24 (2014) 6438–6446.
- [14] A. Ahmed, S. I. Sandler. Hydration Free Energies of Multifunctional Nitroaromatic Compounds. *J. Chem. Theory Comput.* **9** no. 6 (2013) 2774–2785.
- [15] J. P. M. Jämbeck, A. P. Lyubartsev. Update to the general amber force field for small solutes with an emphasis on free energies of hydration. *J. Phys. Chem. B* **118** no. 14 (2014) 3793–3804.
- [16] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case. Development and testing of a general amber force field. *J. Comput. Chem.* **25** no. 9 (2004) 1157–1174.
- [17] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **118** no. 45 (1996) 11225–11236.
- [18] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. MacKerell, JR. CHARMM general force field: A force field for drug-like

- molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31** no. 4 (2010) 671–690.
- [19] A. K. Rappe, W. A. Goddard. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.* **95** no. 8 (1991) 3358–3363.
- [20] S. W. Rick, S. J. Stuart, B. J. Berne. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.* **101** no. 7 (1994) 6141–6156.
- [21] J. L. Banks, G. A. Kaminski, R. Zhou, D. T. Mainz, B. J. Berne, R. A. Friesner. Parametrizing a polarizable force field from ab initio data. I. The fluctuating point charge model. *J. Chem. Phys.* **110** no. 2 (1999) 741–754.
- [22] S. Patel, C. L. Brooks. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **25** no. 1 (2004) 1–15.
- [23] S. Patel, A. D. Mackerell, C. L. Brooks. CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **25** no. 12 (2004) 1504–1514.
- [24] S. Patel, C. L. Brooks. Fluctuating charge force fields: recent developments and applications from small molecules to macromolecular biological systems. *Mol. Simulat.* **32** 3-4 (2006) 231–249.
- [25] G. Lamoureux, B. Roux. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.* **119** no. 6 (2003) 3025–3039.
- [26] J. A. Lemkul, J. Huang, B. Roux, A. D. Mackerell. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **116** no. 9 (2016) 4983–5013.
- [27] P. E. M. Lopes, J. Huang, J. Shim, Y. Luo, H. Li, B. Roux, A. D. Mackerell, JR. Force Field for Peptides and Proteins based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **9** no. 12 (2013) 5430–5449.
- [28] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, P. Ren. The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **9** no. 9 (2013) 4046–4063.
- [29] N. A. Mohamed, R. T. Bradshaw, J. W. Essex. Evaluation of solvation free energies for small molecules with the AMOEBA polarizable force field. *J. Comput. Chem.* **37** no. 32 (2016) 2749–2758.
- [30] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, T. Head-Gordon. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **114** no. 8 (2010) 2549–2564.
- [31] J. A. Lemkul, B. Roux, D. van der Spoel, A. D. Mackerell, JR. Implementation of extended Lagrangian dynamics in GROMACS for polarizable simulations using the classical Drude oscillator model. *J. Comput. Chem.* **36** no. 19 (2015) 1473–1479.
- [32] D. S. Cerutti, J. E. Rice, W. C. Swope, D. A. Case. Derivation of fixed partial charges for amino acids accommodating a specific water model and implicit polarization. *J. Phys. Chem. B* **117** no. 8 (2013) 2328–2338.
- [33] C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges. *J. Phys. Chem.* **97** no. 40 (1993) 10269–10280.
- [34] R. S. D. Meine. *2-Substituierte Indol-3-carbonitrile als neue Inhibitoren der Proteinkinase DYRK1A*. Dissertation. Technische Universität Braunschweig, Braunschweig, Germany (2018).
- [35] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79** no. 2 (1983) 926–926.
- [36] J. L. F. Abascal, C. Vega. A general purpose model for the condensed phases of water. *J. Chem. Phys.* **123** no. 23 (2005) 234505–234505.
- [37] J. Wang, W. Wang, P. A. Kollman, D. A. Case. Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc.* **222** (2001) U403.
- [38] D. A. Case, J. T. Berryman, R. M. Betz, D. S. Cerutti, Cheatham, T. E., III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Gotz, N. Homeyer, S. Izadi, P. A. Janowski, J. W. Kaus, A. Kovalenko, T. Lee, S.

- Le Grand, P. Li, T. Luchko, R. Luo, B. D. Madej, K. M. Merz, G. Monard, H. Needham, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. E. Roitberg, R. Salomon-Ferrer, C. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, D. M. York, P. A. Kollman. *AMBER*, University of California: San Francisco, CA 2015.
- [39] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl. *GROMACS. SoftwareX* **1-2** (2015) 19–25.
- [40] H. J. C. Berendsen, D. van der Spoel, R. van Drunen. *GROMACS - a message-passing parallel molecular-dynamics implementation. Comput. Phys.* **91** 1-3 (1995) 43–56.
- [41] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl. *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J. Chem. Theory Comput.* **4** no. 3 (2008) 435–447.
- [42] E. Lindahl, B. Hess, D. van der Spoel. *GROMACS 3.0. J. Mol. Model.* **7** no. 8 (2001) 306–317.
- [43] S. Páll, M. J. Abraham, C. Kutzner, B. Hess, E. Lindahl. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS in *Solving Software Challenges for Exascale*, S. Markidis and E. Laure, Eds., vol. 8759, Springer International Publishing, Cham 2015 3–27.
- [44] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl. *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics* **29** no. 7 (2013) 845–854.
- [45] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, Berendsen, Herman J. C. *GROMACS: Fast, flexible, and free. J. Comput. Chem.* **26** no. 16 (2005) 1701–1718.
- [46] A. W. Sousa da Silva, W. F. Vranken. *ACPYPE - AnteChamber PYthon Parser interface. BMC research notes* **5** (2012) 367–367.
- [47] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, Montgomery, J. A., Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox. *Gaussian 09, Revision A.02*, Gaussian, Inc., Wallingford CT 2009.
- [48] R. Ditchfield. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **54** no. 2 (1971) 724–724.
- [49] W. J. Hehre. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **56** no. 5 (1972) 2257–2257.
- [50] P.C. Hariharan, J.A. Pople. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoret. Chim. Acta* **28** no. 3 (1973) 213–222.
- [51] P. C. Hariharan, J. A. Pople. Accuracy of AH n equilibrium geometries by single determinant molecular orbital theory. *Molecular Physics* **27** no. 1 (1974) 209–214.
- [52] M. S. Gordon. The isomers of silacyclopropane. *Chemical Physics Letters* **76** no. 1 (1980) 163–168.
- [53] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, J. A. Pople. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **77** no. 7 (1982) 3654–3665.
- [54] R. C. Binning, L. A. Curtiss. Compact contracted basis sets for third-row atoms. *J. Comput. Chem.* **11** no. 10 (1990) 1206–1216.
- [55] J.-P. Blaudeau, M. P. McGrath, L. A. Curtiss, L. Radom. Extension of Gaussian-2 (G2) theory to molecules containing third-row atoms K and Ca. *J. Chem. Phys.* **107** no. 13 (1997) 5016–5016.

- [56] V. A. Rassolov, J. A. Pople, M. A. Ratner, T. L. Windus. 6-31G* basis set for atoms K through Zn. *J. Chem. Phys.* **109** no. 4 (1998) 1223–1223.
- [57] V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern, L. A. Curtiss. 6-31G* basis set for third-row atoms. *J. Comput. Chem.* **22** no. 9 (2001) 976–984.
- [58] C. C. J. Roothaan. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **23** no. 2 (1951) 69–89.
- [59] A. Jakalian, B. L. Bush, D. B. Jack, C. I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **21** no. 2 (2000) 132–146.
- [60] A. Jakalian, D. B. Jack, C. I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23** no. 16 (2002) 1623–1641.
- [61] T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90** no. 2 (1989) 1007–1007.
- [62] R. A. Kendall, T. H. Dunning, R. J. Harrison. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **96** no. 9 (1992) 6796–6796.
- [63] D. E. Woon, T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **98** no. 2 (1993) 1358–1371.
- [64] K. A. Peterson, D. E. Woon, T. H. Dunning. Benchmark calculations with correlated molecular wave functions. IV. The classical barrier height of the $H+H_2 \rightarrow H_2+H$ reaction. *J. Chem. Phys.* **100** no. 10 (1994) 7410–7410.
- [65] A. K. Wilson, T. van Mourik, T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. VI. Sextuple zeta correlation consistent basis sets for boron through neon. *J. Mol. Struct.* **388** (1996) 339–349.
- [66] M. J. Frisch, M. Head-Gordon, J. A. Pople. A direct MP2 gradient method. *Chemical Physics Letters* **166** no. 3 (1990) 275–280.
- [67] M. J. Frisch, M. Head-Gordon, J. A. Pople. Semi-direct algorithms for the MP2 energy and gradient. *Chemical Physics Letters* **166** no. 3 (1990) 281–289.
- [68] M. Head-Gordon, T. Head-Gordon. Analytic MP2 frequencies without fifth-order storage. Theory and application to bifurcated hydrogen bonds in the water hexamer. *Chemical Physics Letters* **220** 1-2 (1994) 122–128.
- [69] M. Head-Gordon, J. A. Pople, M. J. Frisch. MP2 energy evaluation by direct methods. *Chemical Physics Letters* **153** no. 6 (1988) 503–506.
- [70] S. Sæbø, J. Almlöf. Avoiding the integral storage bottleneck in LCAO calculations of electron correlation. *Chemical Physics Letters* **154** no. 1 (1989) 83–89.
- [71] B. Mennucci, R. Cammi, J. Tomasi. Excited states and solvatochromic shifts within a nonequilibrium solvation approach. *J. Chem. Phys.* **109** no. 7 (1998) 2798–2807.
- [72] K. A. Peterson, D. Figgen, E. Goll, H. Stoll, M. Dolg. Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post- d group 16–18 elements. *J. Chem. Phys.* **119** no. 21 (2003) 11113–11123.
- [73] K. A. Peterson, B. C. Shepler, D. Figgen, H. Stoll. On the spectroscopic and thermochemical properties of ClO, BrO, IO, and their anions. *J. Phys. Chem. A* **110** no. 51 (2006) 13877–13883.
- [74] K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, T. L. Windus. Basis set exchange. *J. Chem. Inf. Model.* **47** no. 3 (2007) 1045–1052.
- [75] D. Feller. The role of databases in support of computational chemistry calculations. *J. Comput. Chem.* **17** no. 13 (1996) 1571–1586.
- [76] R. A. Messerly, S. M. Razavi, M. R. Shirts. Configuration-Sampling-Based Surrogate Models for Rapid Parameterization of Non-Bonded Interactions. *J. Chem. Theory Comput.* **14** no. 6 (2018) 3144–3162.
- [77] L.-P. Wang, T. J. Martinez, V. S. Pande. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *The journal of physical chemistry letters* **5** no. 11 (2014) 1885–1891.

- [78] A. Mecklenfeld, G. Raabe. Applicability of a thermodynamic cycle approach for a force field parametrization targeting non-aqueous solvation free energies. *J. Comput. Aided Mol. Des.* **34** no. 1 (2019) 71–82.
- [79] R. Faller, H. Schmitz, O. Biermann, F. Müller-Plathe. Automatic parameterization of force fields for liquids by simplex optimization. *J. Comput. Chem.* **20** no. 10 (1999) 1009–1017.
- [80] J. A. Nelder, R. Mead. A Simplex Method for Function Minimization. *Comput. J.* **7** no. 4 (1965) 308–313.
- [81] C. G. Mayne, J. Saam, K. Schulten, E. Tajkhorshid, J. C. Gumbart. Rapid parameterization of small molecules using the Force Field Toolkit. *J. Comput. Chem.* **34** no. 32 (2013) 2757–2770.
- [82] J. C. Fogarty, S.-W. Chiu, P. Kirby, E. Jakobsson, S. A. Pandit. Automated optimization of water-water interaction parameters for a coarse-grained model. *J. Phys. Chem. B* **118** no. 6 (2014) 1603–1611.
- [83] L. Vlcek, A. A. Chialvo. Rigorous force field optimization principles based on statistical distance minimization. *J. Chem. Phys.* **143** no. 14 (2015) 144110–144110.
- [84] L.-P. Wang, J. Chen, T. van Voorhis. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *J. Chem. Theory Comput.* **9** no. 1 (2013) 452–460.
- [85] L.-P. Wang(11/2011), <https://github.com/leeping/forcebalance/blob/master/src/optimizer.py>.
- [86] N. Goga, A. J. Rzepiela, A. H. de Vries, S. J. Marrink, H. J. C. Berendsen. Efficient Algorithms for Langevin and DPD Dynamics. *J. Chem. Theory Comput.* **8** no. 10 (2012) 3637–3649.
- [87] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81** no. 8 (1984) 3684–3684.
- [88] M. Parrinello, A. Rahman. Crystal Structure and Pair Potentials. *Phys. Rev. Lett.* **45** no. 14 (1980) 1196–1199.
- [89] S. Parameswaran, D. L. Mobley. Box size effects are negligible for solvation free energies of neutral solutes. *J. Comput. Aided Mol. Des.* **28** no. 8 (2014) 825–829.
- [90] L. Martinez, R. Andrade, E. G. Birgin, J. M. Martinez. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30** no. 13 (2009) 2157–2164.
- [91] S. Miyamoto, P. A. Kollman. Settle. *J. Comput. Chem.* **13** no. 8 (1992) 952–962.
- [92] A. Villa, A. E. Mark. Calculation of the free energy of solvation for neutral analogs of amino acid side chains. *J. Comput. Chem.* **23** no. 5 (2002) 548–553.
- [93] G. Duarte Ramos Matos, D. Y. Kyu, H. H. Loeffler, J. D. Chodera, M. R. Shirts, D. L. Mobley. Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. *J. Chem. Eng. Data* (2017).
- [94] M. R. Shirts, J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129** no. 12 (2008) 124105–124105.
- [95] P. V. Klimovich, M. R. Shirts, D. L. Mobley. Guidelines for the analysis of free energy calculations. *J. Comput. Aided Mol. Des.* **29** no. 5 (2015) 397–411.
- [96] T. T. Pham, M. R. Shirts. Identifying low variance pathways for free energy calculations of molecular transformations in solution phase. *J. Chem. Phys.* **135** no. 3 (2011) 34114–34114.
- [97] D. K. Shenfeld, H. Xu, M. P. Eastwood, R. O. Dror, D. E. Shaw. Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations. *Phys. Rev. E* **80** no. 4 (2009) 46705–46705.
- [98] B. Efron. Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist.* **7** no. 1 (1979) 1–26.
- [99] A. V. Marenich, C. P. Kelly, J. D. Thompson, G. D. Hawkins, C. C. Chambers, D. J. Giesen, P. Winget, C. J. Cramer, D. G. Truhlar. Minnesota Solvation Database – version 2012, University of Minnesota, Minneapolis.
- [100] D. L. Mobley, J. P. Guthrie. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J. Comput. Aided Mol. Des.* **28** no. 7 (2014) 711–720.

- [101] A. Mecklenfeld, G. Raabe. *Efficient Molecular Simulations of the Free Energy of Solvation - Jahrestreffen der ProcessNet-Fachgruppe Molekulare Modellierung*, Frankfurt/Main, Germany 03/09/2017.
- [102] H. Landolt, K.-H. Hellwege, O. Madelung. *Zahlenwerte und Funktionen aus Naturwissenschaften und Technik*, Springer, Berlin 1974.
- [103] A. García-Abuín, D. Gómez-Díaz, M. D. La Rubia, J. M. Navaza, R. Pacheco. Density, Speed of Sound, and Isentropic Compressibility of Triethanolamine (or N -Methyldiethanolamine) + Water + Ethanol Solutions from $t = (15 \text{ to } 50) \text{ }^\circ\text{C}$. *J. Chem. Eng. Data* **54** no. 11 (2009) 3114–3117.
- [104] G. Sivaramprasad, M. V. Rao, D. H. L. Prasad. Density and viscosity of ethanol + 1,2-dichloroethane, ethanol + 1,1,1-trichloroethane, and ethanol + 1,1,2,2-tetrachloroethane binary mixtures. *J. Chem. Eng. Data* **35** no. 2 (1990) 122–124.
- [105] M. T. Zafarani-Moattar, N. Tohidifar. Vapor–Liquid Equilibria, Density, and Speed of Sound for the System Poly(ethylene glycol) 400 + Methanol at Different Temperatures. *J. Chem. Eng. Data* **51** no. 5 (2006) 1769–1774.
- [106] D. L. Cunha, J. A. P. Coutinho, J. L. Daridon, R. A. Reis, M. L. L. Paredes. Experimental Densities and Speeds of Sound of Substituted Phenols and Their Modeling with the Prigogine–Flory–Patterson Model. *J. Chem. Eng. Data* **58** no. 11 (2013) 2925–2931.
- [107] R. Rosal, I. Medina, E. Forster, J. MacInnes. Viscosities and densities for binary mixtures of cresols. *Fluid Phase Equilibria* **211** no. 2 (2003) 143–150.
- [108] J. A. Al-Kandary, A. S. Al-Jimaz, A.-H. M. Abdul-Latif. Viscosities, Densities, and Speeds of Sound of Binary Mixtures of Benzene, Toluene, o -Xylene, m -Xylene, p -Xylene, and Mesitylene with Anisole at (288.15, 293.15, 298.15, and 303.15) K. *J. Chem. Eng. Data* **51** no. 6 (2006) 2074–2082.
- [109] J. N. Nayak, M. I. Aralaguppi, T. M. Aminabhavi. Density, Viscosity, Refractive Index, and Speed of Sound in the Binary Mixtures of Ethyl Chloroacetate with Aromatic Liquids at 298.15, 303.15, and 308.15 K. *J. Chem. Eng. Data* **47** no. 4 (2002) 964–969.
- [110] M. A. Varfolomeev, I. T. Rakipov, B. N. Solomonov, W. Marczak. Speed of Sound, Density, and Related Thermodynamic Excess Properties of Binary Mixtures of 2-Pyrrolidone and N -Methyl-2-pyrrolidone with Acetonitrile and Chloroform. *J. Chem. Eng. Data* **61** no. 3 (2016) 1032–1046.
- [111] M. A. Varfolomeev, K. V. Zaitseva, I. T. Rakipov, B. N. Solomonov, W. Marczak. Speed of Sound, Density, and Related Thermodynamic Excess Properties of Binary Mixtures of Butan-2-one with C1–C4 n -Alkanols and Chloroform. *J. Chem. Eng. Data* **59** no. 12 (2014) 4118–4132.
- [112] T. M. Aminabhavi, K. Banerjee. Density, Viscosity, Refractive Index, and Speed of Sound in Binary Mixtures of Dimethyl Carbonate with Methanol, Chloroform, Carbon Tetrachloride, Cyclohexane, and Dichloromethane in the Temperature Interval (298.15–308.15) K. *J. Chem. Eng. Data* **43** no. 6 (1998) 1096–1101.
- [113] J. N. Nayak, M. I. Aralaguppi, T. M. Aminabhavi. Density, Viscosity, Refractive Index, and Speed of Sound in the Binary Mixtures of Ethyl Chloroacetate + Cyclohexanone, + Chlorobenzene, + Bromobenzene, or + Benzyl Alcohol at (298.15, 303.15, and 308.15) K. *J. Chem. Eng. Data* **48** no. 3 (2003) 628–631.
- [114] G. A. Torín-Ollarves, J. J. Segovia, M. C. Martín, M. A. Villamañán. Density, Viscosity, and Isobaric Heat Capacity of the Mixture (1-Butanol + 1-Hexene). *J. Chem. Eng. Data* **58** no. 10 (2013) 2717–2723.
- [115] A. M. Kerimov, T. A. Apaev. Experimental values of density of 1-hexene, 1-octene, cyclohexene, cyclohexane and methylcyclohexane in dependence on temperature and pressure. *Teplofiz.Svoistva Vesh.Mater.* (1972) 26–46.
- [116] D. I. Sagdeev, M. G. Fomina, G. K. Mukhamedzyanov, I. M. Abdulagatov. Experimental Study and Correlation Models of the Density and Viscosity of 1-Hexene and 1-Heptene at Temperatures from (298 to 473) K and Pressures up to 245 MPa. *J. Chem. Eng. Data* **59** no. 4 (2014) 1105–1119.
- [117] J. N. Nayak, M. I. Aralaguppi, U. S. Toti, T. M. Aminabhavi. Density, Viscosity, Refractive Index, and Speed of Sound in the Binary Mixtures of Tri- n -butylamine + Triethylamine, + Tetrahydrofuran, +

- Tetradecane, + Tetrachloroethylene, + Pyridine, or + Trichloroethylene at (298.15, 303.15, and 308.15) K. *J. Chem. Eng. Data* **48** no. 6 (2003) 1483–1488.
- [118] L.-C. Wang, H. Ding, J.-H. Zhao, C.-Y. Song, J.-S. Wang. Density and Viscosity of (4-Picoline + Water) Binary Mixtures from T = (298.15 to 338.15) K. *J. Chem. Eng. Data* **54** no. 3 (2009) 1000–1003.
- [119] W. Marczak. Speed of Ultrasound, Density, and Adiabatic Compressibility for 3-Methylpyridine + Heavy Water in the Temperature Range 293–313 K. *J. Chem. Eng. Data* **41** no. 6 (1996) 1462–1465.
- [120] G. I. Egorov, D. M. Makarov. Densities and Molar Isobaric Thermal Expansions of the Water + Formamide Mixture over the Temperature Range from 274.15 to 333.15 K at Atmospheric Pressure. *J. Chem. Eng. Data* **62** no. 4 (2017) 1247–1256.
- [121] J. M. Bernal-García, A. Guzmán-López, A. Cabrales-Torres, A. Estrada-Baltazar, G. A. Iglesias-Silva. Densities and Viscosities of (N, N -Dimethylformamide + Water) at Atmospheric Pressure from (283.15 to 353.15) K. *J. Chem. Eng. Data* **53** no. 4 (2008) 1024–1027.
- [122] S. Mrad, C. Lafuente, M. Hichri, I. Khattech. Density, Speed of Sound, Refractive Index, and Viscosity of the Binary Mixtures of N, N -dimethylacetamide with Methanol and Ethanol. *J. Chem. Eng. Data* **61** no. 9 (2016) 2946–2953.
- [123] A. A. Dyshin, O. V. Eliseeva, M. G. Kiselev. Density and Viscosity of N -Methylacetamide–Calcium Chloride Mixtures over the Temperature Range from 308.15 to 328.15 K at Atmospheric Pressure. *J. Chem. Eng. Data* **62** no. 12 (2017) 4128–4132.
- [124] K. Hofmann. *The Chemistry of Heterocyclic Compounds, Imidazole and Its Derivatives*, John Wiley & Sons, Hoboken 1953.
- [125] M. T. Khimenko, N. N. Gritsenko. Determination of the Polarisabilities and Radii of the Acetonitrile and Dimethylacetamide Molecules. *Zh. Fiz. Khim.* **54** (1980) 198–199.
- [126] Y. Lei, Z. Chen, X. An, M. Huang, W. Shen. Measurements of Density and Heat Capacity for Binary Mixtures { x Benzonitrile + (1 – x) (Octane or Nonane)} †. *J. Chem. Eng. Data* **55** no. 10 (2010) 4154–4161.
- [127] E. Vercher, F. J. Llopis, M. V. González-Alfaro, A. Martínez-Andreu. Density, Speed of Sound, and Refractive Index of 1-Ethyl-3-methylimidazolium Trifluoromethanesulfonate with Acetone, Methyl Acetate, and Ethyl Acetate at Temperatures from (278.15 to 328.15) K. *J. Chem. Eng. Data* **55** no. 3 (2010) 1377–1388.
- [128] N. Deenadayalu, P. Bhujrajh. Density, Speed of Sound, and Derived Thermodynamic Properties of Ionic Liquids [EMIM] + [BETI] – or ([EMIM] + [CH 3 (OCH 2 CH 2) 2 OSO 3] – + Methanol or + Acetone) at T = (298.15 or 303.15 or 313.15) K. *J. Chem. Eng. Data* **53** no. 5 (2008) 1098–1102.
- [129] D. Zhu, D. Gao, H. Zhang, B. Winter, P. Lücking, H. Sun, H. Guan, H. Chen, J. Shi. Geometric Structures of Associating Component Optimized toward Correlation and Prediction of Isobaric Vapor–Liquid Equilibria for Binary and Ternary Mixtures of Ethanal, Ethanol, and Ethanoic Acid. *J. Chem. Eng. Data* **58** no. 1 (2013) 7–17.
- [130] W. A. Felsing, A. R. Thomas. Vapor Pressures and Other Physical Constants of Methylamine and Methylamine Solutions. *Ind. Eng. Chem.* **21** no. 12 (1929) 1269–1272.
- [131] R. H. Capps, W. M. Jackson. Density, Vapor Pressure And Heat Of Vaporization Of 2,2,3-Trichloro-Heptafluorobutane. *J. Phys. Chem.* **60** no. 6 (1956) 811–812.
- [132] A. Jouyban. *Handbook of solubility data for pharmaceuticals*, CRC Press, Taylor & Francis Group, Boca Raton, Fla. 2010.
- [133] V. Jaana, S. Nallani. THERMODYNAMIC AND TRANSPORT PROPERTIES OF BINARY LIQUID MIXTURES OF N-METHYLACETAMIDE WITH ALKYL (METHYL, ETHYL, n-PROPYL AND n-BUTYL) ACETATES AT 308.15 K. *Rasayan J. Chem.* **1** no. 3 (2008) 602–608.
- [134] Å. U. Burman, K. H. U. Ström. Density for (Water + Ethylenediamine) at Temperatures between (283 and 353) K. *J. Chem. Eng. Data* **53** no. 10 (2008) 2307–2310.
- [135] H. A. Zarei, M. Z. Lavasani, H. Iloukhani. Densities and Volumetric Properties of Binary and Ternary Liquid Mixtures of Water (1) + Acetonitrile (2) + Dimethyl Sulfoxide (3) at Temperatures from (293.15 to 333.15) K and at Ambient Pressure (81.5 kPa). *J. Chem. Eng. Data* **53** no. 2 (2008) 578–585.

- [136] T. Nitta, J. Fujio, T. Katayama. Solubilities of nitrogen in binary solutions. Mixtures of ethanol with benzene, ethyl acetate, and diethyl ether. *J. Chem. Eng. Data* **23** no. 2 (1978) 157–159.
- [137] J.-D. Ye, C.-H. Tu. Densities, Viscosities, and Refractive Indices for Binary and Ternary Mixtures of Diisopropyl Ether, Ethanol, and Methylcyclohexane. *J. Chem. Eng. Data* **50** no. 3 (2005) 1060–1067.
- [138] K. T. Thomas, R. A. McAllister. Densities of liquid-acetone-water solutions up to their normal boiling points. *AIChE J.* **3** no. 2 (1957) 161–164.

©2020 by the authors; licensee IAPC, Zagreb, Croatia. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>) 