Journal of Molecular Graphics and Modelling 118 (2023) 108312

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/jmgm

Force fields optimized against experimental data for large compound families using CombiFF: Validation considering non-target properties and polyfunctional compounds

Marina P. Oliveira, Philippe H. Hünenberger *

Laboratorium für Physikalische Chemie, ETH Zürich, ETH-Hönggerberg, HCI, CH-8093, Zürich, Switzerland

ARTICLE INFO

Keywords:
Molecular dynamics
Force-field
Parameter optimization
Validation

ABSTRACT

The CombiFF scheme is a workflow for the automated calibration of force-field parameters against condensed-phase experimental data considering simultaneously entire classes of organic molecules. The main steps of this scheme are: (*i*) selection of a molecule family; (*ii*) enumeration of all isomers; (*iii*) query for experimental data; (*iv*) automatic construction of the molecular topologies; (*v*) iterative refinement of the force-field parameters considering the entire family. In two recent articles, CombiFF was applied to the design of GROMOS-compatible united-atom force fields for the saturated acyclic haloalkanes and for saturated acyclic compounds involving eight common chemical functional groups of oxygen and nitrogen. This calibration and the subsequent initial validation involved two limitations: (*i*) the experimental data considered was restricted to values for the pure-liquid density ρ_{liq} and the vaporization enthalpy ΔH_{vap} of the compounds; (*ii*) beyond monofunctional compounds, the training set only involved homo-polyhaloalkanes (possibly mixing halogen types) in the first study, and homo-polyfunctional compounds of the considered oxygen or nitrogen functional groups (no mixing of different group types) in the second one. The goal of this article is to further test the accuracy of CombiFF-generated force fields by extending the validation to: (*i*) nine additional properties that were not used as optimization targets (pure-liquid thermodynamic, dielectric and transport properties, as well as solvation properties); (*ii*) hetero-polyfunctional molecules that were not included in the calibration and initial validation sets. The results for the nine additional properties show good agreement with experiment, except for the shear viscosity and the dielectric permittivity. There, larger discrepancies are observed, likely due to the united-atom representation adopted for the aliphatic groups and to the implicit treatment of electronic polarization effects. The results for the hetero-polyfunctional molecules also show reasonable agreement with experiment in terms of the monitored properties.

1. Introduction

The CombiFF scheme is a workflow recently introduced by our group for performing the automated calibration of force-field parameters against experimental condensed-phase data, considering entire classes of organic molecules constructed using a fragment library *via* combinatorial isomer enumeration [1,2] (Fig. 1). The main steps of the scheme are: (*i*) definition of a molecule family; (*ii*) enumeration of all isomers; (*iii*) query for experimental data; (*iv*) automatic construction of the molecular topologies by fragment assembly; (*v*) iterative refinement of the force-field parameters considering the entire family.

Until now, the CombiFF approach has been used to calibrate the non-bonded interaction parameters of GROMOS-compatible united-atom force fields for two families of molecules, based on target experimental data for the pure-liquid density ρ_{liq} and the vaporization enthalpy ΔH_{vap}. In Ref. 1, a force field was constructed for the saturated acyclic haloalkane family (HAL), considering 486 compounds and 749 data points for ρ_{liq} or ΔH_{vap} (calibration plus validation sets). The resulting errors relative to experiment in terms of root-mean-square deviation (RMSD) were 49.8 (27.6) $kg{\cdot}m^{-3}$ for ρ_{liq} and 2.7 (1.8) $kJ{\cdot}mol^{-1}$ for ΔH_{vap} considering the calibration (validation) set. In Ref. 2, parameters were refined for saturated acyclic compounds involving eight common chemical functional groups of oxygen and nitrogen. The corresponding family (O+N) is representative for the ether, aldehyde, ketone, ester,

* Corresponding author.
E-mail address: phil@igc.phys.chem.ethz.ch (P.H. Hünenberger).

https://doi.org/10.1016/j.jmgm.2022.108312
Received 17 June 2022; Received in revised form 9 August 2022; Accepted 16 August 2022
Available online 3 September 2022

alcohol, carboxylic acid, amine, and amide functional groups. This work relied on 1175 compounds and 1712 data points for ρ_{liq} or ΔH_{vap} (calibration plus validation sets). The resulting RMSD values relative to experiment were 29.9 (22.4) kg·m^{-3} for ρ_{liq} and 4.1 (5.5) kJ·mol^{-1} for ΔH_{vap} considering the calibration (validation) set.

For both the HAL and O+N families, the compounds considered involved one to ten carbon atoms. In the calibrations and subsequent initial validations, monofunctional as well as homo-polyfunctional compounds were considered, including up to four occurrences of the same chemical function within a given molecule (here, a halogenated compound is still considered homo-polyfunctional when the halogen atoms are of different types). However, neither hetero-polyfunctional compounds nor condensed-phase properties beyond ρ_{liq} and ΔH_{vap} were considered. The present article investigates the adequation of the CombiFF-generated parameters for extrapolative calculations, *i.e.* concerning compounds and properties not included in the force-field calibration and initial validation.

To this purpose, we consider 1771 molecules from the HAL and O+N families along with 685 molecules defining a hetero-polyfunctional (MIX) family that combines all types of functional groups of the two other families. For these 2456 compounds, eleven pure-liquid thermodynamic, dielectric and transport properties, as well as solvation properties, are calculated and compared to experiment. In addition to the parametrization targets ρ_{liq} and ΔH_{vap}, the nine other properties monitored are the surface-tension coefficient γ, the isothermal compressibility κ_T, the isobaric thermal-expansion coefficient α_P, the molar isobaric heat capacity c_P, the static relative dielectric permittivity ε, the self-diffusion coefficient D, the shear viscosity η, the hydration free energy ΔG_{wat}, and the free energy of solvation ΔG_{che} in cyclohexane.

2. Methodology

2.1. Compounds and data selection

The 2456 molecules considered are distributed into the HAL (490), O+N (1281), and MIX (685) families. The corresponding structures, along with identifiers and key properties, are provided in Suppl. Mat. Sec. S.1 (Figs. S1-S3 and Table. S1).

The collection of reference experimental data was performed similarly to the previous CombiFF applications [1,2]. First, the isomer-enumerator program [3] (ENU) was employed to list all possible constitutional isomers of up to ten carbon atoms. The generated SMILES strings were then used to search for other identifiers (molecule name, CAS registry number, and standard InChIKey) in the PubChem database [4] using its Application Programming Interface (API). Finally, the resulting identifiers served as a basis to collect reference experimental data from the database [1] (DBS) maintained by our group. This involved a query for the eleven properties over the three families. The main sources accessed *via* DBS were Refs. [5–25], and the corresponding experimental values are reported in Suppl. Mat. Sec. S.2 (Tables. S.2-S.12). This data is also provided in a freely downloadable format under Ref. [26].

2.2. Force field and optimized parameters

The potential-energy function used in this work is that of the GROMOS force field [27] in its 2016H66 variant [28]. The same applies to the standard simulation parameters [29]. One main difference, however, is that the atomic partial charges are not specified explicitly within topology building blocks for the various functional groups, but determined implicitly using an electronegativity-equalization (EE) scheme. Accordingly, atomic hardness and electronegativity values are specified instead of atomic partial charges. These parameters permit to take induction effects into account (*i.e.* they are expected to depend less on the covalent environment of an atom compared to the charges themselves) and enable the definition of neutral charge groups. For the HAL family, charge flows are only allowed along the carbon-halogen bonds. For the O+N family, the charge groups employed are shown in Figure 3 of Ref. [2]. The same conventions are used for the MIX family.

Given the use of a geometric-mean combination rule [30,31] for the Lennard-Jones (LJ) interactions [32], the pairwise LJ coefficients are constructed from two atomic parameters as well. This results in atom types that are characterized by an EE-type, determining the hardness η and the electronegativity χ, along with a LJ-type, specifying the collision diameter σ and the well depth ε. These force-field parameters were calibrated using the CombiFF approach as described in Ref. 1 (HAL) and

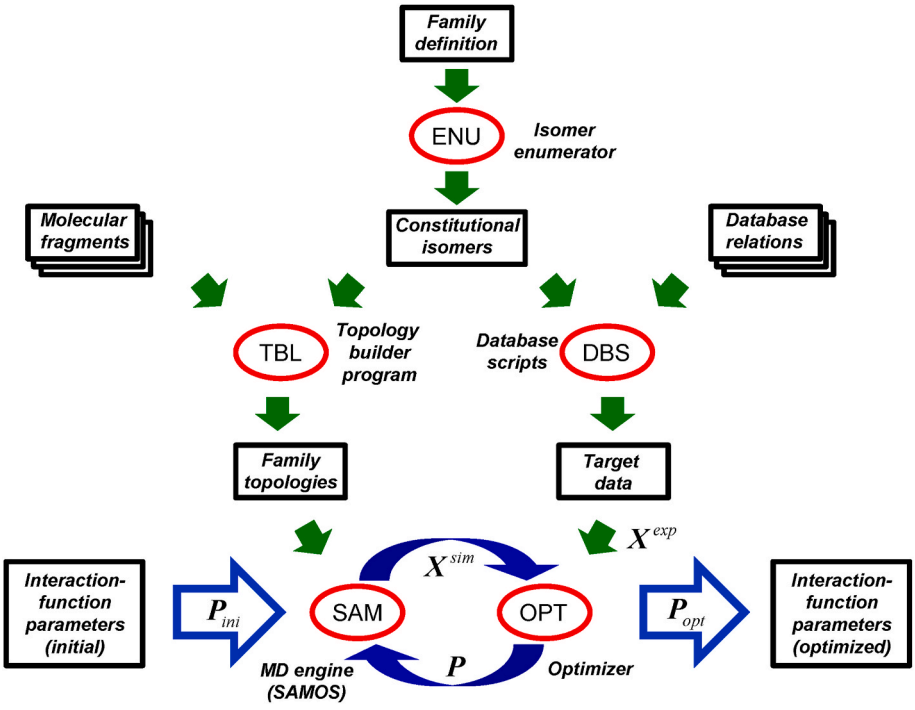

Fig. 1. CombiFF workflow for calibrating the parameters of a force field based on experimental data concerning a given compound family. Based on the definition of the family, the program ENU enumerates all possible constitutional isomers. The program TBL is then used to construct the molecular topologies of the corresponding compounds by linking molecular fragments, and the DBS scripts to extract available target experimental data pertaining to these molecules from an in-house database. By alternating simulations (MD engine SAM) to calculate the vector $\mathbf{X}^{sim}$ of simulated observables (as well as its derivative relative to all force-field parameters), and variations of the parameter vector $\mathbf{P}$ (optimization script OPT) designed to bring these simulated observables closer to their experimental target vector $\mathbf{X}^{exp}$, the initial parameters $\mathbf{P}_{ini}$ are progressively refined into optimal ones $\mathbf{P}_{opt}$.

Ref. 2 (O+N). For the ease of reference, the optimized values [1,2] are repeated in Suppl. Mat. Sec. S.3 (Tables. S.13-S.16). The molecular-topology files for GROMOS were generated automatically for all molecules based on their SMILES strings using the topology-builder program [1] (TBL). These files, along with equilibrated coordinate files, are provided in a freely downloadable format under Ref. [26].

2.3. Simulation protocols

All simulations were performed using the GROMOS program [33,34]. The pure-liquid simulations relied on molecular dynamics (MD) and were carried out under periodic boundary conditions based on cubic computational boxes containing 512 molecules. Unless otherwise specified, they were performed in the isothermal-isobaric ensemble at the reference pressures P and temperatures T listed in the tables of Suppl. Mat. Sec. S.2. The calculations of ΔG_{wat} and ΔG_{che} involved cubic periodic boxes containing one solute and 1000 water or 300 cyclohexane molecules. They were carried out at standard pressure $P^\circ = 1$ bar and temperature $T^- = 298.15$ K. In all cases, the temperature was maintained close to its reference value using a Nosé-Hoover thermostat [35] with a coupling time of 0.1 ps, and the pressure was maintained close to its reference value using a weak-coupling barostat [36] with a coupling time of 0.5 ps and an isothermal compressibility set to $4.575 \cdot 10^{-4}$ kJ^{-1} mol nm^3. The ideal-gas simulations (required for ΔH_{vap}) involved a single molecule and relied on stochastic dynamics [37–41] (SD) with a friction coefficient set to 2 ps^{-1}. They were performed at the same reference temperatures T as the corresponding pure-liquid simulations.

Newton's equation of motion (MD) or Langevin's equation of motion (SD) were integrated using the leap-frog scheme [37,42] with a timestep of 2 fs. Constraints on all bond lengths were enforced using the SHAKE procedure [43] with a relative geometric tolerance of 10^{-5}. The non-bonded interactions were calculated using a twin-range scheme [44] based on charge-group distances, with short- and long-range cutoff radii set to 0.8 and 1.4 nm, respectively, and an update frequency of 5 timesteps for the short-range pairlist and intermediate-range interactions. The mean effect of the omitted electrostatic interactions beyond the long-range cutoff was reintroduced using a reaction-field correction [45,46]. The corresponding static relative dielectric permittivities were set to the experimental permittivity ε for the pure-liquid simulations (Suppl. Mat. Table. S1), or to one for the ideal-gas simulations.

2.4. Property calculations

The calculation procedures for the eleven properties monitored are similar to those employed in previous work [29,30] and are only briefly summarized here. Three independent repeats were carried out for each analysis, involving different initial coordinates and velocities. The average value is reported, along with an uncertainty estimate provided by the corresponding error on the mean with a 95% confidence interval.

The density ρ_{liq} and vaporization enthalpy ΔH_{vap} were calculated based on pure-liquid and ideal-gas simulations carried out at the P,T-points listed in Suppl. Mat. Tables. S.2 and S.3, respectively. For ρ_{liq}, this only required a pure-liquid simulation at the indicated P,T-point. For ΔH_{vap}, this also required an ideal-gas simulation at the same temperature T. Each of the three simulation repeats in either phase involved 0.3 ns equilibration followed by 5 ns production. The value of ρ_{liq} was calculated from the pure-liquid simulation as the ratio of the mass of the computational box to the corresponding average volume. The value of ΔH_{vap} was calculated from the pure-liquid and ideal-gas simulations as the difference between the average potential energies per molecule in the two phases (gas minus liquid), expressed on a per-mole basis and increased by RT, where R is the gas constant.

The surface-tension coefficient was calculated at the P,T-points listed in Suppl. Mat. Table. S4. For each of the three repeats, the system was first equilibrated for 5 ns at constant pressure. The box was then extended by a factor 5 in the z-direction, thereby generating a system with two liquid-vacuum interfaces. The value of γ was calculated from a subsequent 5 ns constant-volume simulation as

$$\gamma = \frac{L_z}{2}\langle P_{zz} - \frac{1}{2}(P_{xx} + P_{yy})\rangle, \tag{1}$$

where L_z is the box length in the z-direction and $P_{\alpha\alpha}$ ($\alpha = x, y, z$) are the diagonal elements of the pressure tensor.

The isothermal compressibility κ_T was calculated at the P,T-points listed in Suppl. Mat. Table. S5. For each of the three repeats, the system was first equilibrated for 5 ns at constant pressure. Constant-volume simulations were then performed at $\rho_1 = \rho_{liq} - 10$ kg·m^{-3} and $\rho_2 = \rho_{liq} + 10$ kg·m^{-3}, each involving 0.25 ns equilibration followed by 5 ns production. The value of κ_T was calculated based on these two simulations *via* finite difference, as

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \approx \frac{\ln \rho_2 - \ln \rho_1}{\langle P\rangle_{\rho_2} - \langle P\rangle_{\rho_1}}. \tag{2}$$

The isobaric thermal-expansion coefficient α_P was calculated at the P,T-points listed in Suppl. Mat. Table. S6. For each of the three repeats, constant-pressure simulations were performed at $T_1 = T - 10$ K and $T_2 = T + 10$ K, each involving 0.25 ns equilibration followed by 5 ns production. The value of α_P was calculated based on these two simulations *via* finite difference, as

$$\alpha_P = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P \approx -\frac{\ln \langle\rho\rangle_{T_2} - \ln \langle\rho\rangle_{T_1}}{T_2 - T_1}. \tag{3}$$

The molar isobaric heat capacity c_P was calculated at the P,T-points listed in Suppl. Mat. Table. S7. For each of the three repeats, it was determined based on the same simulations as for α_P *via* finite difference, as

$$c_P = \left(\frac{\partial H}{\partial T}\right)_P \approx \left(\frac{\partial U}{\partial T}\right)_P \approx \frac{\langle U\rangle_{T_2} - \langle U\rangle_{T_1}}{T_2 - T_1}, \tag{4}$$

where U is the total energy per molecule, expressed on a per-mole basis. The contribution of the pressure-volume term was neglected and no quantum corrections were applied.

The static relative dielectric permittivity ε was calculated at the P,T-points listed in Suppl. Mat. Table. S8. For each of the three repeats, ε was obtained from a 50 ns constant-pressure simulation using the Neuman relation [47].

$$\varepsilon = \frac{3\varepsilon_o(2\varepsilon_{RF}+1)RT\langle V\rangle + \varepsilon_{RF}\left[\langle \mathbf{M}^2\rangle - \langle \mathbf{M}\rangle^2\right]}{3\varepsilon_o(2\varepsilon_{RF}+1)RT\langle V\rangle - \left[\langle \mathbf{M}^2\rangle - \langle \mathbf{M}\rangle^2\right]}, \tag{5}$$

where $\mathbf{M}$ is the box dipole-moment vector, ε_o is the permittivity of vacuum, and ε_{RF} is the reaction-field permittivity (Suppl. Mat. Table. S1).

The self-diffusion coefficient D was calculated at the P,T-points listed in Suppl. Mat. Table. S9. For each of the three repeats, it was evaluated based on a 5 ns constant-pressure simulation from the mean-square displacement of the molecules using the Einstein relation [48].

$$D = \lim_{t\to\infty} \frac{\langle(\mathbf{r}_i(\tau + t) - \mathbf{r}_i(\tau))^2\rangle_{i,\tau}}{6t}, \tag{6}$$

where $\mathbf{r}_i$ is the instantaneous position of the center of geometry of molecule i (following molecules across periodic boundaries) and the average is performed over all molecules i and time origins τ. The estimation of D relied in practice on a least-squares fit over the interval from 0 to 3 ns.

The shear viscosity η (also called dynamic or absolute viscosity) was calculated at the P,T-points listed in Suppl. Mat. Table. S10. For each of the three repeats, it was evaluated based on a 2 ns constant-volume simulation (with pressure tensor saved to file every timestep) using

the Einstein-type relation

$$\eta = \frac{V}{2RT}\lim_{t\to\infty}\frac{d}{dt}\langle \Delta P^2_{\alpha\beta}(\tau,\tau+t)\rangle_\tau, \quad (7)$$

where V is the box volume and $\Delta P^2_{\alpha\beta}$ is the square displacement of the off-diagonal components of the pressure tensor

$$\Delta P_{\alpha\beta}(\tau,\tau+t) = \int_\tau^{\tau+t} dt' P_{\alpha\beta}(t'). \quad (8)$$

The estimation of η relied in practice on a least-squares fit over the interval from 5 to 20 ps.

Finally, the calculation of the hydration free energy ΔG_{wat} and of the solvation free energy ΔG_{che} in cyclohexane at $P^\circ = 1$ bar and $T^- = 298.15$ K followed the same protocol as in Ref. [28]. The solute-solvent interactions were perturbed using a coupling parameter λ based on a soft-core scheme [49], where $\lambda = 0$ corresponds to full and $\lambda = 1$ to absent solute-solvent interactions. The electrostatic and Lennard-Jones soft-core parameters were set to 0.5 nm^2 and 0.5, respectively. For both solvents and for each of the three independent repeats, the solvation free energies were calculated using thermodynamic integration [50] based on 21 equidistant λ-points and a trapezoidal integration. In each repeat, the initial configuration at $\lambda = 0$ was equilibrated for 0.4 ns, and the configuration obtained after 50 ps of the simulation at a given λ-point was used as the initial configuration for the simulation at the next λ-point. For each λ-value, the system was equilibrated for 0.5 ns followed by a production simulation of 15 ns.

3. Results and discussion

The correlation between experimental values and simulation results in terms of the eleven properties monitored are displayed in Figs. 2 and 3. The corresponding numerical values are reported in Suppl. Mat. Sec. S.2 (Tables. S.2-S.12). The statistics in terms of root-mean-square deviation (RMSD) and average deviation (AVED) are presented in Table 1 for the entire set of molecules (ALL), as well as separately over the HAL, O+N, and MIX families separately. The statistics sorted by chemical function is illustrated in Figs. 4–6. The corresponding numerical data can be found in Suppl. Mat. Sec. S.4 (Tables. S.17-S.20).

In terms of overall agreement between simulation and experiment, one may roughly distinguish three main categories of observables: (*i*) for ρ_{liq}, the agreement is generally excellent; (*ii*) for ΔH_{vap}, γ, κ_T, α_P, c_P, D, ΔG_{wat} and ΔG_{che}, the agreement is good; (*iii*) for ε and η, there are more pronounced and systematic discrepancies. In the following, the eleven properties are discussed in turn. Comparison is also made with the results of two previous investigations reporting the calculation of the same properties over large sets of compounds: (*i*) the validation of the GROMOS-compatible 2016H66 parameter set based on 62 organic molecules [28] (see also Ref. [29]); (*ii*) the benchmark evaluation of the AMBER-GAFF [51,52] and OPLS-AA [53] force fields based on 146 organic molecules [14].

Considering ρ_{liq}, which was a target property in the calibration [1,2], the force-field parameters lead to an overall RMSD of 45.6 kg·m^{-3} (4.4%) and a corresponding AVED of −5.4 kg·m^{-3}. The RMSD value mostly results from comparatively large deviations associated with a few outliers. The RMSD and AVED values are of somewhat higher magnitudes for the molecules in the MIX family compared to the two other families, but the agreement with experiment still remains very good in this case. Thus, in terms of ρ_{liq}, the transferability of the force-field parameters from monofunctional and homo-polyfunctional compounds to hetero-polyfunctional ones appears to represent a valid assumption.

Considering ΔH_{vap}, which was also a target property in the calibration [1,2], a number of compounds with large deviations result in an overall RMSD of 8.1 kJ·mol^{-1} (17.0%) and an AVED of 2.7 kJ·mol^{-1}. The positive AVED suggests is a slight systematic trend towards an overestimation of ΔH_{vap}. The discrepancies increase when the parameters are transferred from the monofunctional and homo-polyfunctional compounds to hetero-polyfunctional ones. The largest discrepancies are observed for the combinations of chemical functions F+RCOH, ROR+RCOOH, RCOR+RCOOH, RCOOR+ROH and RCOOH+NR, and are particularly pronounced for molecules involving carboxylic-acid functional group. The lower accuracy for hetero-polyfunctional molecules may be at least in part due to the following reasons: (*i*) the EE scheme applied with the constraint of forming neutral charge-groups may fail to entirely capture induction effects in these molecules; (*ii*) the assignment of dihedral-angle force-field parameters using the closest available types may not always be appropriate, in particular at the junction between two distinct functional groups; (*iii*) there may be higher inaccuracies in the reference experimental data for these more complex molecules.

For the surface-tension coefficient γ, an excellent agreement between calculated and experimental values is observed, with a RMSD of 6.4 mN·m^{-1} (23.8%) and an AVED of −1.5 mN·m^{-1} (for comparison, the experimental γ values range between 12.6 and 133.2 mN·m^{-1}). The deviations are homogeneously distributed across the different functional groups, with only a few exceptions. The observation of an excellent agreement for γ certainly results from the strong connection between this observable and the optimization target ΔH_{vap}, both observables characterizing the cohesive forces within the pure liquid. Indeed a high correlation between ΔH_{vap} and γ, as well as between the associated

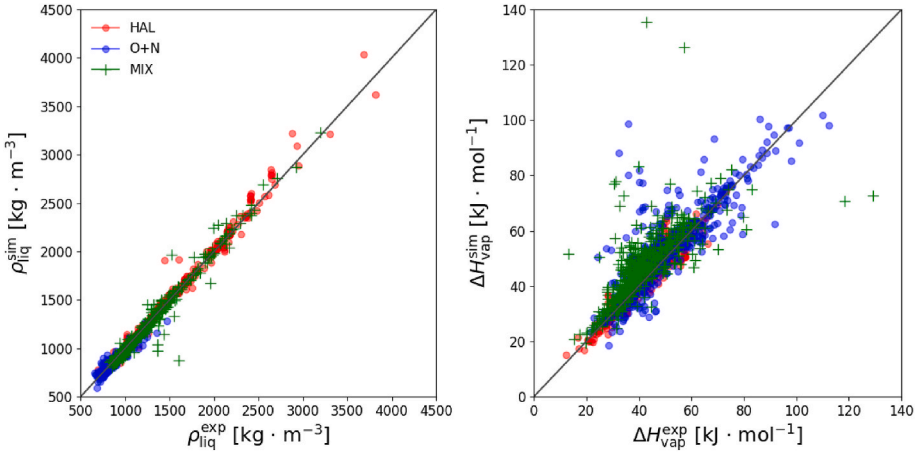

Fig. 2. Simulation results (sim) versus experimental values (exp) for the parametrization targets ρ_{liq} (left) and ΔH_{vap} (right). The correlations are shown for compounds of the HAL, O+N, and MIX families. The corresponding numerical values can be found in Suppl. Mat. Tabs. S.2 and S.3, and the associated statistical information in Table 1. A similar correlation for the nine other properties can be found in Fig. 3. The diagonal solid lines indicate perfect agreement with experiment.

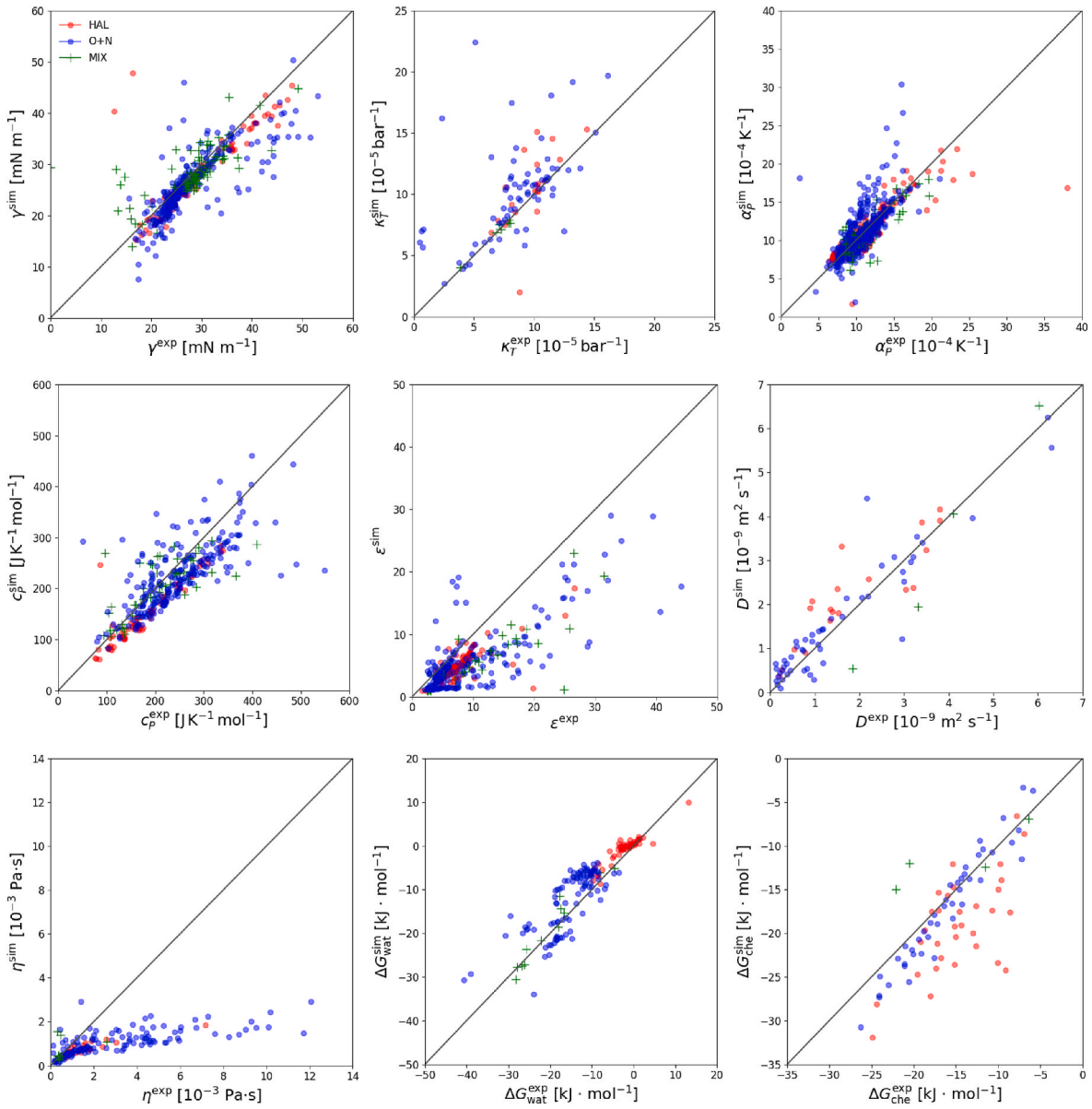

Fig. 3. Simulation results (sim) versus experimental values (exp) for the nine properties that were not included as parametrization targets. The correlations are shown for compounds of the HAL, O+N, and MIX families. The properties are the surface-tension coefficient γ, the isothermal compressibility κ_T, the isobaric thermal-expansion coefficient α_P, the molar isobaric heat capacity c_P, the static relative dielectric permittivity ε, the self-diffusion coefficient D, the shear viscosity η, the hydration free energy ΔG_{wat}, and the solvation free energy ΔG_{che} in cyclohexane. The corresponding numerical values can be found in Suppl. Mat. Tabs. S.4-S.12 and the associated statistical information in Table 1. A similar correlation for ρ_{liq} and ΔH_{vap} can be found in Fig. 2. The diagonal solid lines indicate perfect agreement with experiment. The representative points of a few molecules are not visible in the graphs due to large deviations (See Suppl. Mat. Sec. S.2 for the corresponding values). These are: I1201, L3301 and D3401 for γ, N2101 for κ_T, E7212 for c_P, D1201, M2201, M3201, M4202, M4203 and M6203 for ε, as well as L2201, L3201, L3202, L4202, L4204, L4205, L4206, L5209, L5211, L6201, L6223, L6225, L7214, L8220, L8210, L9210, L0213, L0210, L3301 and L6302 for η.

force-field errors, has been reported previously [14,28,29], and γ has also been used as an alternative target to ensure an appropriate energetics in force-field calibration [54]. It has been suggested [14,28,55–57] that the application of a long-range LJ correction or the use of a lattice-sum LJ interaction scheme are required for the accurate reproduction of γ, and that the inclusion of explicit polarization is essential for such interfacial properties. However, the results obtained here as well as in Ref. [28] indicate that this is not necessarily true, and that a careful parametrization against ΔH_{vap} may be sufficient to produce C_6 dispersion coefficients that appropriately compensate in a mean-field fashion for the omitted long-range LJ and electronic-polarization effects. Still, for a few molecules, the deviations in terms of γ are somewhat larger, namely fluoroalkanes, iodoalkanes, hetero-haloalkanes, amides, and molecules with two amino groups. For a few molecules (F1301, F1401, A5201, L4301, D3401, N3101, S2502, S1402, S2310, S1201) the deviations become very high (above 50%), and similarly to the results of Ref. [28], deviations of about 20–30% are observed for the molecules L4101, N2101 and M2201 (with codes 2M2P, EAN and AMD, respectively, in Ref. [28]).

Considering the isothermal compressibility κ_T, the calculated results are in reasonable agreement with the experimental values, with a RMSD of 3.5 · 10^{-5} bar^{-1} (34.9%) and an AVED of 1.4 · 10^{-5} bar^{-1} (for comparison, the experimental κ_T values range between 0.5 and 16.1 · 10^{-5} bar^{-1}). The RMSD values for AMBER-GAFF and OPLS-AA in Ref. [14] and for GROMOS-2016H66 in Ref. [28] are on the order of 2.8–3.0 · 10^{-5} bar^{-1}. As indicated by the positive AVED, the present

Table 1
Statistics concerning the discrepancies between simulated and experimental values for the different properties, considering the molecules of the HAL, O+N, and MIX families. The root-mean-square deviation (RMSD), along with the RMSD value normalized by the mean of the simulation results and expressed in percent, and the average deviation (AVED) are reported for the N data points in terms of eleven properties. These properties are: the pure-liquid density ρ_{liq}, the vaporization enthalpy ΔH_{vap} the surface-tension coefficient γ, the isothermal compressibility κ_T, the isobaric thermal-expansion coefficient α_P, the isobaric heat capacity c_P, the static relative dielectric permittivity ε, the self-diffusion coefficient D, the shear viscosity η, the hydration free energy ΔG_{wat}, and the solvation free energy ΔG_{che} in cyclohexane. The different molecule sets considered are the halohydrocarbons (HAL), the compounds involving functional groups of oxygen and nitrogen (O+N), the compounds containing simultaneously different functional groups from the HAL and O+N families (MIX), and the entire set (ALL). The corresponding correlations are displayed in Figs. 2 and 3.

Group	N	RMSD	%Err	AVED	N	RMSD	%Err	AVED	N	RMSD	%Err	AVED
	ρ_{liq} [kg·m^{-3}]				ΔH_{vap} [kJ·mol^{-1}]							
HAL	468	53.6	3.9	11.9	313	3.4	7.9	−0.1				
O+N	1087	27.8	3.3	−9.4	725	7.5	15.0	2.1				
MIX	461	66.1	5.5	−13.3	524	10.5	22.3	5.3				
ALL	2016	45.6	4.4	−5.4	1562	8.1	17.0	2.7				
	γ [mN·m^{-1}]				κ_T [10^{-5} bar^{-1}]				α_P [10^{-4}K^{-1}]			
HAL	89	5.7	19.8	−1.5	21	2.4	23.3	0.7	305	1.7	16.3	−0.3
O+N	279	7.2	27.7	−2.1	71	3.9	38.3	1.7	735	1.7	16.9	−0.0
MIX	106	4.7	16.3	0.1	7	0.2	3.3	0.0	84	1.6	14.3	−0.3
ALL	474	6.4	23.8	−1.5	99	3.5	34.9	1.4	1124	1.7	16.5	−0.1
	c_P [J·K^{-1}·mol^{-1}]				ε				D [10^{-9} m^2·s^{-1}]			
HAL	72	41.3	27.8	−31.0	103	3.8	77.6	−2.8	19	0.7	30.6	0.4
O+N	229	68.4	30.3	−22.5	222	18.7	315.2	−6.3	50	0.5	32.7	0.1
MIX	54	48.0	24.2	−4.5	34	7.1	122.5	−5.3	4	1.0	29.9	−0.5
ALL	355	60.9	29.6	−21.5	359	15.0	267.2	−5.2	73	0.6	32.4	0.1
	η [mPa·s]				ΔG_{wat} [kJ·mol^{-1}]				ΔG_{che} [kJ·mol^{-1}]			
HAL	51	0.9	163.8	−0.5	50	2.2	-	1.4	32	6.0	-	−4.4
O+N	176	109.7	6477.2	−17.4	100	5.3	-	2.9	43	2.3	-	−1.0
MIX	9	0.7	102.6	0.1	12	2.4	-	0.9	4	5.6	-	3.6
ALL	236	94.7	6704.1	−13.1	162	4.4	-	2.3	79	4.4	-	−2.1

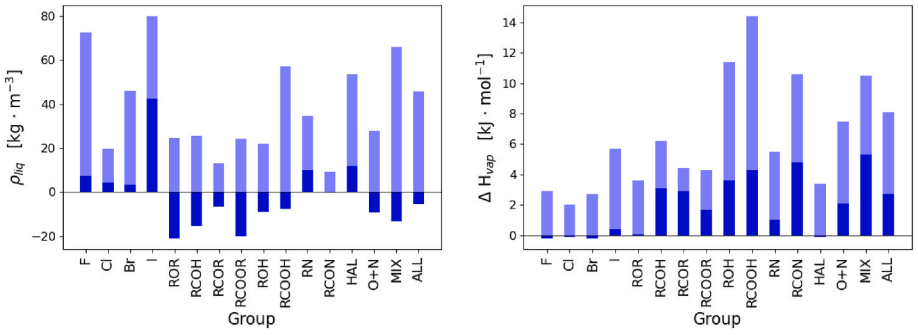

Fig. 4. Statistics concerning the root-mean-square deviation (RMSD, semi-transparent bar) and average deviation (AVED, solid bar) per group of molecules in terms of ρ_{liq} (left) and ΔH_{vap} (right). The corresponding numerical values can be found in Table 1 and Suppl. Mat. Table. S17.

simulations tend to somewhat overestimate the value of κ_T, in particular for ethers, aldehydes, carboxylic acids, and amines. In practice, the balance between density and compressibility in a simulated liquid is largely related to the functional form selected for the repulsive component of the van der Waals interactions. In previous work [14,28], good agreement with experiment in terms of both ρ_{liq} and κ_T suggested that the inverse-twelfth power assigned to the repulsive component of the LJ function [32] is adequate to reproduce both properties. As discussed in Ref. [28], this was an interesting observation, considering that the latter choice is *ad hoc* and has sometimes been considered too steep, some force fields relying on an inverse-ninth power instead [30,31,58–61]. The overestimation of κ_T observed here is thus somewhat surprising, especially because if the inverse-twelfth power was too steep, this would be expected to reduce the magnitude of the compressibility rather than increase it.

Considering the isobaric thermal-expansion coefficient α_P, the overall agreement between calculated and experimental values is reasonable, with a RMSD of $1.7 \cdot 10^{-4}$ K^{-1} (16.5%) and an AVED of $-0.1 \cdot 10^{-4}$ K^{-1} (for comparison, the experimental α_P values range between 2.5 and $38.1 \cdot 10^{-4}$ K^{-1}). The RMSD values for AMBER/GAFF and OPLS/AA in Ref. [14], and for GROMOS-2016H66 in Ref. [28], are on the order of $3.0–4.4 \cdot 10^{-4}$ K^{-1}. Here, fluoroalkanes tend to present much higher RMSD values compared to the other compounds. Similarly to the results reported in Ref. [28], the α_P values are overestimated by 20–50% for L1101, D2201, N2201 and M3201 (the corresponding codes in Ref. [28] are MTL, ACA, EDAN and LNMA). However, the α_P value for M3201 is overestimated by about a factor of four in Ref. [28], whereas the error is reduced here to 19%. On the other hand, the error for D2201 reported in Ref. [28] is 31%, while the error observed here is 44%.

In terms of c_P, large deviations are observed, with a RMSD of 60.9 J K^{-1} mol^{-1} (29.6%) and an AVED of −21.5 J K^{-1} mol^{-1} (for comparison, the experimental c_P values range between 50.6 and 744.6 J K^{-1} mol^{-1}). Similarly to the results reported in Ref. [28], in most cases, c_P is underestimated. The exceptions are fluoroalkanes, carboxylic acids, amines and amides, with deviations that are positive or close to zero. These discrepancies can at least in part be attributed to the absence of quantum corrections for values calculated at the classical level [14,62–64]. Concerning this issue, it is interesting to compare the present observations and those of Ref. [28] (see Figure 3 therein) to those on Ref. [14] (see Figure 7 therein), where the classical molar heat-capacities calculated using the all-atom force fields AMBER-GAFF and OPLS-AA were found to systematically overestimate the

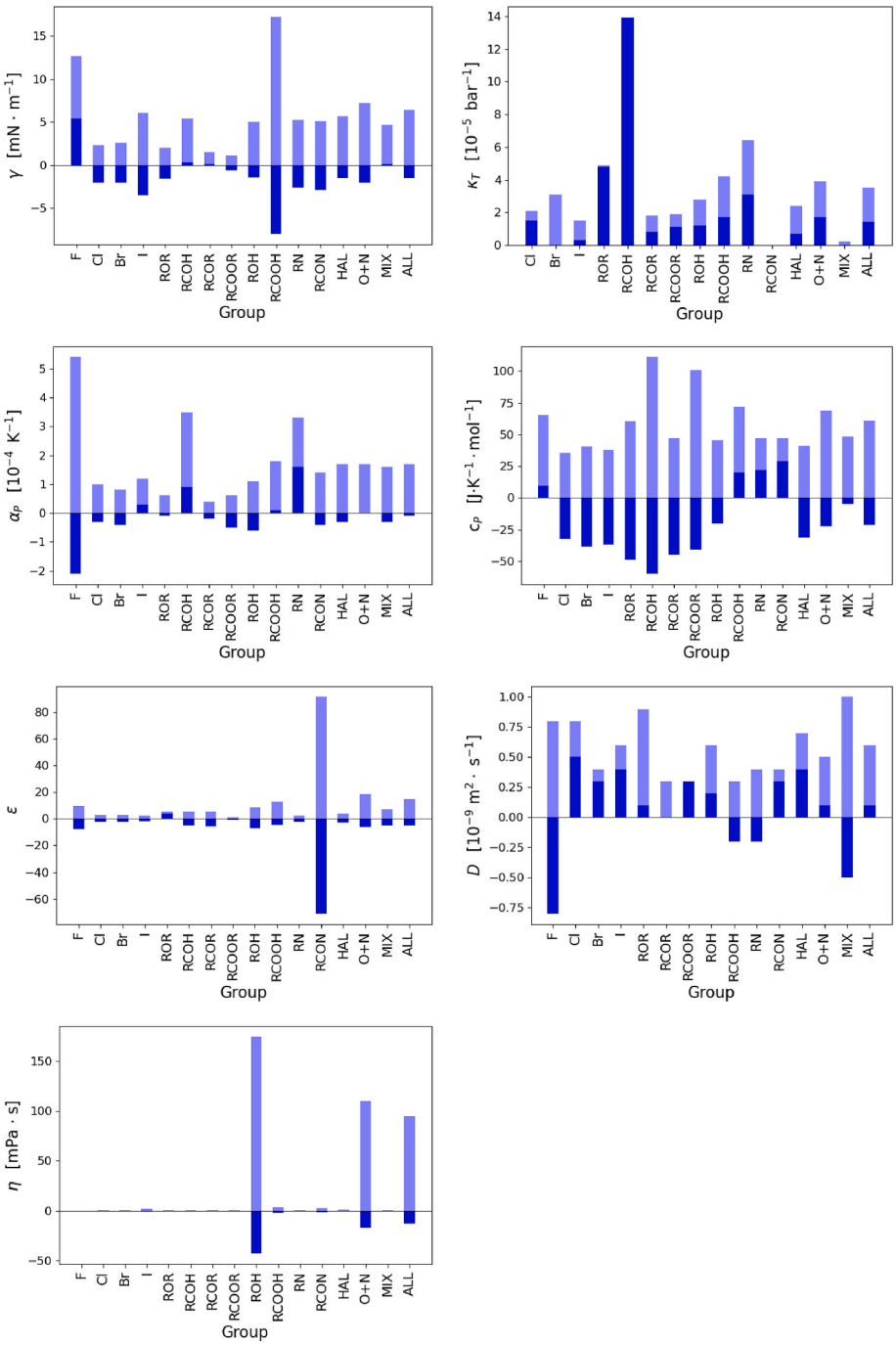

Fig. 5. Statistics concerning the root-mean-square deviation (RMSD, semi-transparent bar) and average deviation (AVED, solid bar) per group of molecules in terms of γ, κ_T, α_P, c_P, ε, D and η. The corresponding numerical values can be found in Table 1 and in Suppl. Mat. Tabs. S.18, S.19 and S.20.

experimental values by about a factor of two. For the two all-atom force fields, only when applying quantum corrections did the results become closer to experiment, with RMSD values in the range of 10.4–19.8 J K^{-1} mol^{-1}. Using united-atoms eliminates the contribution of high-frequency vibrations involving the aliphatic hydrogen atoms, leading to more realistic classical c_P estimates. This interpretation is in line with the overestimation observed in Ref. [28] for the aromatic compounds in GROMOS-2016H66, for which the hydrogen atoms are explicit. As pointed out in this work [28], although classical heat capacities may be of moderate relevance in terms of a detailed comparison with experimental data in the absence of quantum corrections, they nevertheless determine the temperature derivatives of the thermodynamic properties within the force field. In this sense, qualitative agreement with experiment suggests a good transferability of the force-field parameters to temperatures differing from room temperature (within reasonable bounds), which is an advantage of the united-atom over an all-atom representation.

In terms of the dielectric permittivity ε, the overall RMSD is 15.0 and the AVED is −5.2 (for comparison, the experimental ε values range between 1.6 and 172.2). The deviations are particularly small for haloalkanes, likely due to the low polarity of these compounds. Among these, the fluoroalkanes show the largest deviations. Conversely, alcohols, carboxylic acids, and amides are the most polar compounds, and show the most pronounced deviations. For nearly all the molecules considered here, the calculated permittivity value is underestimated, with the exception of ethers. The underestimation is particularly severe

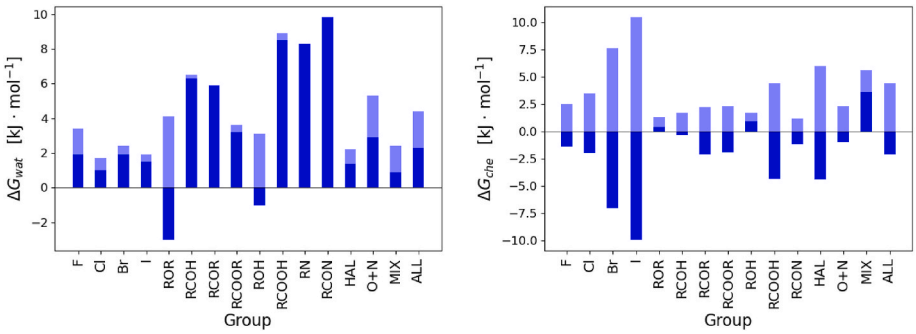

Fig. 6. Statistics concerning the root-mean-square deviation (RMSD, semi-transparent bar) and average deviation (AVED, solid bar) per group of molecules in terms of ΔG_{wat} and ΔG_{che}. The corresponding numerical values can be found in Table 1 and Suppl. Mat. Table. S20.

for amides, similarly to what was observed previously in Ref. [28]. As discussed in this article [28], the representation of amides and, in particular, of their dielectric properties within classical force fields is a challenging problem that has attracted considerable attention over the years [65–70]. The origin of the amide peculiarity has been attributed by different authors to the formation of hydrogen-bonded microclusters or chains [71,72] stabilized by strong and cooperative [73] NH··· O hydrogen bonds (except for *N*-disubstituted amides) as well as weaker CH··· O interactions [67] ("improper" hydrogen bonds [65]), or to the large magnitude of electronic polarizability effects for this chemical function [68,70]. Note that, the dielectric permittivity is the most difficult property to reproduce, in large part due to the neglect of electronic-polarization effects and to the slow convergence of this property with simulation time, a feature enhanced by the high viscosity of some molecules [14] such as alcohols. Despite all these issues, the results obtained with the present force field give a RMSD on the same order as the results with the AMBER-GAFF and OPLS-AA force fields [14] (7.6–15.9).

In terms of the transport properties, good agreement is observed for D, whereas pronounced and systematic discrepancies are observed for η. The overall RMSD for D is $0.6 \cdot 10^{-9}$ $m^2 \cdot s^{-1}$ (32.4%) and the AVED is $0.1 \cdot 10^{-9}$ $m^2 \cdot s^{-1}$ (for comparison, the experimental D values range between 0.12 and 6.3 10^{-9} $m^2 \cdot s^{-1}$). The corresponding values for η are 94.7 and −13.1 mPa·s (for comparison, the experimental η values range between 0.1 and 1095 mPa·s). For the latter quantity, the correlation between experimental and calculated values only follows a straight line at low viscosity, and the slope becomes significantly lower than one as the experimental viscosity increases. The viscosities of alcohols, carboxylic acids and amides tend to be strongly underestimated, as was observed in Ref. [28]. No results for D and η were reported in Ref. [14].

Considering ΔG_{wat}, the simulation results lead to an overall RMSD of 4.4 $kJ \cdot mol^{-1}$ and an AVED of 2.3 $kJ \cdot mol^{-1}$ (for comparison, the experimental ΔG_{wat} values range between − 40.6 and 13.2 $kJ \cdot mol^{-1}$). The discrepancies are distributed rather heterogeneously across the different chemical functions. The haloalkanes show the smallest deviations in terms of RMSD, and carboxylic acids, amines, and amides show the highest deviations. The hydration free energy is overestimated for most of the chemical functions, in particular for aldehydes, ketones, acids, amines and amides, while it is underestimated for ethers. Similar results are reported in Ref. [28], where ΔG_{wat} is systematically overestimated for the amides and, to a lesser extent, for the amines.

Considering ΔG_{che}, the simulation results lead to a RMSD of 4.4 $kJ \cdot mol^{-1}$ and an AVED of −2.2 $kJ \cdot mol^{-1}$ (for comparison, the experimental ΔG_{che} values range between − 26.3 and 0.2 $kJ \cdot mol^{-1}$). For this quantity, the force-field parameters yield results in good agreement with experiment for most types of molecules, with the exception of haloalkanes involving bromine and iodine. The deviations are less systematic than for ΔG_{wat}. Except for the problematic haloalkane cases, the average deviations are on the order of the thermal energy, with a slight tendency towards underestimated values. In comparison to the results reported in Ref. [28], similar values are obtained here, and the deviations are always below 4 $kJ \cdot mol^{-1}$.

In the calibration of the force-field parameters for the HAL and O+N families, only two types of target properties were considered, ρ_{liq} and ΔH_{vap}. Interactions with water or cyclohexane, as probed by ΔG_{wat} and ΔG_{che}, were not considered. It was expected that the quality of the reoptimized force field with respect to these two properties would be at least in part preserved relative to the 2016H66 set [28], used as a starting point. In view of the results obtained here, this assumption largely holds, as the deviations relative to experiment are comparable between the 2016H66 and the reoptimized set. Additionally, the accuracy of the present parameter set in terms ΔG_{wat} (RMSD of 4.4) remains comparable to that of the AMBER-GAFF force field [74,75] (4–5 $kJ \cdot mol^{-1}$), which is often considered representative of the accuracy that can be expected for this property from the best current physical models [75].

The structures of the main outliers for the different properties are shown in Suppl. Mat. Sec. S.5 (Figs. S4-S14). The most common types of molecules with the highest discrepancies are alcohols, carboxylic acids, and compounds in the MIX family.

4. Conclusion

The goal this article was to test the force fields developed in Refs. [1,2] using the CombiFF approach for the HAL and O+N families. More precisely, the appropriateness of using these parameters in an extrapolative fashion was investigated, *i.e.* for compounds and properties not included in the force-field calibration and initial validation.

Pure-liquid thermodynamic, dielectric, and transport properties, as well as solvation properties, were calculated for 2456 molecules of the HAL, O+N, and MIX families, and compared to experiment. In addition to ρ_{liq} and ΔH_{vap}, the nine properties considered are: the surface-tension coefficient γ, the isothermal compressibility κ_T, the isobaric thermal-expansion coefficient α_P, the isobaric heat capacity c_P, the static relative dielectric permittivity ε, the self-diffusion coefficient D, the shear viscosity η, the hydration free energy ΔG_{wat}, and the free energy of solvation ΔG_{che} in cyclohexane.

The final RMSD values for these eleven properties over the entire set of compounds are 45.6 $kg \cdot m^{-3}$, 8.1 $kJ \cdot mol^{-1}$, 6.4 $mN \cdot m^{-1}$, $3.9 \cdot 10^{-5}$ bar^{-1}, $1.7 \cdot 10^{-4}$ K^{-1}, 60.9 J K^{-1} mol^{-1}, 15.0, 0.6 10^{-9} $m^2 \cdot s^{-1}$, 94.7 mPa·s, 4.4 $kJ \cdot mol^{-1}$ and 4.4 $kJ \cdot mol^{-1}$, respectively. In terms of overall agreement between simulation and experiment, one may roughly distinguish three main categories of observables: (*i*) for ρ_{liq}, the agreement is generally excellent; (*ii*) for ΔH_{vap}, γ, κ_T, α_P, c_P, D, ΔG_{wat} and ΔG_{che}, the agreement is good; (*iii*) for ε and η, there are more pronounced and systematic discrepancies. The observations made in the present study considering the deviations relative to experiment are qualitatively (trends) and quantitatively (magnitudes) comparable to those made previously using the AMBER-GAFF and OPLS-AA force fields[14,74,75] as well as the GROMOS-2016H66 force field [28].

A high level of agreement for ρ_{liq} and ΔH_{vap} is expected considering that these were the target properties in the parameter optimization. The results in terms of ρ_{liq} are in excellent agreement with the reference data, apart from a few outliers. However, for ΔH_{vap} somewhat larger discrepancies are observed. In practice, the degree of transferability of these force-field parameters may be limited, probably in large part due to shortcomings of the EE scheme with charge-group constraints, the unavailability of appropriate torsional interaction parameters, and possible problems with the reference data for these more complex molecules.

The high agreement for γ largely follows from its high correlation with ΔH_{vap}, both observables probing the magnitude of the cohesive forces within the liquid. Good agreement for κ_T largely follows from the excellent reproduction of ρ_{liq} given an appropriate choice for the repulsive part of the van der Waals interactions. For α_P, the agreement is good as well, and deviations are mostly due to outliers. Concerning c_P, a small degree of underestimation is observed, related to the absence of quantum corrections. Note, however, that all-atom models lead to positive deviations of significantly larger magnitudes [14]. The quantity ε is the most difficult to reproduce (slow convergence, lack of polarizability, effect of viscosity). The agreement with experiment remains nevertheless acceptable. Finally, the analysis of D and η reveals a slightly enhanced diffusivity (a well-known feature of united-atom models [28]), along with a pronounced underestimation of the viscosity.

Work is in progress to expand the CombiFF calibration to other chemical families. In addition, liquid properties other that the density and vaporization enthalpy could be used not only for validation purposes (as done here) but directly during the calibration (as well as properties concerning the gas and solid phases). The inclusion of polyfunctional compounds in the calibration is also likely to enhance the transferability of the resulting force fields. From a methodological perspective, a number of force-field representation choices may also have to be addressed in future work, and in particular: (*i*) the parameters that were not optimized (in the first place, the torsions); (*ii*) the level of details of the selected atom-type set; (*iii*) the choice of a model resolution, i.e. united-*vs.* all-atom; (*iv*) the restriction of EE charge flows to small neutral charge groups. Efforts are currently in progress along these different lines. A possible further development of CombiFF would be the design of a polarizable force field of the fluctuating-charge type.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have shared the data in the supplementary material.

Acknowledgment

Financial support by the Swiss National Science Foundation (Grant No.200021-175944) is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmgm.2022.108312.

References

[1] M.P. Oliveira, M. Andrey, S.R. Rieder, L. Kern, D.F. Hahn, S. Riniker, B.A.C. Horta, P.H. Hünenberger, Systematic optimization of a fragment-based force-field based on experimental pure-liquid properties considering large compound families: application to the saturated haloalkanes, J. Chem. Theor. Comput. 16 (2020) 7525.
[2] M.P. Oliveira, P.H. Hünenberger, Systematic optimization of a fragment-based force-field against experimental pure-liquid properties considering large compound families: application to oxygen and nitrogen compounds, Phys. Chem. Chem. Phys. 23 (2021), 17774.
[3] S.R. Rieder, M.P. Oliveira, S. Riniker, P.H. Hünenberger, Development of an open-source software for isomer enumeration, J. Cheminf. (2022).
[4] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B.A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E.E. Bolton, PubChem 2019 update: improved access to chemical data, Nucleic Acids Res. 47 (2019) D1102.
[5] M. Frenkel, X. Hong, R.C. Wilhoit, K.R. Hall, In Thermodynamic properties of organic compounds and their mixtures, in: K.R. Landolt-Börnstein Series - Hall, K. N. Marsh (Eds.), Densities of Alcohols IV/8G, Springer-Verlag, Berlin/Heidelberg, Deutschland, 2000.
[6] M. Frenkel, X. Hong, R.C. Wilhoit, K.R. Hall, In Thermodynamic properties of organic compounds and their mixtures, in: K.R. Landolt-Börnstein Series - Hall, K. N. Marsh (Eds.), Densities of Esters and Ethers IV/8H, Springer-Verlag, Berlin/Heidelberg, Deutschland, 2001.
[7] M. Frenkel, X. Hong, Q. Dong, X. Yan, R.D. Chirico, Thermodynamic properties of organic compounds and their mixtures, in: K.R. Landolt-Börnstein Series - Hall, K. N. Marsh (Eds.), Densities of Phenols, Aldehydes, Ketones, Carboxylic Acids, Amines, Nitriles, and Nitrohydrocarbons IV/8I, Springer-Verlag, Berlin/Heidelberg, Deutschland, 2002.
[8] A.R. Katritzky, I. Tulp, D.C. Fara, A. Lauria, U. Maran, W.E. Acree Jr., A general treatment of solubility. 3. Principal component analysis (PCA) of the solubilities of diverse solutes in diverse solvents, J. Chem. Inf. Model. 45 (2005) 913.
[9] M. Frenkel, X. Hong, Q. Dong, X. Yan, R.D. Chirico, In Thermodynamic properties of organic compounds and their mixtures, in: M. Landolt-Börnstein Series - Frenkel, K.N. Marsh (Eds.), Densities of Halohydrocarbons IV/8J, Springer-Verlag, Berlin/Heidelberg, Deutschland, 2003.
[10] M. Frenkel, R.D. Chirico, V. Diky, Q. Dong, K.N. Marsh, J.H. Dymond, W. A. Wakeham, S.E. Stein, E. Königsberger, A.R.H. Goodwin, XML-based IUPAC standard for experimental, predicted, and critically evaluated thermodynamic property data storage and capture (ThermoML), Pure Appl. Chem. 78 (2006) 541.
[11] C. Wohlfahrt, In Static Dielectric Constants of Pure Liquids and Binary Liquid Mixtures, IV/17, Springer, Berlin, Germany, 2008.
[12] I. Tulp, D.A. Dobchev, A.R. Katritzky, W. Acree Jr., U. Maran, A general treatment of solubility 4. Description and analysis of a PCA model for Ostwald solubility coefficients, J. Chem. Inf. Model. 50 (2010) 1275.
[13] M. Zábranský, Z. Kolská, V. Růžička Jr., E.S. Domalski, Heat capacity of liquids: critical review and recommended values. Supplement II, J. Phys. Chem. Ref. Data 39 (2010), 013103/1.
[14] C. Caleman, P.J. van Maaren, M. Hong, J.S. Hub, L.T. Costa, D. van der Spoel, Force field benchmark of organic liquids: density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant, J. Chem. Theor. Comput. 8 (2012) 61.
[15] 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, 2009. http://comp.chem.umn.edu/mnsol.
[16] C.L. Yaws, In Thermophysical Properties of Chemicals and Hydrocarbons, second ed., Gulf Professional Publishing (Elsevier), Oxford, UK, 2014.
[17] C. Wohlfahrt, In Static Dielectric Constants of Pure Liquids and Binary Liquid Mixtures, IV/17S, Springer, Berlin, Germany, 2015.
[18] O. Suárez-Iglesias, I. Medina, M.A. Sanz, C. Pizarro, J.L. Bueno, Self-diffusion in molecular fluids and noble gases: available data, J. Chem. Eng. Data 60 (2015) 2757.
[19] W. Acree Jr., J.S. Chickos, Phase transition enthalpy measurements of organic and organometallic compounds. Sublimation, vaporization and fusion enthalpies from 1880 to 2015. Part 1. C_1-C_{10}, J. Phys. Chem. Ref. Data 45 (2016), 033101/1.
[20] G.D.R. 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 62 (2017) 1559.
[21] Springer Materials Database, 2018. https://materials.springer.com.
[22] NIST 103a Database, 2018. https://www.nist.gov/mml/acmd/trc/thermodata-engine/srd-nist-tde-103a.
[23] W.E. Acree Jr., Large Collection of Solute-Solvent Mutual Infinite-Dilution Solvation Free Energies, *Personal* **2018**, Communication, July 2018.
[24] J.R. Rumble, In CRC Handbook of Chemistry and Physics, ninety eighth ed., CRC Press/Taylor and Francis, Boca Raton, USA, 2018.
[25] Dechema Detherm Database, 2018. https://dechema.de/en/detherm.html.
[26] P.H. Hünenberger, CombiFF Data Collection in the ETHZ Research Collection (Tar-file CombiFF_validation_HAL_and_O+N_compounds Version 1.0 Corresponds to the Published Article), 2022, https://doi.org/10.3929/ethz-b-000445271.
[27] N. Schmid, A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, W. F. van Gunsteren, Definition and testing of the GROMOS force-field versions 54A7 and 54B7, Eur. Biophys. J. 40 (2011) 843.
[28] B.A.C. Horta, P.T. Merz, P. Fuchs, J. Dolenc, S. Riniker, P.H. Hünenberger, A GROMOS-compatible force field for small organic molecules in the condensed phase: the 2016H66 parameter set, J. Chem. Theor. Comput. 12 (2016) 3825.
[29] Y.M.H. Gonçalves, C. Senac, P.F.J. Fuchs, P.H. Hünenberger, B.A.C. Horta, Influence of the treatment of non-bonded interactions on the thermodynamic and transport properties of pure liquids calculated using the 2016H66 force field, J. Chem. Theor. Comput. 15 (2019) 1806.
[30] A.T. Hagler, E. Huler, S. Lifson, Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals, J. Am. Chem. Soc. 96 (1974) 5319.

[31] S. Lifson, A.T. Hagler, P. Dauber, Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 1. Carboxylic acids, amides, and the C=O ···H hydrogen bonds, J. Am. Chem. Soc. 101 (1979) 5111.
[32] J.E. Lennard-Jones, The equation of state of gases and critical phenomena, Physica 4 (1937) 941.
[33] M. Christen, P.H. Hünenberger, D. Bakowies, R. Baron, R. Bürgi, D.P. Geerke, T. N. Heinz, M.A. Kastenholz, V. Kräutler, C. Oostenbrink, C. Peter, D. Trzesniak, W. F. van Gunsteren, The GROMOS software for biomolecular simulation: GROMOS05, J. Comput. Chem. 26 (2005) 1719.
[34] N. Schmid, C.D. Christ, M. Christen, A.P. Eichenberger, W.F. van Gunsteren, Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation, Comput. Phys. Commun. 183 (2012) 890.
[35] W.G. Hoover, Canonical dynamics: equilibrium phase-space distributions, Phys. Rev. A 31 (1985) 1695.
[36] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. di Nola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684.
[37] W.F. van Gunsteren, H.J.C. Berendsen, J.A.C. Rullmann, Stochastic dynamics for molecules with constraints. Brownian dynamics of *n*-alkanes, Mol. Phys. 44 (1981) 69.
[38] E. Guàrdia, J.A. Padró, Generalized Langevin dynamics simulation of interacting particles, J. Chem. Phys. 83 (1985) 1917.
[[39]] W.F. van Gunsteren, H.J.C. Berendsen, A leap-frog algorithm for stochastic dynamics, Mol. Simulat. 1 (1988) 173.
[40] S. Yun-yu, W. Lu, W.F. van Gunsteren, On the approximation of solvent effects on the conformation and dynamics of cyclosporin A by stochastic dynamics simulation techniques, Mol. Simulat. 1 (1988) 369.
[41] W.F. van Gunsteren, Molecular dynamics and stochastic dynamics simulation: a primer, in: W.F. van Gunsteren, P.K. Weiner, A.J. Wilkinson (Eds.), Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications, vol. 2, ESCOM Science Publishers, B.V., Leiden, The Netherlands, 1993, pp. 3–36.
[42] R.W. Hockney, The potential calculation and some applications, Methods Comput. Phys. 9 (1970) 135.
[43] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes, J. Comput. Phys. 23 (1977) 327.
[44] H.J.C. Berendsen, W.F. van Gunsteren, H.R.J. Zwinderman, R.G. Geurtsen, Simulations of proteins in water, Ann. N. Y. Acad. Sci. 482 (1986) 269.
[45] J.A. Barker, R.O. Watts, Monte Carlo studies of the dielectric properties of water-like models, Mol. Phys. 26 (1973) 789.
[46] I.G. Tironi, R. Sperb, P.E. Smith, W.F. van Gunsteren, A generalized reaction field method for molecular dynamics simulations, J. Chem. Phys. 102 (1995) 5451.
[47] M. Neumann, Dipole moment fluctuation formulas in computer simulations of polar systems, Mol. Phys. 50 (1983) 841.
[48] A. Einstein, Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, Ann. Phys. 322 (1905) 549.
[49] T.C. Beutler, A.E. Mark, R. van Schaik, P.R. Gerber, W.F. van Gunsteren, Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations, Chem. Phys. Lett. 222 (1994) 529.
[50] J.G. Kirkwood, Statistical mechanics of fluid mixtures, J. Chem. Phys. 3 (1935) 300.
[51] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D. C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179.
[52] J.M. 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 (2004) 1157.
[53] W.L. Jorgensen, J. Tirado-Rives, Potential energy functions for atomic-level simulations of water and organic and biomolecular systems, Proc. Natl. Acad. Sci. USA 102 (2005) 6665.
[54] Y. Qiu, P.S. Nerenberg, T. Head-Gordon, L.-P. Wang, Systematic optimization of water models using liquid/vapor surface tension data, J. Phys. Chem. B 123 (2019) 7061.
[55] J.M. Míguez, M.M. Piñeiro, F.J. Blas, Influence of the long-range corrections on the interfacial properties of molecular models using Monte Carlo simulation, J. Chem. Phys. 138 (2013), 034707/1.
[56] R.A. Zubillaga, A. Labastida, B. Cruz, J.C. Martínez, E. Sánchez, J. Alejandre, Surface tension of organic liquids using the OPLS/AA force field, J. Chem. Theor. Comput. 9 (2013) 1611.
[57] N.M. Fischer, P.J. van Maaren, J.C. Ditz, A. Yildirim, D. van der Spoel, Properties of organic liquids when simulated with long-range Lennard-Jones interactions, J. Chem. Theor. Comput. 11 (2015) 2938.
[58] A.T. Hagler, S. Lifson, Energy functions for peptides and proteins. II. Amide hydrogen bond and calculation of amide crystal properties, J. Am. Chem. Soc. 96 (1974) 5327.
[59] A.T. Hagler, S. Lifson, A procedure for obtaining energy parameters from crystal packing, Acta Crystallogr. B 30 (1974) 1336.
[60] A.T. Hagler, S. Lifson, P. Dauber, Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 2. A benchmark for the objective comparison of alternative force fields, J. Am. Chem. Soc. 101 (1979) 5122.
[61] A.T. Hagler, P. Dauber, S. Lifson, Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 3. The C=O ···H hydrogen bonds and the analysis of the energetics and packing of carboxylic acids, J. Am. Chem. Soc. 101 (1979) 5131.
[62] P.H. Berens, D.H.J. Mackay, G.M. White, K.R. Wilson, Thermodynamics and quantum corrections from molecular dynamics for liquid water, J. Chem. Phys. 79 (1983) 2375.
[63] J.P.M. Postma, In *MD of H2O*, a Molecular Dynamics Study of Water, Thesis, University of Groningen, 1985.
[64] Q. Waheed, O. Edholm, Quantum corrections to classical molecular dynamics simulations of water and ice, J. Chem. Theor. Comput. 7 (2011) 2903.
[65] M. Kang, P.E. Smith, A Kirkwood-Buff derived force field for amides, J. Comput. Chem. 27 (2006) 1477.
[66] T.W. Whitfield, G.J. Martyna, S. Allison, S.P. Bates, H. Vass, J. Crain, Structure and hydrogen bonding in neat *N*-methylacetamide: classical molecular dynamics and Raman spectroscopy studies of a liquid of peptide fragments, J. Phys. Chem. B 110 (2006) 3624.
[67] R. Zhang, D. Zheng, Y. Pan, S. Luo, W. Wu, H. Li, All-atom simulation and excess properties study on intermolecular interactions of amide-water system, J. Mol. Struct. 875 (2008) 96.
[68] E. Harder, V.M. Anisimov, T. Whitfield, A.D. MacKerell Jr., B. Roux, Understanding the dielectric properties of liquid amides from a polarizable force field, J. Phys. Chem. B 112 (2008) 3509.
[69] B.A.C. Horta, Z. Lin, W. Huang, S. Riniker, W.F. van Gunsteren, P.H. Hünenberger, Reoptimized interaction parameters for the peptide-backbone model compound N-methylacetamide in the GROMOS force field: influence on the folding properties of two beta-peptides in methanol, J. Comput. Chem. 33 (2012) 1907.
[70] B. Lin, P.E.M. Lopes, B. Roux, A.D. MacKerell Jr., Kirkwood-Buff analysis of aqueous N-methylacetamide and acetamide solutions modeled by the CHARMM additive and Drude polarizable force fields, J. Chem. Phys. 139 (2013), 084509/1.
[71] S. Trabelsi, M. Bahri, S. Nasr, X-ray scattering and density-functional theory calculations to study the presence of hydrogen-bonded clusters in liquid N-methylacetamide, J. Chem. Phys. 122 (2005), 024502/1.
[72] C. Gainaru, S. Bauer, E. Vynokur, H. Wittkamp, W. Hiller, R. Richert, R. Böhmer, Dynamics in supercooled secondary amide mixtures: dielectric and hydrogen bond specific spectroscopies, J. Phys. Chem. B 119 (2015), 15769.
[73] X.-N. Jiang, C.-S. Wang, Rapid prediction of hydrogen bond cooperativity in N-nethylacetamide chains, ChemPhysChem 10 (2009) 3330.
[74] D.L. Mobley, C.I. Bayly, M.D. Cooper, M.R. Shirts, K.A. Dill, Small molecule hydration free energies in explicit solvent: an extensive test of fixed-charge atomistic simulations, J. Chem. Theor. Comput. 5 (2009) 350.
[75] J. Zhang, B. Tuguldur, D. van der Spoel, Force field benchmark of organic liquids. 2. Gibbs energy of solvation, J. Chem. Inf. Model. 55 (2015) 1192.