Title
Improving Force Field Accuracy by Training against Condensed-Phase Mixture Properties

Permalink
https://escholarship.org/uc/item/01q1n6vx

Journal
Journal of Chemical Theory and Computation, 18(6)

ISSN
1549-9618

Authors
Boothroyd, Simon
Madin, Owen C
Mobley, David L
et al.

Publication Date
2022-06-14

DOI
10.1021/acs.jctc.1c01268

Peer reviewed

Published in final edited form as:
J Chem Theory Comput. 2022 June 14; 18(6): 3577–3592. doi:10.1021/acs.jctc.1c01268.

Improving force field accuracy by training against condensed phase mixture properties

Simon Boothroyd[†,#], Owen C. Madin[‡,#], David L. Mobley[¶,§], Lee-Ping Wang[‖], John D. Chodera[⊥], Michael R. Shirts[‡]

[†]Boothroyd Scientific Consulting Ltd., 71-75 Shelton Street, London, Greater London, United Kingdom, WC2H 9JQ

[‡]Department of Chemical & Biological Engineering, University of Colorado Boulder, Boulder, CO 80309

[¶]Department of Chemistry, University of California, Irvine, CA 92617, United States

[§]Department of Pharmaceutical Sciences, University of California, Irvine, California, USA 92617

[‖]Department of Chemistry, University of California, Davis, CA, USA 95616

[⊥]Computational & Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065

Abstract

Developing a sufficiently accurate classical force field representation of molecules is key to realizing the full potential of molecular simulation as a route to gaining fundamental insight into a broad spectrum of chemical and biological phenomena. This is only possible, however, if the many complex interactions between molecules of different species in the system are accurately captured by the model.

Historically, the intermolecular van der Waals (vdW) interactions have primarily been trained against densities and enthalpies of vaporization of pure (single-component) systems, with

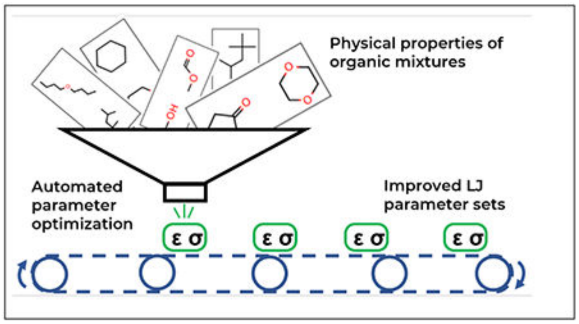
michael.shirts@colorado.edu .
[#]These authors contributed equally to this work

Author contributions
Conceptualization: S.B., O.M., M.R.S, D.L.M. and J.D.C.
Methodology: S.B., O.M. and L.P.W
Software: S.B. and L.P.W.
Investigation: S.B. and O.M.
Validation: S.B.
Formal Analysis: S.B. and O.M.
Data Curation: S.B. and O.M.
Writing - Original Draft: O.M. and S.B.
Writing - Review & Editing: O.M., S.B., M.R.S., J.D.C., and D.L.M.
Visualization: O.M. and S.B.
Supervision: M.R.S.
Project Administration: M.R.S, D.L.M, J.D.C. and L.P.W
Funding Acquisition: M.R.S, D.L.M., J.D.C. and L.P.W

Supporting Information Available
Details of test set molecule selection, parameter coverage of training set, values of initial force field parameters, optimization objective functions, values of optimized force field parameters and training set performance, test set property performance, error in one experimental data point.

occasional usage of hydration free energies. In this study, we demonstrate how including physical property data of binary mixtures can better inform these parameters, encoding more information about the underlying physics of the system in complex chemical mixtures. To demonstrate this, we re-train a select number of the Lennard-Jones parameters describing the vdW interactions of the OpenFF 1.0.0 (Parsley) fixed charge force field against training sets composed of densities and enthalpies of mixing for binary liquid mixtures as well as densities and enthalpies of vaporization of pure liquid systems, and assess the performance of each of these combinations. We show that retraining against the mixture data improves the force field's ability to reproduce mixture properties, including solvation free energies, correcting some systematic errors that exist when training vdW interactions against properties of pure systems only.

Graphical Abstract


Physical properties of organic mixtures
Automated parameter optimization
Improved LJ parameter sets
ε σ
ε σ
ε σ
ε σ

Introduction

Atomistic molecular simulations are a popular and effective method for examining biomolecular systems *in silico*, revealing molecular insights in protein folding, protein-ligand binding, membrane transport, and many other phenomena. For many of these use cases, quantitative accuracy is required for meaningful predictions. One critical example is binding free energy calculations for protein-ligand compounds. These calculations are an important step in the computational drug discovery process, but are only useful to medicinal chemists if predictions are sufficiently accurate and rapid.[1] Consequently, there has been much interest in producing improved parameter sets for the simple fixed charge functional forms common to most modern force fields. One key type of parameters are the parameters that specify the Lennard-Jones (LJ) interaction terms, which are used in standard organic and biomolecular force fields to capture the short-range attractive and repulsive non-bonded interactions that drive many important biomolecular processes.

The simplest method for obtaining LJ parameters is estimation from experimental correlations,[2] as in the original CHARMM[3] and GROMOS[4] force fields. This method has a low computational overhead but very limited accuracy. Training LJ parameters against experimental properties is more computationally expensive, but became the predominant method in small molecule force fields, facilitated by the increase in computational power required to simulate those properties. This method has been used by many force fields, including OPLS,[5] CGenFF,[6] GAFF,[7] and GROMOS.[8] The dominant parameterization paradigm is to train the LJ parameters against liquid density (ρ_l) and heat of vaporization

(ΔH_{vap}) measurements, as in the original OPLS parameterization by Jorgensen et al.[9] These two physical property targets are used because they are simple to calculate from simulation,[10] are dependent on the molecular volume and attractive forces, and together constrain the LJ ϵ and σ parameters better than they do individually. Densities and enthalpies are related to the derivatives of free energy with respect to volume and temperature, respectively; accurately reproducing the free energy is the target for most force fields. We note that while this is the dominant choice, alternatives exist; notably, the GROMOS 53A6[8] and 2016H66[11] force fields both use solvation free energies in addition to ρ_l and ΔH_{vap}. Additionally, *ab initio* calculations can be used to inform parameterization, for example, using rare-gas interaction energies and geometries to produce initial parameter estimates subsequently refined with physical property data.[12,13] More recently, methods to produce LJ parameters entirely from *ab initio* data, using atom-in-molecule electron density partitioning,[14,15] or the exchange hole dipole model[16] have been proposed. Still, parameterization against small molecule ρ_l and ΔH_{vap} is the dominant paradigm.[17,18]

Training against ΔH_{vap} in particular has some problematic aspects. Using fixed charge force fields, predictions of ΔH_{vap} require performing simulations in both liquid phase and gas phase, which means that the same parameters must capture two different polarization states[19,20] to reproduce experimental measurements of ΔH_{vap}. There has been significant discussion on how to account for this polarization cost, which also arises in the calculation of hydration and solvation free energies.[20–22] Methods suggested include calculating an explicit polarization cost[20] or using semi-polarized charges,[14,23] but the issue has not been definitively resolved. Additionally, some compounds, such as acids, can form clusters in the gas phase,[24,25] which are not generally represented in gas phase simulations used to predict ΔH_{vap}.

Another major issue is the availability of modern experimental ΔH_{vap} data. The NIST ThermoML Archive[26] is the one of the largest open databases for physical property measurements, and contains roughly 500 total ΔH_{vap} data points, where a “data point” in this context is defined as an experimental measurement for a specific compound at a given temperature T, pressure p, and mole fraction x. In contrast, the ThermoML Archive contains over 60,000 measurements of pure densities. The ThermoML Archive is certainly not the only location of ΔH_{vap} data (it lacks data prior to the year 2003, and many measurements of ΔH_{vap} date to the mid-20th century), but it is challenging to obtain uncertainty estimates,[27] rigorous provenance,[28] or fully computer-readable forms for these older measurements. This makes it difficult to systematically vet the experimental procedures and outputs when curating large scale datasets for parameter optimizations.

The limitations of parameterizing intermolecular interactions based off of pure properties alone have been noted previously; given two molecules A and B, accurate prediction of A-A interactions and B-B interactions does not imply accurate prediction of A-B interactions. Simulated ΔH_{vap} measurements for A and B, calculated as in equation 1, can measure the cohesive energies of A-A and B-B systems, but are unlikely to capture A-B interactions unless A and B are very similar molecules. In equation 1, ΔV refers to the difference in molar volume between the liquid and gas phases.

$$\Delta H_{vap} = \langle U_{gas} \rangle - \langle U_{liquid} \rangle + P\Delta V \tag{1}$$

Another illustrative example is the work of Kamath *et al.*[29] on azeotropes of acetone/methanol and chloroform/methanol systems, where force fields that accurately reproduced the pure components of these systems were unable to predict the azeotropic phase behavior until reparameterized against simulation of those mixture systems. In addition, Statistical Associating Fluid Theory[30] (SAFT) can be used to predict the behavior of mixtures; in particular, the SAFT-γ group contribution method has been used by Mueller and coworkers[31–34] to produce coarse-grained force fields for molecular fluid simulations and accurately predict the behavior of mixtures.[33] Another important example is the Kirkwood-Buff Force Field[35–37] (KBFF) of Smith and coworkers, which aims to achieve better treatment of solute-solvent interactions by capturing the concentration-dependent activities via Kirkwood-Buff theory.[36] Using Kirkwood-Buff integrals, the macroscopic activity can be related to microscopic solution structures obtained from simulation. Their efforts have focused on adjusting charge parameters, along with some LJ parameters, to match Kirkwood-Buff integral values and better capture solute-solvent interactions. This method has been used to parameterize a wide range of systems, from simple systems[35] to a complete protein and peptide force field.[38]

For a fixed charge small molecule force field geared towards biomolecular systems in heterogeneous condensed phase, our approach to capturing mixture interactions must be general, transferable, and focused on the LJ parameters, as charges for small molecules ligands are often generated from semi-empirical methods like AM1-BCC rather than being determined a priori. To ensure transferability, we need high-quality sources of diverse data to train against. Therefore, properties of mixtures such as the densities ($\rho_l(x)$) and enthalpies of mixing ($\Delta H_{mix}(x)$) of binary mixtures are an attractive alternative to the properties of pure systems for several reasons:

1. Properties of mixtures, especially in the cases of mixtures that deviate strongly from ideality, are sensitive to interactions between functional groups that are not generally present in the pure substances used to train LJ parameters.[39,40] Calculated as in equation 2, simulated enthalpies of mixing directly capture the A-B interactions that enthalpies of vaporization miss. This is especially important for capturing solute-solvent interactions.

$$\Delta H_{mix}(x_1, x_2) = H_{mix} - x_1 H_1 - x_2 H_2 \tag{2}$$

2. The nature of mixture data allows users to more easily include a diverse spectrum of interactions in their training sets. For example, mixtures of drug-like molecules with pharmaceutically relevant solvents or amino acid analogues can in principle be readily included in training sets to allow the LJ parameters of solvents, ligands, and bio-polymers to be self-consistently trained.

3. Although computing some properties of mixtures may require multiple simulations, most such properties (including those studied here) do not require simulations in different phases, minimizing error caused by polarization

differences. There may be some difference in polarization of molecules between more polar and less polar liquids, but this difference is significantly less than the difference between two phases, especially since liquid mixtures are, by definition, miscible and the components must therefore have dielectric constants that are not completely dissimilar.

4. Including mixture data adds the ability to vary training set data by composition; data points can be selected at (T, P, x) rather than just (T, P), probing the balance between pure and mixture interactions.

5. Many data points for mixture properties are available in modern sources such as the ThermoML Archive. In particular, binary $\Delta H_{mix}(x)$ measurements are much more abundant in the ThermoML archive compared to pure ΔH_{vap}. For the moieties and conditions of interest in our study, there are 382 binary mixtures with $\Delta H_{mix}(x)$ measurements (generally available at multiple concentrations), compared to 24 single-component ΔH_{vap} measurements that fit the same criteria. For density measurements, both mixture and pure component data points are relatively abundant, with 4000 data points for pure substances and 900 binary mixtures matching our criteria. We estimate that there is sufficient binary mixture training data to parameterize small molecules containing carbon, hydrogen, oxygen, nitrogen, chlorine, and bromine.

In this study, we aim to rigorously assess whether it is more beneficial to train the intermolecular LJ parameters of a force field on solely pure substance data, binary mixture data, or a combination of both, with an emphasis here on density-related properties (ρ_l, $\rho_l(x)$) and enthalpic properties (ΔH_{vap}, $\Delta H_{mix}(x)$). A combination of density and enthalpic data should be generally sufficient to constrain the LJ σ and ε parameters, with densities providing the most information about σ and enthalpic properties providing information on ε via the cohesive forces between molecules, though there is of course some partial cross-correlation between parameters.[41]

Starting with the OpenFF 1.0.0 (Parsley) force field[42], we use this data to train 12 Lennard-Jones parameters (σ and ε for 6 LJ types) against data for alcohols, esters, ethers, ketones, acids, and alkanes, with property measurements chosen from four training sets containing different combinations of physical properties. To test the performance of the refitted force fields, we benchmark the results of this optimization against a larger test set of physical property measurements for the same moieties, consisting of $\rho_l(x)$, $\Delta H_{mix}(x)$, ρ_l, and ΔH_{vap} measurements.

Methods

Optimization strategy

The studies proposed are constructed with the following workflow, as shown in Figure 1.

1. Sourcing a training set of molecules and selecting particular measurements for each molecule (or pair of molecules) of interest.

2. Optimizing only the selected LJ parameters against the training set using ForceBalance[43] in combination with the OpenFF Evaluator framework[44], starting from the OpenFF 1.0.0 (Parsley)[45] force field parameters.

3. Assessing the performance of the trained force field against a test set of measurements using the OpenFF Evaluator framework.

The goal of the study was to assess whether training the LJ parameters against properties of mixtures, as well as combinations of pure/mixture properties, is more beneficial than training against properties of pure systems. Other force field parameters, namely the valence and electrostatic parameters, were not optimized.

Organic Mixture Studies—We selected four combinations of physical property data types (densities of pure compounds and binary mixtures, heats of vaporization of pure compounds, and enthalpies of mixing of binary mixtures) to optimize against (shown in Table 1).

1. (ρ_l, ΔH_{vap}) **("pure only")**: Includes only density ρ_l, and enthalpy of vaporization ΔH_{vap}, data points. This is the type of training set which has most commonly been used[5–7] for training the non-bonded interaction force field parameters, and is therefore included as a historical baseline.

2. ($\Delta H_{mix}(x)$, $\rho_l(x)$) **("mixture only")**: Includes only density $\rho_l\,(x)$ and enthalpy of mixing $\Delta H_{mix}\,(x)$ data points measured for binary mixtures. This data set allows us to explore whether mixture data alone is sufficient to constrain the non-bonded force field parameters during training, and if force field trained without pure compound data points will be able to accurately reproduce pure compound data.

3. ($\Delta H_{mix}(x)$, $\rho_l(x)$, ρ_l) **("mixtures + pure density")**: A combination of $\rho_l\,(x)$, $\Delta H_{mix}\,(x)$, and ρ_l data points. This extension of the "mixture only" training set is included to explore whether including the density of pure systems helps to constrain the optimization, or whether $\rho_l(x)$ alone is sufficient.

4. ($\Delta H_{mix}(x)$, $\rho_l(x)$, ρ_l, ΔH_{vap}) **("pure and mixture")**: A combination of the "pure only" and the "mixture only" training sets. This data set tests whether including pure ΔH_{vap} alongside $\Delta H_{mix}(x)$ improves the parameterization of the cohesive energies between molecules, or whether $\Delta H_{mix}(x)$ alone is sufficient.

The measurements in the training set are for molecules composed of carbon, hydrogen and oxygen only (including alcohols, esters, ethers, ketones, acids and alkanes). These compounds cover a wide range of fluid phase polarizabilities, with relative permittivities ranging from 1.9 (hexane[46]) to 35.7 (methanol[47]).

Data set selection

All training sets considered here are composed of only alcohols, esters, ethers, ketones, acids and alkanes that have ample density and enthalpic data available, and contain only data points measured at near-ambient conditions (288.15–323.15K, 0.95–1.05 atm). This set of moieties, containing only carbons, hydrogens and oxygens, was chosen to limit the scope of the study and focus specifically on the choice of training data for a set of molecules. The molecules included exercise a total of 9 LJ types, of which 6 are optimized, shown in Table 2. A table showing which LJ types are exercised by each molecule in the training set is available in Supporting Information Section S2.1. The three parameters included that are not optimized are all hydrogen parameters; an explanation of why they are not optimized is given in the “Parameters optimized” section.

We enforce the criteria that all measurements in the data set contain only the molecules in Figure 2. This criteria controls for the identity of molecules used in the optimization; whether the measurements used in fitting are from pure substance or binary mixtures, they are restricted to the same set of molecules. We note that some values for $\rho_l(x)$ are obtained through the conversion of $V_{excess}(x)$ and ρ_l where $\rho_l(x)$ is not directly available.

Pure substance training set—The “pure only” training set is composed of one ρ_l and one ΔH_{vap} measurement for each of the selected molecules (Figure 2). These molecules were manually chosen to include a selection of esters, ethers, ketones, alcohols and alkanes which included long/short chain, branched/unbranched, and cyclic/acyclic characteristics, where data was available. The ρ_l measurements were sourced from the NIST ThermoML[26] archive. The ΔH_{vap} measurements were sourced directly from the literature, as very limited data for the moieties of interest is available in the ThermoML Archive. Many data points were curated from the Majer et al. review[27], where care was taken to select data points which were deemed as reliable by the authors, and for which at least three independent measurements had been made and were in reasonable agreement. In total, 28 molecules were chosen for a total of 56 data points (28 ρ_l data points[48–72] and 28 ΔH_{vap} data points[25,73–83]). For ΔH_{vap} of acids, measurements were sourced which correspond to an infinitely dilute gas (as computed in[25]), which corresponds to the gas we simulate. This is done because carboxylic acids tend to associate in the gas phase.

Mixture training set—The binary mixtures selected for the mixture training set (Figure 3) are composed of the molecules included in the pure training set, and were manually chosen to include a diverse set of interactions. These property measurements were sourced directly from the NIST ThermoML[26] archive using the OpenFF Evaluator’s built-in data selection tools. Where available, three $\rho_l(x)$ and three $\Delta H_{mix}(x)$ data points were included for each binary mixture, one each at 25%, 50%, and 75% composition, or as close to these values as possible given data availability. These compositions were chosen so as to ensure that the set included both components in excess to the other as well as in close to equal amounts. Compositions between 25–75% should capture most of the relevant information, as deviations from ideality for many mixtures are maximized near an equal mixture. Mixtures with compositions close to pure (e.g. > 0.9) were excluded, as when the concentration of one component becomes small, our simulation boxes (1000 total molecules) would have a very low number of molecules of that component. In total, measurements made for 33 binary mixtures were selected for a total of 195 data points. This is significantly more than the 56 total data points in the pure data set, but it is drawn from a number of mixtures similar to the number of compounds in the pure training set. We note that after training was complete, we discovered that one $\Delta H_{mix}(x)$ data point in the mixture training set was transcribed into ThermoML incorrectly (described in Supporting Information, Section 4).

Test set—The test set was chosen to include measurements of $\rho_l\,(x)$, $\Delta H_{mix}\,(x)$, ρ_l, and ΔH_{vap} as in the training set. Additionally, a set of non-aqueous solvation free energy (ΔG_{solv}) measurements for the same moieties included in the training set was sourced from the MNSolv database.[84] Unlike the training set, we do not require that all pure substance and binary mixture measurements in the test set must be sourced from the same set of molecules. Instead, given the limited amount of diverse $\Delta H_{mix}\,(x)$ and ΔH_{vap} data for the selected moieties, focus was given to selecting as diverse a test set as possible which maximally exercised the re-trained parameters. Data points from pure substances included in the training set were excluded from the test set, as well as mixture data points from mixtures included in the test set. The test set did include binary mixtures for which one of the two components was present in the training set; for example, a mixture of ethanol and pentanol would be permissible in the test set even if data points for ethanol/propanol and butanol/pentanol were both included in the training set. This expands the test set to types of mixtures that were not included in the training set; for example, mixtures containing either an alcohol or ketone are in the training set, but alcohol/ketone mixtures are only included in the test set. The set was also selected to contain substances as distinct as possible from the training set, and from other molecules in the test set. Mixtures including carboxylic acids were not included in the test set due to low data availability.

In order to select a maximally diverse test set from the pool of molecules available in the ThermoML Archive or MNSolv Database, a distance metric based on molecular fingerprints was defined to determine how distinct any two substances are. Then, binary mixtures were selected by a greedy optimization that maximized this distance metric. For a more detailed description of this process, see the Supporting Information Section S1.

The substances included for pure substance (ρ_l and ΔH_{vap}) measurements were then chosen to match the components of the test set mixture properties where available; these were supplemented with measurements for similar molecules that exercise the same LJ parameters. This resulted in a test set consisting of 236 $\Delta H_{mix}\,(x)$ (from 43 unique molecules), 385 $\rho_l\,(x)$ (from 60 unique molecules), and 85 ΔG_{solv} measurements (from 31 unique molecules), which was supplemented with a hand-selected test set of 29 ΔH_{vap} and 29 ρ_l pure component measurements.

Physical property simulations

All estimates of the physical property values were performed using the OpenFF Evaluator[44] package version 0.1.0[85] using the default estimation workflow schemas, which are outlined in detail in the OpenFF Evaluator documentation.[86] Where possible, simulations are reused to calculate physical properties. For example, simulations of a pure liquid phase can be reused in calculations of ρ_l, ΔH_{vap} and ΔH_{mix}.

Pure Liquid Simulations—Pure liquid properties were calculated by simulation in the NPT ensemble, at the temperature and pressure from the corresponding physical property reference. These were performed with the default OpenFF Evaluator simulation workflow, in which a box of 1000 molecules of the target substance were placed in a simulation box using PackMol,[87] with parameters then assigned using the OpenFF Toolkit version 0.6.0.[88] An energy minimization and 0.2 ns equilibration run were then performed using OpenMM. Subsequently, the molecules were simulated for 2 ns. For all simulations, a Langevin integrator with BAOAB[89] splitting and a 2 fs timestep, and the default OpenMM Monte Carlo barostat, were employed to ensure simulation in the correct NPT ensemble. Uncorrelated and well-equilibrated snapshots were used to compute the ensemble averages of any observables, according to the procedure outlined by Chodera.[90] All uncertainties in the average observables were computed by bootstrapping with replacement, and propagated through any further calculations, assuming a Gaussian error model. Densities are estimated using ensemble averages from these simulations.

Locations of scripts to run the simulations and reproduce the results in this study are available in the Code and Data Availability section.

Enthalpy of Vaporization Calculations—Enthalpies of vaporization require a pure liquid simulation, as described in Section, as well as a gas phase simulation. This gas phase simulation is performed for a single molecule in the NVT ensemble, with periodic boundaries disabled, using the same Langevin integrator as used with the liquid simulations. These simulations are run for 30 ns instead of the liquid phase 2 ns to converge statistics with only a single molecule. Enthalpies are calculated using Equation 1.

Mixture Properties—Mixture densities were simulated with a similar workflow to the pure liquid simulations, but with the molecules in the initial box split proportionally between the two chemical species according to the experimental mole fraction. Densities of binary mixtures are straightforward to calculate as they do not require more than one simulation; the process is the same as for densities of single component liquids. Binary enthalpies of mixing are calculated according to equation 2, where the enthalpies of the individual simulated components (H_1, H_2) are multiplied by their mole fractions in the mixture, and then subtracted from the enthalpy of the simulated mixture $H_{mix}(x_1, x_2)$.

Enthalpies used in this calculated were simulated with a set of 3 simulations: one for each pure component, and one for the mixture. Each of these simulations followed the standard workflow for a pure or mixture property.

Solvation Free Energies—Solvation free energies were calculated using the default OpenFF evaluator workflows, along with the YANK software package version 0.25.2[91,92] for performing alchemical free energy calculations. The alchemical cycle used in the calculation is the same as described in Shivakumar et al.,[93] and involves 1) the removal of a solute molecule from a box of solvent and 2) the annihilation of the solute molecule in gas phase. Calculation of step 1) involves an alchemical pathway along which non-bonded interactions are gradually turned off. Values of the λ variable that describes this pathway are automatically determined by YANK. Liquid phase simulations are set up in a similar fashion to those used in our other simulations, but with 2000 molecules rather than 1000 to reduce statistical uncertainty; gas phase simulations use the same settings as in the calculation of enthalpies of vaporization.

Optimization

For stochastic gradient descent optimizations, we need to estimate gradients of the observables of interest as a function of force field parameters. In this paper, gradients are calculated using a reweighted finite difference scheme, where the derivative $\frac{dO}{dx}$ of an observable O with respect to a parameter x is calculated using the central difference method with a relative step ($\Delta x/x$) size of $h = 10^{-4}$. Values of O at $x - h$ and $x + h$ are estimated by reweighting from the sampled ensemble using `pymbar`,[94] which is accurate for the properties of interest over the small step size h. All optimizations were performed using the ForceBalance software package using the built-in OpenFF Evaluator target.[43,44] Optimizations were run using the Levenberg-Marquardt[95] non-linear least squares algorithm with adaptive trust radius[96,97] to iteratively minimize the objective function until it was observed to fluctuate around a minimum value in each optimization. This algorithm has been used successfully with ForceBalance for force field optimization previously.[43,98] In all cases 12 iterations was sufficient to meet this criteria. Each iteration consists of 1) estimation of each physical property measurement in the training set using the current force field parameters, 2) comparison of those estimated values to the experimental values in ThermoML, 3) adjustment of the target parameters with the ForceBalance optimizer. A weighted least squares objective function, χ, was used to measure deviations of the reference and estimated physical property values. An L2 penalty function based on the norm of the parameter displacement vector (from the initial parameters) is used to regularize the optimization, with a prior over the ForceBalance mathematical parameters[43] of 0.1 for ε and 1.0 for σ.

$$\chi(\theta) = \sum_{n=1}^{N} \frac{1}{M_n} \sum_{m=1}^{M_n} \left(\frac{y_m^{ref} - y_m(\theta)}{d_n} \right)^2 \tag{3}$$

where N is the number of types of properties (e.g. density, enthalpy of vaporization, etc.), M_n is the number of data points of type n, y_m^{ref} is the experimental value of data point m and $y_m(\theta)$ is the estimated value of data point m using the current force field parameters. The denominator d_n is an inverse weight with the same units as property type n chosen so that that each property type contributed approximately equally to the objective function. For example, for the pure training set, ~ 50% of the objective function value is due to ρ_l data, and ~ 50% is due to ΔH_{vap}. This *a priori* approximation was made as it is unclear that any one type of property should be weighted more than another.

Parameters optimized—Both the training and test sets, each containing only molecules composed of carbon, hydrogen, and oxygen, exercise a total of 18 SMIRNOFF LJ parameters (9 different SMIRKS parameter types with one ε and σ per SMIRKS). These LJ parameters in OpenFF 1.0.0 have not been optimized since their inception in the first SMIRNOFF format force field,[99] and are taken chiefly from AMBER parm94,[100] with the exception of the hydroxyl hydrogen parameter discussed below. Of these parameters, 12 were optimized, with the remaining 6 held constant at their initial OpenFF 1.0.0 values. The parameters held constant (all for hydrogens) were not optimized because either the parameter correspond to a specific context that was not sufficiently constrained by the training data set or, in the case of [#1:1]-[#8] (hydroxyl hydrogen), the OpenFF 1.0.0 ε value is explicitly set to a very small nonzero value ($\varepsilon = 5.27 \times 10^{-5}$) and not reoptimized. This is a slight modification of the AMBER hydroxyl hydrogen parameter[100] (HO, $\varepsilon = 0$) to avoid unphysical effects caused by the AMBER parameterization.[101] Here each parameter is uniquely identified by a SMIRKS pattern which encodes the chemical environment to which the parameter will be applied.[99] These parameters, along with brief descriptions, are listed in Table 2.

Testing

Tests of force field performance were performed by taking the final force fields produced from each optimization and estimating each data point in the test set using OpenFF Evaluator. All property calculations were made using the same property prediction workflows as in the optimizations.

To assess the improvement of the refitted force fields relative to OpenFF 1.0.0, we calculate the mean shift in absolute error for each of the physical property types in the benchmark set. This metric describes the average improvement (or regression) in a refitted force field's ability to reproduce test set physical properties compared to a reference force field, and is described in Equation 4 for a generic observable O.

$$\overline{\Delta\left(\Delta O_{sim-exp}\right)_{\mathrm{ff0}\to\mathrm{ff1}}} = \frac{1}{N}\sum_{n=1}^{N}\left(\left|O_{sim,\mathrm{ff1}} - O_{exp}\right| - \left|O_{sim,\mathrm{ff0}} - O_{exp}\right|\right)_n \tag{4}$$

In this equation, $O_{sim,\mathrm{ff}x}$ is the simulation estimate of O with a given force field, and O_{exp} is the experimental value of O. The reference force field ff0 is always chosen as OpenFF 1.0.0 in this analysis. The average is taken over the test set of N physical properties of one specific type (e.g. over the 236 $\Delta H_{mix}(x)$ measurements in our test set). When bootstrapped 95% confidence intervals are calculated with this metric, bootstrapping is performed over paired measurements in two force fields, capturing the correlation between force fields that is lost when bootstrapped errors are calculated individually.

We also calculate kernel density estimates (KDE)[102,103] of the distribution of individual shifts $(|O_{sim,\mathrm{ff1}} - O_{exp}| - |O_{sim,\mathrm{ff0}} - O_{exp}|)_n$ to visualize the differences in improvement for the different force fields. KDE plots are generated using the seaborn[104] 0.11.2 data visualization package, with a Gaussian kernel and bandwith calculated with the method of Scott.[105]

Results & Discussion

Optimization

Parameter Changes—The objective function was observed to decrease by 50–70% for each of the four optimizations performed, indicating improvements against the training set in all cases (see Supporting Information Section S2.1). This improvement was achieved with relatively small changes in the target parameters, as most of the refitted parameters changed only slightly from their initial values, varying less than 5% in most cases (Figure

4). A notable exception is ε for `[#1:1]-[#6X4]` (hydrogen attached to tetravalent carbon), which changes up to 40% depending on the optimization. We also note that the σ for `[#8X2H1+0:1]` (hydroxyl oxygen) changes much more when trained against mixture data (−0.4 % for "pure only" vs. −1.7–2.8% for sets containing mixture data).

Training Set Property RMSE—We examine the performance of the trained force fields on the training set, as well as the changes in parameters after optimizations. This detailed look at the optimization sheds light on which parameter changes are driving the specific property improvements that result in an improved force field. Using the RMSE for each target property as a metric and grouping by property and chemical environment, it is clear that most of the different moieties in the training set are improving when trained against either pure or mixture data. This is evident when training against both the "pure only" data set in Figure 5 and the "mixture only" data set in Figure 6. Improvements in both pure and mixture training data for the other two (mixed) optimizations were also observed, which are shown in supporting information (Section S2.6.2,S2.7.2).

One notable exception is ketones, as pure ketone densities and "Ketone > Ether" binary densities were both degraded upon training. Given that this occurs for both pure and mixture training data, it is unlikely that it is a symptom of the training sets selected. We also note that ketone ΔH_{vap} RMSEs are improved, alongside both densities and ΔH_{vap} RMSEs for esters, which utilize the same `[#8:1]` generic carbon parameter. It is likely that these properties are improved at the expense of ketone densities. By examining the first derivatives of the density contribution to the objective function with respect to the force field parameters, again partitioned by moiety (Figure 7), we see that modifying the `[#1:1]-[#6X4]` (hydrogen attached to tetravalent carbon), `[#6X4]` (tetravalent carbon), and `[#8:1]` (generic oxygen) has an opposite effect on ketone objectives compared to the objective for other moieties. This suggests that the force field lacks the degrees of freedom required to accurately capture carbons and hydrogens in ketone environments alongside the other environments represented by the same SMIRKS patterns. It is possible that including a more specific hydrogen or carbon parameter for this environment might improve prediction of ketone densities. Another possibility is that the LJ parameters are compensating for deficiencies in the AM1-BCC electrostatic model, which was not optimized in this study. This result will be explored in further work as it is beyond the scope of the current study. However, analyses such as these point out how additional interaction types can be motivated by the large sets of data generated by this sort of study.

Test Set Performance

Overall Results—Benchmarking simulations of the test set physical property measurements were performed for OpenFF 1.0.0 and each of the refitted force fields.

Mean shift and and shift distributions (metrics described in the Methods section) for the test sets of each of the four physical properties used in training are shown in Figure 8.

We observe that for both ρ_l and $\rho_l(x)$, the refitted force fields all offer mild improvements over OpenFF 1.0.0, with no significant differences between them. This is consistent with our expectations as densities are generally well predicted in the initial force field. On the other hand, for $\Delta H_{mix}(x)$, all refitted force fields improve relative to OpenFF 1.0.0, but the improvements of the three force fields trained with mixture data (“mixture only”, “mixtures + pure density”, and “pure and mixture”) are significantly larger (0.2 kJ/mol vs 0.1 kJ/mol for “pure only”), indicating that training against mixture data significantly improves performance on our $\Delta H_{mix}(x)$ test set. This is also clearly visible in the KDE plot, where the distributions for the sets containing mixtures are shifted relative to the “pure only” set. A significant number of measurements are improved by >0.75 kJ/mol when trained against mixtures, whereas almost no measurements achieve this improvement when trained against the “pure only” set. Similarly, for our ΔH_{vap} test set, we observe that the two force fields trained against sets that include ΔH_{vap} data (“pure only” and “pure and mixture”) offer significant improvements over OpenFF 1.0.0, whereas the two sets that do not include ΔH_{vap} (“mixture only”, “mixtures and pure density”) do not improve relative to the initial force field. Again, this can be seen clearly in the KDE plot, where the peak of the distributions for force fields trained with ΔH_{vap} data are shifted left compared to the other force fields.

This data shows, perhaps unsurprisingly, that force fields trained against ΔH_{vap} and $\Delta H_{mix}(x)$ will do better at reproducing those respective properties. In this view, one could assume that training against the “pure and mixture” set, which contains both types of enthalpy data, is the best strategy. However, the utility of improved ΔH_{vap} predictions is questionable for a force field intended to be used for biomolecular systems where vaporization does not typically occur.

With this in mind, benchmarking on the non-aqueous ΔG_{solv} test set serves as a more neutral test of the different force fields’ abilities to capture the appropriate interaction strengths between molecules. A plot of the mean shifts for the ΔG_{solv} test set, as well as a KDE plot of the shift distribution, is shown in Figure 9.

The mean shift of ΔG_{solv} absolute errors relative to OpenFF 1.0.0 show that training against the “mixture only” set provides an improvement over the initial force field, whereas training against the “pure only” set degrades ΔG_{solv} predictions. This is also reflected in the KDE plot, where the peak of the shift distribution is shifted right for the two sets that contain ΔH_{vap} (“pure only”, “pure and mixture”) compared to the two that do not (“mixture only” and “mixtures + pure density”), suggesting that refitting to ΔH_{vap} hinder attempts to reproduce properties like ΔG_{solv}. It is important to note that the initial LJ parameters used in this force field were fitted to ΔH_{vap} simulation when originally determined,[106?] and that the RMSE of OpenFF 1.0.0 on the ΔG_{solv} test set is 3.3 kJ/mol, so reasonably accurate predictions can be obtained with LJ parameters trained against ΔH_{vap}. However, training against mixture data can offer additional improvements to performance.

These results indicate that mixture properties can replace physical properties of pure systems as a target for training LJ parameters, particularly in cases where more and more chemically diverse data is available for mixtures. Training against the “pure only” set does lead a significant improvement to ΔH_{mix} (x) against the baseline; however, training directly against the “mixture only” set yields a much larger improvement. It appears that training against properties of mixtures alone sufficiently constrains the optimization, and includes enthalpic information that the traditional pure dataset alone does not. We also note that augmenting a traditional pure data training set with mixture data (such as the “pure and mixture” set) can improve treatment of mixture properties without degrading performance on pure properties.

Results by chemical environment—Notably, training against the mixture properties appears to have corrected a systematic error in the enthalpy of mixing, which training against pure properties alone is not able to correct. This can be inferred from the KDE plot for $\Delta H_{mix}(x)$ in Figure 4, where the shift of the secondary peak indicates that performance is improving for a subset of molecules. More specific evidence is obtained from a simulation/experiment parity plot for $\Delta H_{mix}(x)$, where a systematic underprediction of alcohol/ester (green points) and alcohol/ketone (orange points) mixture enthalpies is corrected (Figure 10). The improvement in the treatment of alcohol/ketone mixtures was achieved without directly including these mixtures in the training set.

This is particularly significant as alcohol/ester and alcohol/ketone mixture enthalpies have strong deviations from ideal solution behavior. Namely, ketone and esters are both hydrogen bond acceptors only and thus do not form hydrogen bonds in the pure phase. However when mixed with a hydrogen bond donor (an alcohol) they do. This change is likely related to the reduction in σ for the `[#8X2H1+0:1]` (hydroxyl oxygen), noted in Figure 4. This reduction is much larger (1.7%–2.8% vs. 0.4% for “pure only”) for force fields refit against mixture data. This is where mixture properties, and especially their ability to more readily capture complementary interactions, appear to be advantageous over pure properties.

Conclusions

Using our automatic data set selection and force field optimization workflow, we re-parameterized select LJ parameters of the OpenFF 1.0.0 force field against training sets containing combinations of pure (ρ_l, ΔH_{vap}) and mixture ($\rho_l(x)$, $\Delta H_{mix}(x)$) properties for alkanes, alcohols, esters, ethers, ketones, and acids. These training sets were controlled such that the same molecules are used in both pure and mixture training sets, to isolate the effect of the different data types used. Through iterative optimization of parameter sets, new force fields were produced that all exceeded the performance of the initial force field on some parts of the test set. Furthermore, we observe that training LJ parameters against mixture data constrains the optimization in a comparable or superior manner to optimizing with the traditional pure properties commonly used in LJ parameterization.

Training against mixture properties, specifically $\Delta H_{mix}\ (x)$, is a compelling alternative for capturing enthalpic contributions to LJ interactions to ΔH_{vap}. Training against ΔH_{vap} is problematic due to limited data coverage and quality, as well as changes in molecular polarization between liquid and gas phase simulations. Mixture property datasets also offer expanded datasets by varying composition, and are more widely available in the ThermoML Archive. Moreover, we have shown here how mixture properties offer significant advantages over pure properties as an optimization target, especially in those cases of interactions which deviate strongly from ideality. These advantages lead to improved LJ parameters sets and better agreement with experiment. Given that we control for the identity of the molecules in the training set, this demonstrates that mixture properties contain information about intermolecular interactions that pure component property measurements do not.

While some parameter sets we demonstrate in this work improved both enthalpies of vaporization and enthalpies of mixing, in our view, improvements in the properties of mixtures are a better metric of force field improvement than pure or phase change properties for force fields intended for use in biomolecular simulations, since simulations typically take place in mixed aqueous or other liquid phases. This is supported by our finding that force fields trained against mixture data improve predictions of ΔG_{solv}, whereas force fields trained against only pure data (including ΔH_{vap}) degrade those predictions. The same interactions captured in solvation free energies should also be informative for properties of pharmaceutical/biomolecular interest, such as binding affinities. For this reason, optimization of LJ parameters against mixture property targets is planned to be the standard going forward for our OpenFF force fields. It is also important to note the scope of the study is limited to LJ parameters, and that other parameters, such as electrostatics, torsions, and 1–4 atomic scalings will impact the accuracy of these mixture properties. We anticipate that the automated property prediction in our parameterization workflow, along with the wider chemistry covered by the mixture properties in the ThermoML Archive, will lead to more accurate LJ parameters for general small molecule force fields.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We thank Christopher Bayly (ORCID: https://orcid.org/0000-0001-9145-64570000-0001-9145-6457) for his insightful discussion and feedback on the results of this study and their conclusions, along with all current and former members of the Open Force Field Initiative and the Open Force Field Scientific Advisory Board.

Funding

We thank the Open Force Field Consortium for funding, including our industry partners as listed at the Open Force Field website (http://openforcefield.org). We also appreciate the Molecular Sciences Software Institute (MolSSI) for their support of the Open Force Field Initiative. Simon Boothroyd also acknowledges both a Joint OpenFF-XtalPi Distinguished Postdoctoral Fellowship. JDC acknowledges support from NSF grant CHE-1738979, NIH grant P30CA008748, NIH grant R01GM132386, and the Sloan Kettering Institute. Research reported in this publication was in part supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM132386. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This material is also, in part, based upon work supported by the National Science Foundation under Grant No. CHE-1738979. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Conflict of Interest Statement

MRS is an Open Science Fellow for Roivant Sciences. DLM is a current member of the Scientific Advisory Board of OpenEye Scientific Software and an Open Science Fellow for Roivant Sciences. SB is a director of Boothroyd Scientific Consulting Ltd. JDC is a current member of the Scientific Advisory Board of OpenEye Scientific Software, Redesign Science, and Interline Therapeutics, and has equity interests in Redesign Science and Interline Therapeutics. The Chodera laboratory receives or has received funding from multiple sources, including the National Institutes of Health, the National Science Foundation, the Parker Institute for Cancer Immunotherapy, Relay Therapeutics, Entasis Therapeutics, Silicon Therapeutics, EMD Serono (Merck KGaA), AstraZeneca, Vir Biotechnology, Bayer, XtalPi, Interline Therapeutics, the Molecular Sciences Software Institute, the Starr Cancer Consortium, the Open Force Field Consortium, Cycle for Survival, a Louis V. Gerstner Young Investigator Award, and the Sloan Kettering Institute. A complete funding history for the Chodera lab can be found at http://choderalab.org/funding.

Data and Code Availability

Scripts to run the simulations and reproduce the results in this study are available at https://github.com/SimonBoothroyd/binary-mixture-publication.

The training and test data sets used in this publication are also available in this repository in .csv and .json formats.

To provide feedback on performance of the OpenFF force fields, we highly recommend using the issue tracker at http://github.com/openforcefield/openforcefields. For toolkit feedback, use http://github.com/openforcefield/openforcefield. Alternatively, inquiries may be e-mailed to support@openforcefield.org, though responses to e-mails sent to this address may be delayed and GitHub issues receive higher priority. For information on getting started with OpenFF, please see the documentation linked at http://github.com/openforcefield/openforcefield, and note the availability of several introductory examples.

References

(1). Bottaro S; Lindorff-Larsen K Biophysical Experiments and Biomolecular Simulations: A Perfect Match? Science 2018, 361, 355–360. [PubMed: 30049874]

(2). Slater JC; Kirkwood JG The Van Der Waals Forces in Gases. Physical Review 1931, 37, 682–697.

(3). Brooks BR; Bruccoleri RE; Olafson BD; States DJ; Swaminathan S; Karplus M CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. Journal of Computational Chemistry 1983, 4, 187–217.

(4). Van Gunsteren WF; Karplus M Effect of Constraints on the Dynamics of Macromolecules. Macromolecules 1982, 15, 1528–1544.

(5). Jorgensen WL; Maxwell DS; Tirado-Rives J Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. Journal of the American Chemical Society 1996, 118, 11225–11236.

(6). Vanommeslaeghe K; Hatcher E; Acharya C; Kundu S; Zhong S; Shim J; Darian E; Guvench O; Lopes P; Vorobyov I; Mackerell AD CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. Journal of Computational Chemistry 2010, 31, 671–690. [PubMed: 19575467]

(7). Wang J; Wolf RM; Caldwell JW; Kollman PA; Case DA Development and Testing of a General Amber Force Field. Journal of Computational Chemistry 2004, 25, 1157–1174. [PubMed: 15116359]

(8). Oostenbrink C; Villa A; Mark AE; Gunsteren WFV A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. Journal of Computational Chemistry 2004, 25, 1656–1676. [PubMed: 15264259]

(9). Jorgensen WL; Madura JD; Swenson CJ Optimized Intermolecular Potential Functions for Liquid Hydrocarbons. Journal of the American Chemical Society 1984, 106, 6638–6646.

(10). Monticelli L, Salonen E, Eds. Biomolecular Simulations: Methods and Protocols; Methods in Molecular Biology; Humana Press, 2013.

(11). Horta BAC; Merz PT; Fuchs PFJ; Dolenc J; Riniker S; Hünenberger PH A GROMOS-Compatible Force Field for Small Organic Molecules in the Condensed Phase: The 2016H66 Parameter Set. Journal of Chemical Theory and Computation 2016, 12, 3825–3850. [PubMed: 27248705]

(12). Yin D; MacKerell AD Combined Ab Initio/Empirical Approach for Optimization of Lennard–Jones Parameters. Journal of Computational Chemistry 1998, 19, 334–348.

(13). Chen IJ; Yin D; MacKerell AD Combined Ab Initio/Empirical Approach for Optimization of Lennard-Jones Parameters for Polar-Neutral Compounds. Journal of Computational Chemistry 2002, 23, 199–213. [PubMed: 11924734]

(14). Cole DJ; Vilseck JZ; Tirado-Rives J; Payne MC; Jorgensen WL Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning. Journal of Chemical Theory and Computation 2016, 12, 2312–2323. [PubMed: 27057643]

(15). Kantonen SM; Muddana HS; Schauperl M; Henriksen NM; Wang L-P; Gilson MK Data-Driven Mapping of Gas-Phase Quantum Calculations to General Force Field Lennard-Jones Parameters. Journal of Chemical Theory and Computation 2020, 16, 1115–1127. [PubMed: 31917572]

(16). Mohebifar M; Johnson ER; Rowley CN Evaluating Force-Field London Dispersion Coefficients Using the Exchange-Hole Dipole Moment Model. Journal of Chemical Theory and Computation 2017, 13, 6146–6157. [PubMed: 29149556]

(17). Dauber-Osguthorpe P; Hagler AT Biomolecular Force Fields: Where Have We Been, Where Are We Now, Where Do We Need to Go and How Do We Get There? Journal of Computer-Aided Molecular Design 2019, 33, 133–203. [PubMed: 30506158]

(18). Hagler AT Force Field Development Phase II: Relaxation of Physics-Based Criteria… or Inclusion of More Rigorous Physics into the Representation of Molecular Energetics. Journal of Computer-Aided Molecular Design 2019, 33, 205–264. [PubMed: 30506159]

(19). Berendsen HJC; Grigera JR; Straatsma TP The Missing Term in Effective Pair Potentials. The Journal of Physical Chemistry 1987, 91, 6269–6271.

(20). Swope WC; Horn HW; Rice JE Accounting for Polarization Cost When Using Fixed Charge Force Fields. I. Method for Computing Energy. The Journal of Physical Chemistry B 2010, 114, 8621–8630. [PubMed: 20540503]

(21). Swope WC; Horn HW; Rice JE Accounting for Polarization Cost When Using Fixed Charge Force Fields. II. Method and Application for Computing Effect of Polarization Cost on Free Energy of Hydration. The Journal of Physical Chemistry B 2010, 114, 8631–8645. [PubMed: 20540502]

(22). Muddana HS; Sapra NV; Fenley AT; Gilson MK The SAMPL4 Hydration Challenge: Evaluation of Partial Charge Sets with Explicit-Water Molecular Dynamics Simulations. Journal of Computer-Aided Molecular Design 2014, 28, 277–287. [PubMed: 24477800]

(23). Cerutti DS; Rice JE; Swope WC; Case DA Derivation of Fixed Partial Charges for Amino Acids Accommodating a Specific Water Model and Implicit Polarization. The Journal of Physical Chemistry B 2013, 117, 2328–2338. [PubMed: 23379664]

(24). Clague ADH; Bernstein HJ The Heat of Dimerization of Some Carboxylic Acids in the Vapour Phase Determined by a Spectroscopic Method. Spectrochimica Acta Part A: Molecular Spectroscopy 1969, 25, 593–596.

(25). Konicek J; Wadsö I; Munch-Petersen J; Ohlson R; Shimizu A Enthalpies of Vaporization of Organic Compounds. VII. Some Carboxylic Acids. Acta Chemica Scandinavica 1970, 24, 2612–2616.

(26). Frenkel M; Chiroco RD; Diky V; Dong Q; Marsh KN; Dymond JH; Wakeham WA; Stein SE; Königsberger E; Goodwin ARH XML-based IU-PAC Standard for Experimental, Predicted, and Critically Evaluated Thermodynamic Property Data Storage and Capture (ThermoML) (IUPAC Recommendations 2006). Pure and Applied Chemistry 2006, 78, 541–612.

(27). Majer V; Svoboda V; Kehiahan H Enthalpies of Vaporization of Organic Compounds: A Critical Review and Data Compilation; Blackwell Scientific Oxford, 1985; Vol. 32.

(28). Pontolillo J; Eganhouse R The Search for Reliable Aqueous Solubility (Sw) and Octanol-Water Partition Coefficient (Kow) Data for Hydrophobic Organic Compounds; DDT and DDE as a Case Study; USGS Numbered Series 2001–4201, 2001.

(29). Kamath G; Georgiev G; Potoff JJ Molecular Modeling of Phase Behavior and Microstructure of Acetone-Chloroform-Methanol Binary Mixtures. The Journal of Physical Chemistry B 2005, 109, 19463–19473. [PubMed: 16853515]

(30). Chapman WG; Gubbins KE; Jackson G; Radosz M SAFT: Equation-of-state Solution Model for Associating Fluids. Fluid Phase Equilibria 1989, 52, 31–38.

(31). Lobanova O; Mejía A; Jackson G; Müller EA SAFT-γ Force Field for the Simulation of Molecular Fluids 6: Binary and Ternary Mixtures Comprising Water, Carbon Dioxide, and n-Alkanes. The Journal of Chemical Thermodynamics 2016, 93, 320–336.

(32). Rahman S; Lobanova O; Jiménez-Serratos G; Braga C; Raptis V; Müller EA; Jackson G; Avendaño C; Galindo A SAFT-γ Force Field for the Simulation of Molecular Fluids. 5. Hetero-Group Coarse-Grained Models of Linear Alkanes and the Importance of Intramolecular Interactions. The Journal of Physical Chemistry B 2018, 122, 9161–9177. [PubMed: 30179489]

(33). Herdes C; Ervik Å; Mejía A; Müller EA Prediction of the Water/Oil Interfacial Tension from Molecular Simulations Using the Coarse-Grained SAFT-γ Mie Force Field. Fluid Phase Equilibria 2018, 476, 9–15.

(34). Zheng L; Bresme F; Trusler JPM; Müller EA Employing SAFT Coarse-Grained Force Fields for the Molecular Simulation of Thermodynamic and Transport Properties of CO2–n-Alkane Mixtures. Journal of Chemical & Engineering Data 2020, 65, 1159–1171.

(35). Weerasinghe S; Smith PE A Kirkwood–Buff Derived Force Field for Sodium Chloride in Water. The Journal of Chemical Physics 2003, 119, 11342–11349.

(36). Ploetz EA; Bentenitis N; Smith PE Developing Force Fields from the Microscopic Structure of Solutions. Fluid Phase Equilibria 2010, 290, 43–47. [PubMed: 20161692]

(37). Ploetz EA; Smith PE A Kirkwood–Buff Force Field for the Aromatic Amino Acids. Physical Chemistry Chemical Physics 2011, 13, 18154–18167. [PubMed: 21931889]

(38). Ploetz EA; Karunaweera S; Bentenitis N; Chen F; Dai S; Gee MB; Jiao Y; Kang M; Kariyawasam NL; Naleem N; Weerasinghe S; Smith PE Kirkwood–Buff-Derived Force Field for Peptides and Proteins: Philosophy and Development of KBFF20. Journal of Chemical Theory and Computation 2021, 17, 2964–2990. [PubMed: 33878263]

(39). Fischer J; Möller D; Chialvo A; Haile JM The Influence of Unlike Molecule Interaction Parameters on Liquid Mixture Excess Properties. Fluid Phase Equilibria 1989, 48, 161–176.

(40). Dai J; Li X; Zhao L; Sun H Enthalpies of Mixing Predicted Using Molecular Dynamics Simulations and OPLS Force Field. Fluid Phase Equilibria 2010, 289, 156–165.

(41). Stroet M; Koziara KB; Malde AK; Mark AE Optimization of Empirical Force Fields by Parameter Space Mapping: A Single-Step Perturbation Approach. Journal of Chemical Theory and Computation 2017, 13, 6201–6212. [PubMed: 29125748]

(42). Qiu Y; Smith DGA; Boothroyd S; Jang H; Hahn DF; Wagner J; Bannan CC; Gokey T; Lim VT; Stern CD; Rizzi A; Tjanaka B; Tresadern G; Lucas X; Shirts MR; Gilson MK; Chodera JD; Bayly CI; Mobley DL; Wang L-P Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. Journal of Chemical Theory and Computation 2021,

(43). Wang L-P; Martinez TJ; Pande VS Building Force Fields: An Automatic, Systematic, and Reproducible Approach. The Journal of Physical Chemistry Letters 2014, 5, 1885–1891. [PubMed: 26273869]

(44). Boothroyd S; Wang L-P; Mobley D; Chodera J; Shirts M The Open Force Field Evaluator: An Automated, Efficient, and Scalable Framework for the Estimation of Physical Properties from Molecular Simulation. ChemRxiv 2021,

(45). Qiu Y; Smith DG; Boothroyd S; Wagner J; Bannan CC; Gokey T; Jang H; Lim VT; Lucas X; Tjanaka B; Shirts MR; Gilson MK; Chodera JD; Bayly CI; Mobley DL; Wang L-P Openforcefield/Openforcefields: Version 1.0.0 ”Parsley”. Zenodo, 2019.

(46). Mopsik FI Dielectric Constant of N-Hexane as a Function of Temperature, Pressure, and Density. Journal of Research of the National Bureau of Standards. Section A, Physics and Chemistry 1967, 71A, 287–292.

(47). Pereira SM; Iglesias TP; Legido JL; Rivas MA; Real JN Relative Permittivity Increments for {xCH3OH+ (1 -x)CH3OCH2(CH2OCH2)3CH2OCH3} fromT= 283.15 K toT= 323.15 K. The Journal of Chemical Thermodynamics 2001, 33, 433–440.

(48). Giner B; Villares A; Martín S; Lafuente C; Royo FM Isothermal Vapour–Liquid Equilibrium for Cyclic Ethers with 1-Chloropentane. Fluid Phase Equilibria 2007, 251, 8–16.

(49). Alcalde R; Aparicio S; Dávila MJ; García B; Leal JM Liquid–Liquid Equilibria of Lactam Containing Binary Systems. Fluid Phase Equilibria 2008, 266, 90–100.

(50). Cháfer A; Lladosa E; de la Torre J; Burguet MC Study of Liquid–Liquid Equilibrium of the Systems Isobutyl Acetate+acetic Acid+water and Isobutyl Alcohol+acetic Acid+water at Different Temperatures. Fluid Phase Equilibria 2008, 271, 76–81.

(51). Wang Y; Gao H; Yan W Excess Molar Enthalpies of Diethyl Malonate+ (1-Butanol, 2-Methyl-1-Propanol, 1-Pentanol, n-Heptane, and Ethyl Acetate) at T= (288.2, 298.2, 313.2, 328.2, 338.2, and 348.2K) and P=101.3kPa. Fluid Phase Equilibria 2010, 291, 8–12.

(52). Cháfer A; de la Torre J; Lladosa E; Montón JB Liquid–Liquid Equilibria of 4-Methyl-2-Pentanone+1-Propanol or 2-Propanol+water Ternary Systems: Measurements and Correlation at Different Temperatures. Fluid Phase Equilibria 2014, 361, 23–29.

(53). Keshapolla D; Singh V; Gupta A; Gardas RL Apparent Molar Properties of Benzyldimethylammonium Based Protic Ionic Liquids in Water and Ethanol at Different Temperatures. Fluid Phase Equilibria 2015, 385, 92–104.

(54). Martínez-Baños L; Embid JM; Otín S; Artal M Vapour–Liquid Equilibrium at T=308.15K for Binary Systems: Dibromomethane+n-Heptane, Bromotrichloromethane+n-Heptane, Bromotrichloromethane+dibromomethane, Bromotrichloromethane+bromochloromethane and Dibromomethane+bromochloromethane. Experimental Data and Modelling. Fluid Phase Equilibria 2015, 395, 1–8.

(55). Requejo PF; Calvar N; Domíınguez Á; Gómez E Application of the Ionic Liquid Tributylmethylammonium Bis(Trifluoromethylsulfonyl)Imide as Solvent for the Extraction of Benzene from Octane and Decane at T = 298.15 K and Atmospheric Pressure. Fluid Phase Equilibria 2016, 417, 137–143.

(56). Ortega J; Navas A; Plácido J Thermodynamic Study of (Alkyl Esters+α,ω-Alkyl Dihalides) IV: HmEandVmE for 25 Binary Mixtures {xCu-1H2u-1CO2CH3+(1-x)α,ω-BrCH2(CH2)v-2CH2Br}, Where U=1 to 5, A=1 and V=ω=2 to 6. The Journal of Chemical Thermodynamics 2007, 39, 128–141.

(57). Dragoescu D; Teodorescu M; Barhala A Isothermal (Vapour+liquid) Equilibria and Excess Gibbs Free Energies in Some Binary (Cyclopentanone+chloroalkane) Mixtures at Temperatures from 298.15K to 318.15K. The Journal of Chemical Thermodynamics 2007, 39, 1452–1457.

(58). Tôrres RB; Ortolan MI; Volpe PLO Volumetric Properties of Binary Mixtures of Ethers and Acetonitrile: Experimental Results and Application of the Prigogine–Flory–Patterson Theory. The Journal of Chemical Thermodynamics 2008, 40, 442–459.

(59). Morávková L; Wagner Z; Linek J Volumetric Behaviour of Binary Liquid Systems Composed of Toluene, Isooctane, and Methyl Tert-Butyl Ether at Temperatures from (298.15 to 328.15)K. The Journal of Chemical Thermodynamics 2009, 41, 591–597.

(60). Dragoescu D; Barhala A; Teodorescu M (Vapour+liquid) Equilibria and Excess Gibbs Free Energies of (Cyclohexanone+1-Chlorobutane And+1,1,1-Trichloroethane) Binary Mixtures at Temperatures from (298.15 to 318.15)K. The Journal of Chemical Thermodynamics 2009, 41, 1025–1029.

(61). Ghanadzadeh Gilani H; Ghanadzadeh Gilani A; Shekarsaraee S; Uslu H (Liquid+liquid) Equilibrium Data of (Water+phosphoric Acid+solvents) Systems at T=(308.2 and 318.2)K. The Journal of Chemical Thermodynamics 2012, 53, 52–59.

(62). Cobos A; Hevia F; González JA; García De La Fuente, I.; Alonso Tristán, C. Thermodynamics of Amide+ketone Mixtures. 1. Volumetric, Speed of Sound and Refractive Index Data for N,N-dimethylformamide+2-Alkanone Systems at Several Temperatures. The Journal of Chemical Thermodynamics 2016, 98, 21–32.

(63). Daoudi H; Ait kaci A; Tafat-Igoudjilene O Volumetric Properties of Binary Liquid Mixtures of Alcohols with 1,2-Dichloroethane at Different Temperatures and Atmospheric Pressure. Thermochimica Acta 2012, 543, 66–73.

(64). Sharma VK; Malik S; Solanki S Thermodynamic Studies of Molecular Interactions in Mixtures Containing Tetrahydropyran, 1,4-Dioxane, and Cyclic Ketones. Journal of Chemical & Engineering Data 2017, 62, 623–632.

(65). Matsuda H; Inaba K; Nishihara K; Sumida H; Kurihara K; Tochigi K; Ochi K Separation Effects of Renewable Solvent Ethyl Lactate on the Vapor–Liquid Equilibria of the Methanol + Dimethyl Carbonate Azeotropic System. Journal of Chemical & Engineering Data 2017, 62, 2944–2952.

(66). Ouyang G; Huang Z; Ou J; Wu W; Kang B Excess Molar Volumes and Surface Tensions of Xylene with 2-Propanol or 2-Methyl-2-propanol at 298.15 K. Journal of Chemical & Engineering Data 2003, 48, 195–197.

(67). George J; Sastry NV Densities, Excess Molar Volumes at T = (298.15 to 313.15) K, Speeds of Sound, Excess Isentropic Compressibilities, Relative Permittivities, and Deviations in Molar Polarizations at T = (298.15 and 308.15) K for Methyl Methacrylate + 2-Butoxyethanol or Dibutyl Ether + Benzene, Toluene, or p-Xylene. Journal of Chemical & Engineering Data 2004, 49, 1116–1126.

(68). Kato M; Kodama D; Sato M; Sugiyama K Volumetric Behavior and Saturated Pressure for Carbon Dioxide + Ethyl Acetate at a Temperature of 313.15 K. Journal of Chemical & Engineering Data 2006, 51, 1031–1034.

(69). Ranjbar S; Momenian SH Densities and Viscosities of Binary and Ternary Mixtures of (Nitrobenzene + 1-Bromobutane), (1-Bromobutane + Methylcyclohexane), (Nitrobenzene + Methylcyclohexane), and (Methylcyclohexane + Nitrobenzene + 1-Bromobutane) from (293.15 to 308.15) K. Journal of Chemical & Engineering Data 2011, 56, 3949–3954.

(70). Dohnal V; Řehák K Thermal and Volumetric Properties of Four Aqueous Aroma Compounds at Infinite Dilution. Journal of Chemical & Engineering Data 2012, 57, 1822–1828.

(71). Zorebski E; Waligóra A Densities, Excess Molar Volumes, and Isobaric Thermal Expansibilities for 1,2-Ethanediol + 1-Butanol, or 1-Hexanol, or 1-Octanol in the Temperature Range from (293.15 to 313.15) K. Journal of Chemical & Engineering Data 2008, 53, 591–595.

(72). Postigo MA; Mariano AB; Jara AF; Zurakoski N Isobaric Vapor-Liquid Equilibria for the Binary Systems Benzene + Methyl Ethanoate, Benzene + Butyl Ethanoate, and Benzene + Methyl Heptanoate at 101.31 kPa. Journal of Chemical & Engineering Data 2009, 54, 1575–1579.

(73). Cihlář J; Hynek V; Svoboda V; Holub R Heats of Vaporization of Alkyl Esters of Formic Acid. Collection of Czechoslovak Chemical Communications 1976, 41, 1–6.

(74). Majer V; Wagner Z; Svoboda V; Čadek V Enthalpies of Vaporization and Cohesive Energies for a Group of Aliphatic Ethers. The Journal of Chemical Thermodynamics 1980, 12, 387–391.

(75). Majer V; Svoboda V; Hála S; Pick J Temperature Dependence of Heats of Vaporization of Saturated Hydrocarbons C5-C8; Experimental Data and an Estimation Method. Collection of Czechoslovak Chemical Communications 1979, 44, 637–651.

(76). Snelson A; Skinner HA Heats of Combustion: Sec-Propanol, 1,4-Dioxan, 1,3-Dioxan and Tetrahydropyran. Transactions of the Faraday Society 1961, 57, 2125–2131.

(77). Svoboda V; Uchytilová V; Majer V; Pick J Heats of Vaporization of Alkyl Esters of Formic, Acetic and Propionic Acids. Collection of Czechoslovak Chemical Communications 1980, 45, 3233–3240.

(78). Majer V; Svoboda V; Uchytilová V; Finke M Enthalpies of Vaporization of Aliphatic C5 and C6 Alcohols. Fluid Phase Equilibria 1985, 20, 111–118.

(79). Uchytilová V; Majer V; Svoboda V; Hynek V Enthalpies of Vaporization and Cohesive Energies for Seven Aliphatic Ketones. The Journal of Chemical Thermodynamics 1983, 15, 853–858.

(80). Byström K; Månsson M Enthalpies of Formation of Some Cyclic 1,3- and 1,4-Di- and Poly-Ethers: Thermochemical Strain in the –O–C–O– and –O–C–C–O– Groups. Journal of the Chemical Society, Perkin Transactions 2 1982, 565–569.

(81). Wolf G Thermochemische Untersuchungen an Cyclischen Ketonen. Helvetica Chimica Acta 1972, 55, 1446–1459.

(82). Wadsö I; Murto M-L; Bergson G; Ehrenberg L; Brunvoll J; Bunnenberg E; Djerassi C; Records R A Heat of Vaporization Calorimeter for Work at 25 Degrees C and for Small Amounts of Substances. Acta Chemica Scandinavica 1966, 20, 536–543.

(83). Lipp SV; Krasnykh EL; Verevkin SP Vapor Pressures and Enthalpies of Vaporization of a Series of the Symmetric Linear N-Alkyl Esters of Dicarboxylic Acids. Journal of Chemical & Engineering Data 2011, 56, 800–810.

(84). Marenich AV; Kelly CP; Thompson JD; Hawkins GD; Chambers CC; Giesen DJ; Winget P; Cramer CJ; Truhlar DG Minnesota Solvation Database (MNSOL) Version 2012. 2020.

(85). Boothroyd S; Madin O; Wagner J; Setiadi J; Thompson M; Rodríguez-Guerra J Openforcefield/Openff-Evaluator: 0.1.0 OpenFF Evaluator. Zenodo, 2020.

(86). Boothroyd S Common Workflows - OpenFF Evaluator Documentation. https://openff-evaluator.readthedocs.io/en/stable/properties/commonworkflows.html#simulation-layer, Accessed April 13th, 2022.

(87). Martínez L; Andrade R; Birgin EG; Martínez JM PACKMOL: A package for building initial configurations for molecular dynamics simulations. Journal of Computational Chemistry 2009, 30, 2157–2164. [PubMed: 19229944]

(88). Wagner J; Mobley DL; Chodera J; Bannan C; Rizzi A; Camila,; Bayly C; Lim NM; Lim V; Sasmal S; Rodríguez-Guerra J; Zhao Y; Lee-Ping, Openforcefield/Openforcefield: 0.6.0 Library Charges. Zenodo, 2019.

(89). Leimkuhler B; Matthews C Rational Construction of Stochastic Numerical Methods for Molecular Sampling. Applied Mathematics Research eXpress 2013, 2013, 34–56.

(90). Chodera JD A Simple Method for Automated Equilibration Detection in Molecular Simulations. Journal of Chemical Theory and Computation 2016, 12, 1799–1805. [PubMed: 26771390]

(91). Rizzi A; Chodera J; Naden L; Beauchamp K; Albanese S; Grinaway P; Prada-Gracia D; Rustenburg B; ajsilveira,; Saladi S; Boehm K; Gmach J; Rodríguez-Guerra J Choderalab/Yank: 0.25.2 - Bugfix Release. Zenodo, 2019.

(92). Wang K; Chodera JD; Yang Y; Shirts MR Identifying Ligand Binding Sites and Poses Using GPU-accelerated Hamiltonian Replica Exchange Molecular Dynamics. Journal of Computer-Aided Molecular Design 2013, 27, 989–1007. [PubMed: 24297454]

(93). Shivakumar D; Williams J; Wu Y; Damm W; Shelley J; Sherman W Prediction of Absolute Solvation Free Energies Using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. Journal of Chemical Theory and Computation 2010, 6, 1509–1519. [PubMed: 26615687]

(94). Shirts MR; Chodera JD Statistically Optimal Analysis of Samples from Multiple Equilibrium States. The Journal of Chemical Physics 2008, 129, 124105. [PubMed: 19045004]

(95). Levenberg K A Method for the Solution of Certain Non-Linear Problems in Least Squares. Quarterly of Applied Mathematics 1944, 2, 164–168.

(96). Moré JJ; Sorensen DC Computing a Trust Region Step. SIAM Journal on Scientific and Statistical Computing 1983, 4, 553–572.

(97). Dennis JE; Gay DM; Walsh RE An Adaptive Nonlinear Least-Squares Algorithm. ACM Transactions on Mathematical Software 1981, 7, 348–368.

(98). Wang L-P; McKiernan KA; Gomes J; Beauchamp KA; Head-Gordon T; Rice JE; Swope WC; Martíınez TJ; Pande VS Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. The Journal of Physical Chemistry B 2017, 121, 4023–4039. [PubMed: 28306259]

(99). Mobley DL; Bannan CC; Rizzi A; Bayly CI; Chodera JD; Lim VT; Lim NM; Beauchamp KA; Slochower DR; Shirts MR; Gilson MK; Eastman PK Escaping Atom Types in Force Fields Using Direct Chemical Perception. Journal of Chemical Theory and Computation 2018, 14, 6076–6092. [PubMed: 30351006]

(100). Cornell WD; Cieplak P; Bayly CI; Gould IR; Merz KM; Ferguson DM; Spellmeyer DC; Fox T; Caldwell JW; Kollman PA A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. Journal of the American Chemical Society 1995, 117, 5179–5197.

(101). Mobley DL; Bannan CC; Rizzi A; Bayly CI; Chodera JD; Lim VT; Lim NM; Beauchamp KA; Shirts MR; Gilson MK; Eastman PK Open Force Field Consortium: Escaping Atom Types Using Direct Chemical Perception with SMIRNOFF v0.1. bioRxiv 2018, 286542.

(102). Rosenblatt M Remarks on Some Nonparametric Estimates of a Density Function. The Annals of Mathematical Statistics 1956, 27, 832–837.

(103). Parzen E On Estimation of a Probability Density Function and Mode. The Annals of Mathematical Statistics 1962, 33, 1065–1076.

(104). Waskom ML Seaborn: Statistical Data Visualization. Journal of Open Source Software 2021, 6, 3021.

(105). Scott DW Multivariate Density Estimation: Theory, Practice, and Visualization, 2nd ed.; John Wiley & Sons: New York, 1992.

(106). Jorgensen WL; Tirado-Rives J The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and

Crambin. Journal of the American Chemical Society 1988, 110, 1657–1666. [PubMed: 27557051]

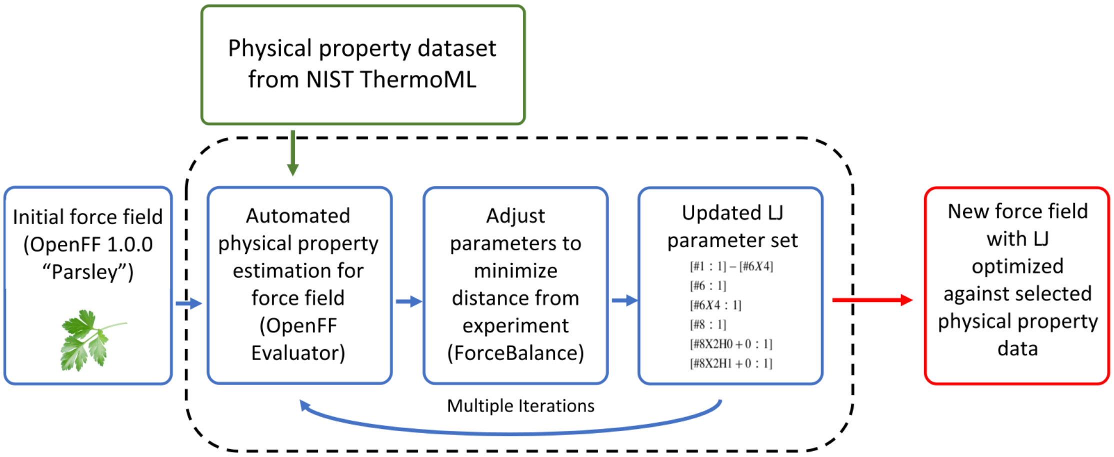

Physical property dataset from NIST ThermoML
Initial force field (OpenFF 1.0.0 “Parsley”)
Automated physical property estimation for force field (OpenFF Evaluator)
Adjust parameters to minimize distance from experiment (ForceBalance)
Updated LJ parameter set
[#1 : 1] – [#6X4]
[#6 : 1]
[#6X4 : 1]
[#8 : 1]
[#8X2H0 + 0 : 1]
[#8X2H1 + 0 : 1]
New force field with LJ optimized against selected physical property data
Multiple Iterations

Figure 1: LJ optimization workflow used in this study.

A training dataset consisting of physical property measurements for organic molecules is selected from the NIST ThermoML database. Starting with the OpenFF 1.0.0 (Parsley) force field, the physical properties in the training dataset are estimated using the force field and the OpenFF Evaluator software package.LJ parameters are then adjusted by minimizing the difference between the simulation results and experimental training data via a regularized least-squares procedure as implemented in the ForceBalance package.[43]

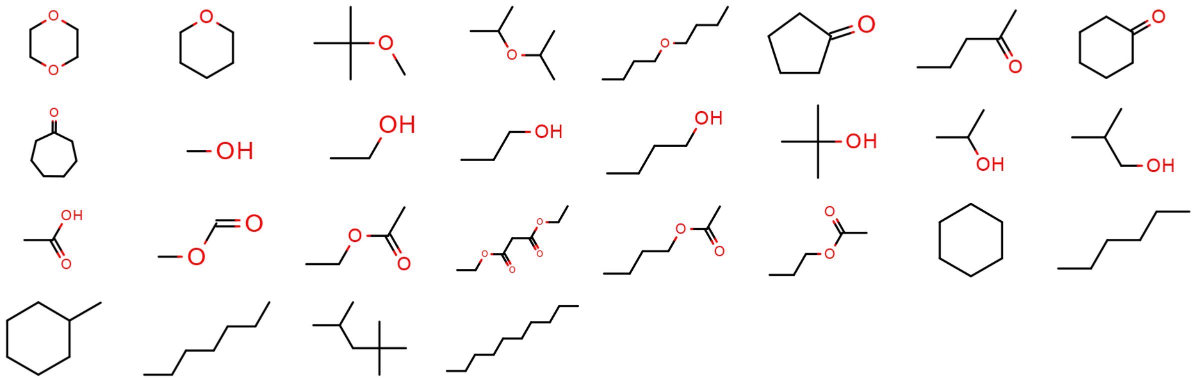


Figure 2: The 28 molecules were included in the “pure only” training set.

The pure data used in our training sets contains one ρ_l and one ΔH_{vap} measurement per molecule, measured at close to ambient conditions (P=1 atm, T=298K), yielding a training set of 28 molecules with 56 data points total.

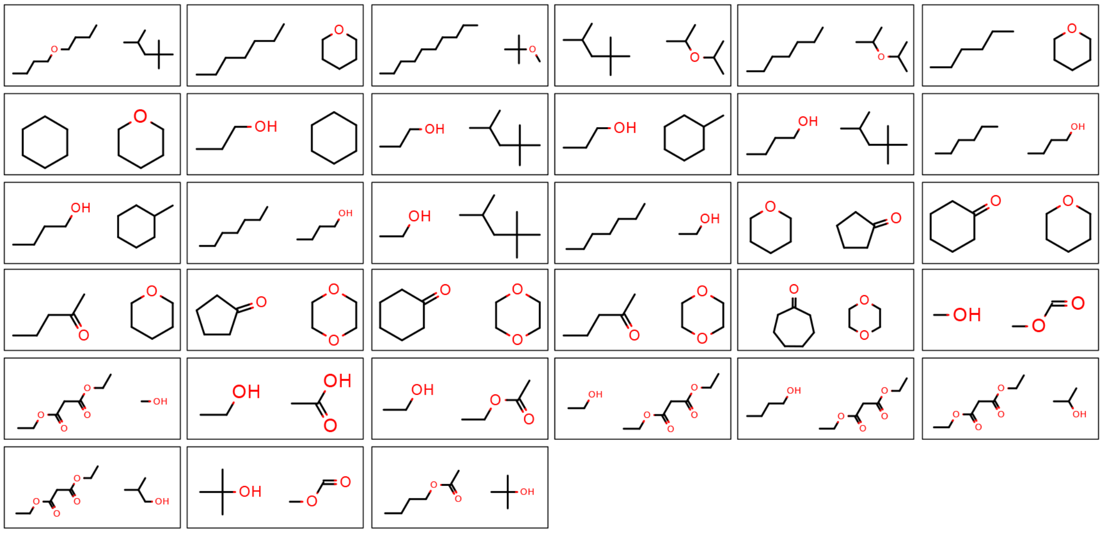


Figure 3: The 33 pairs of molecules (shown as boxed pairs) which were chosen for the mixture training sets.

The mixture data used in our training sets contains one $\rho_l\,(x)$ and one $\Delta H_{mix}\,(x)$ measurement per mixture for three different compositions if multiple compositions were available (close to 25%, 50% and 75%) measured at close to ambient conditions (P=1 atm, T=298 K), yielding a training set of 33 binary mixtures with 187 data points total.

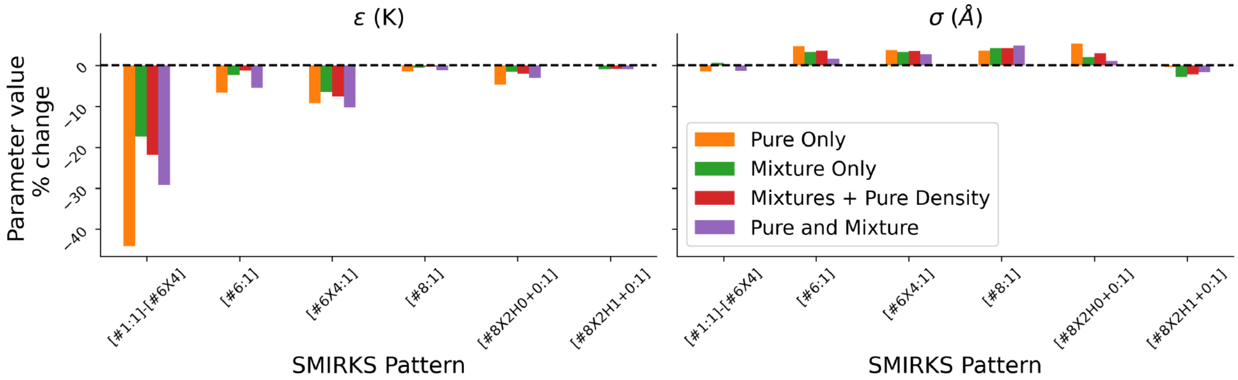


Figure 4: The 4 different training sets generally drive the parameters in the same direction and to similar magnitudes, indicating all data sets encode somewhat similar parameter information.

Changes in parameter values for each of the training sets considered in this paper are shown as bar graphs above. The percent change in the each parameter for each of the training sets relative to their starting value taken from the OpenFF 1.0.0 force field. One notable difference between the "pure only" set and the sets containing mixtures is the `[#8X2H1+0:1]` (hydroxyl oxygen) σ parameter, which is reduced by only 0.4% in the "pure only" (orange) set, but reduced by 1.7–2.8% in the other sets.

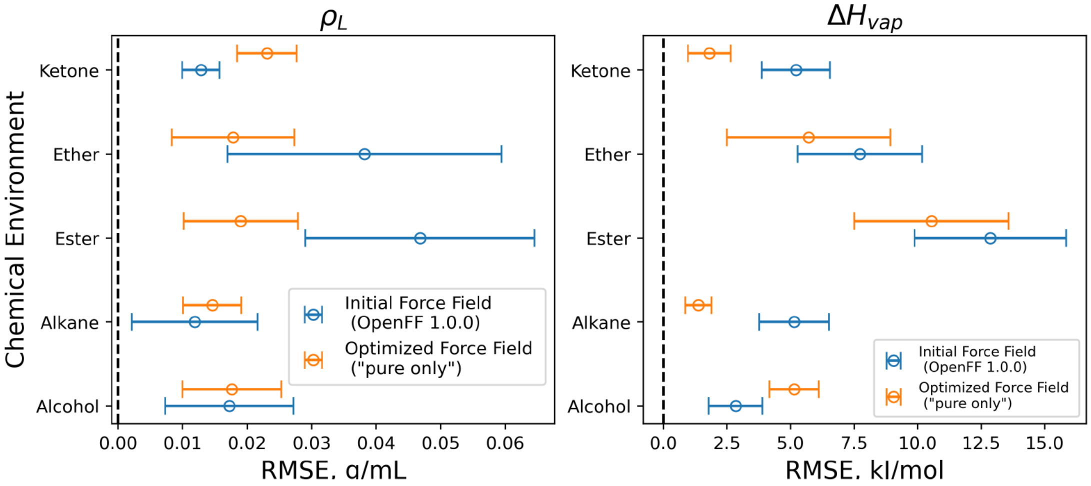


Figure 5: Optimization generally improves training set RMSEs of pure properties for all force fields trained against pure properties.

Figure shows categorized RMSE vs. experiment of ρ_l (left panel) and H_{vap} (right panel) measurements in the "pure only" training set, estimated using the initial parameters (OpenFF 1.0.0, blue points) and the final parameters after 12 optimization iterations ("pure only", orange points). RMSEs are categorized by chemical environment, and error bars represent 95% confidence intervals computed by bootstrapping with replacement for 1000 iterations. The results from the other training sets containing pure properties ("mixtures and pure density", "pure and mixture") are statistically equivalent, with the exception of ketone pure densities (statistically better in the "pure only" set), and alcohol heats of vaporization (statistically inferior in the "pure only" dataset). Figures for other optimization are available in Supporting Information Section S2.6.2,S2.7.2.

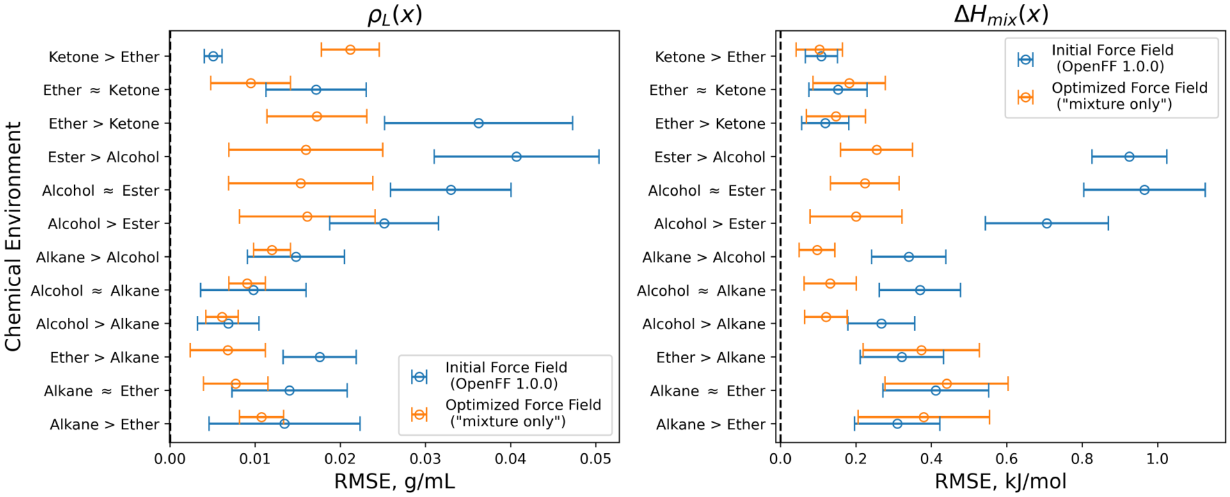


Figure 6: Optimization improves RMSEs of mixture properties for all training sets.

Figure shows categorized RMSE vs. experiment of $\rho_l(x)$ (left panel) and $\Delta H_{mix}(x)$ (right panel) measurements in the “mixture only” training set, estimated using the initial parameters (OpenFF 1.0.0, blue points) and the final parameters after 12 iterations (“mixture only”, orange points). RMSEs are categorized by chemical environment, where “Ether > Ketone” denotes a mixture with ether molecules in excess of ketone molecules, and “Ether ≈ Ketone” denotes a mixture with ether and ketone molecules in roughly equal compositions, etc. Error bars represent 95% confidence intervals computed by bootstrapping with replacement for 1000 iterations. The results from the other training sets containing mixture properties (“mixtures and pure density”, “pure and mixture”) show statistically equivalent improvements in training set RMSEs, and are available in Supporting Information Sections S2.6.2,S2.7.2

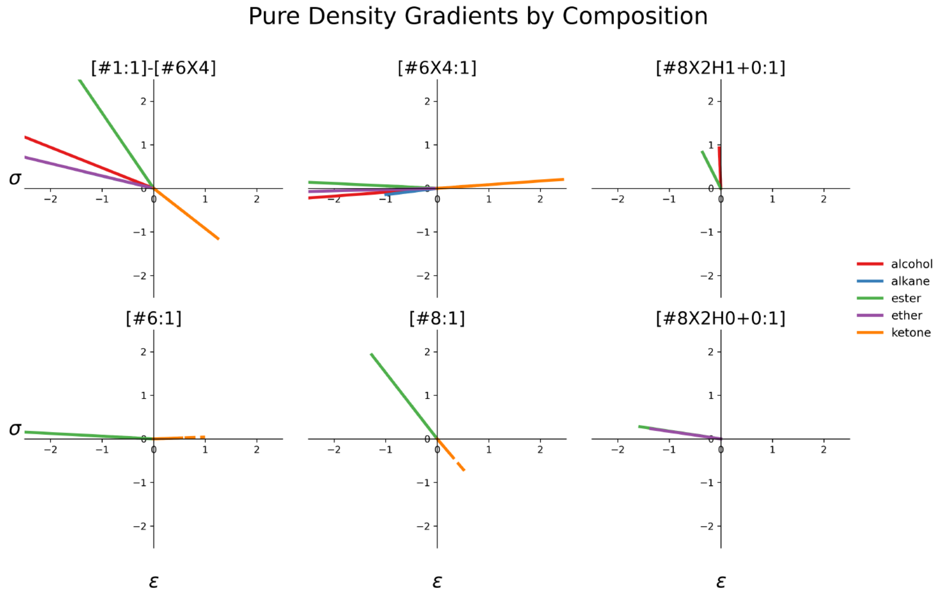


Figure 7: Parameter gradients indicate that ketone measurements targets drive parameters for hydrogen, carbon, and oxygen in opposite directions as other moieties.

The data show the contribution of the first derivatives of the force field parameters to the pure density data portion of the objective function for the "pure only" training set. Dotted lines correspond to the same moieties as solid lines of the same color, and indicate that magnitude of gradient is small, and is shown enlarged to a magnitude of 1 in this figure. The data indicate that the ketone measurements in the training set (orange dotted line) are pulling the hydrogen parameter `[#1:1]-[#6X4]`, general tetravalent carbon parameter `[#6X4]`, and generic oxygen parameter `[#8:1]` in opposite directions from the other chemical environments (all other lines). This suggests that adding a separate parameter (or parameters) to explicitly address ketone environments is likely to improve parameterization.

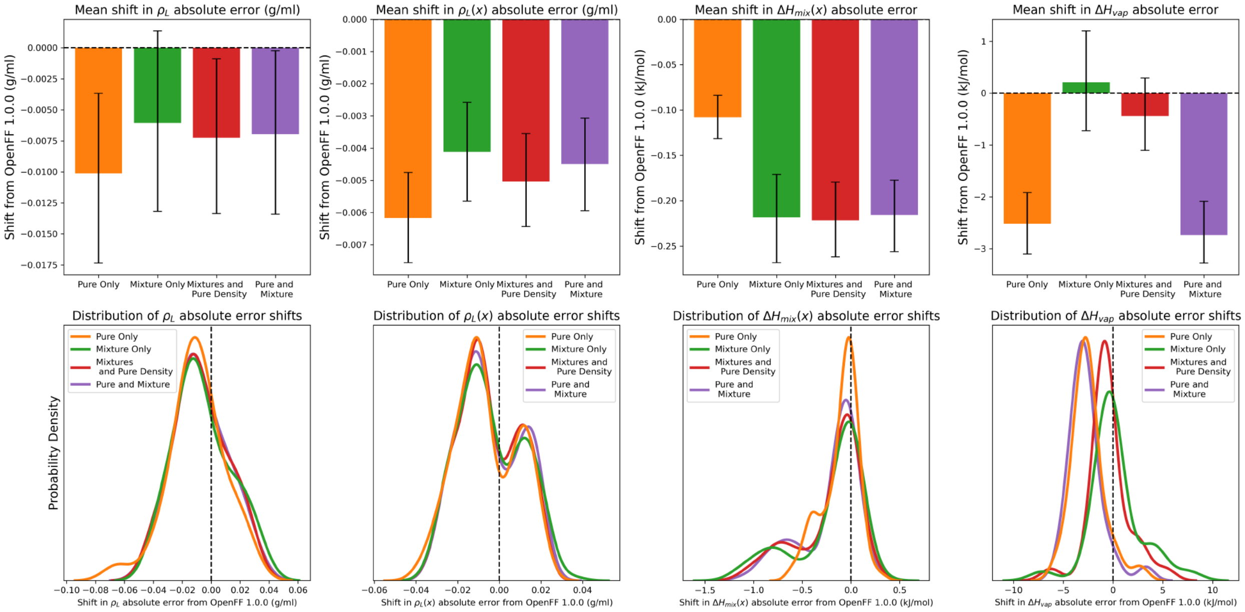


Figure 8: Benchmarking metrics for the test data sets of ρ_l, $\rho_l(x)$, $\Delta H_{mix}(x)$, ΔH_{vap}. Benchmarking indicates that densities are well predicted in all cases, test set ΔH_{vap} is improved when training against ΔH_{vap}, and test set $\Delta H_{mix}(x)$ is improved when training against $\Delta H_{mix}(x)$.

Upper panels show mean shift in absolute error from OpenFF 1.0.0 (the starting point for each of these optimizations). Negative values indicate that the refitted force field's performance on the test set is improved related to OpenFF 1.0.0. Lower panels show kernel density estimates of the distribution of absolute error shifts from OpenFF 1.0.0. Negative values indicate improvement relative to OpenFF 1.0.0, whereas positive values indicate degradation. Error bars in upper panels represent 95% confidence intervals, bootstrapped over pairs of measurements between OpenFF 1.0.0 and the refitted force fields.

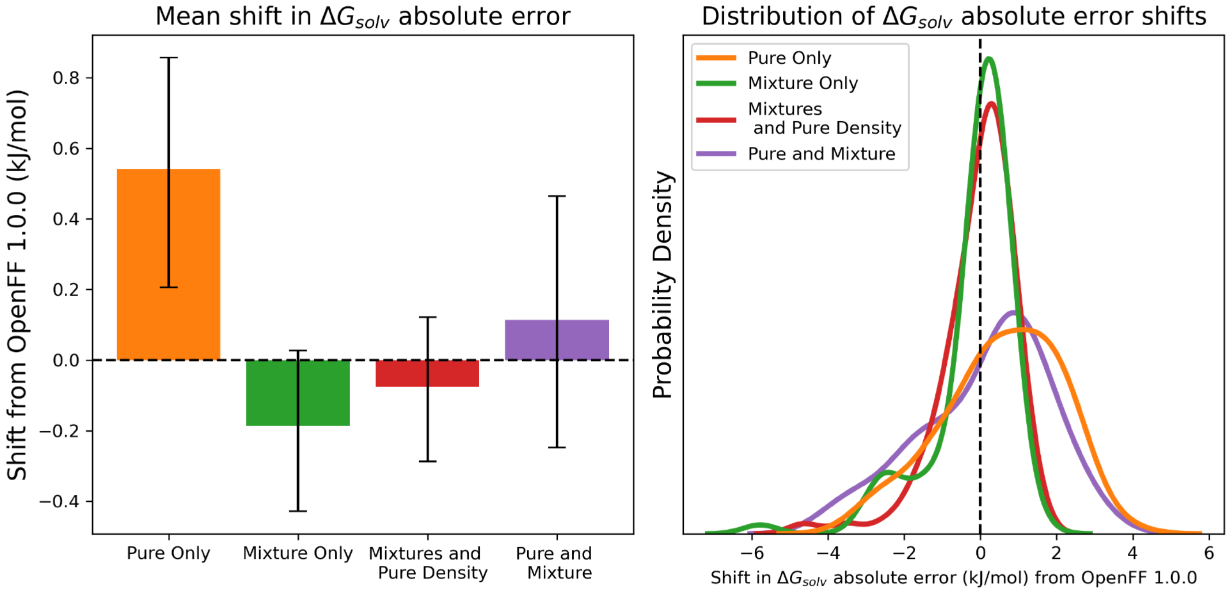


Figure 9: Benchmarking metrics for the ΔG_{solv} test data set shows that training against "mixture only" set improves ΔG_{solv} predictions, whereas training against "pure only" set degrades predictions.

Left panel shows mean shift in absolute error from OpenFF 1.0.0 (the starting point for each of these optimizations). Negative values indicate that the refitted force field's performance on the test set is improved related to OpenFF 1.0.0. Right panel shows kernel density estimate of the distribution of absolute error shifts from OpenFF 1.0.0. Negative values indicate improvement relative to OpenFF 1.0.0, whereas positive values indicate degradation. Error bars in upper panels represent 95% confidence intervals, bootstrapped over pairs of measurements between OpenFF 1.0.0 and the refitted force fields.

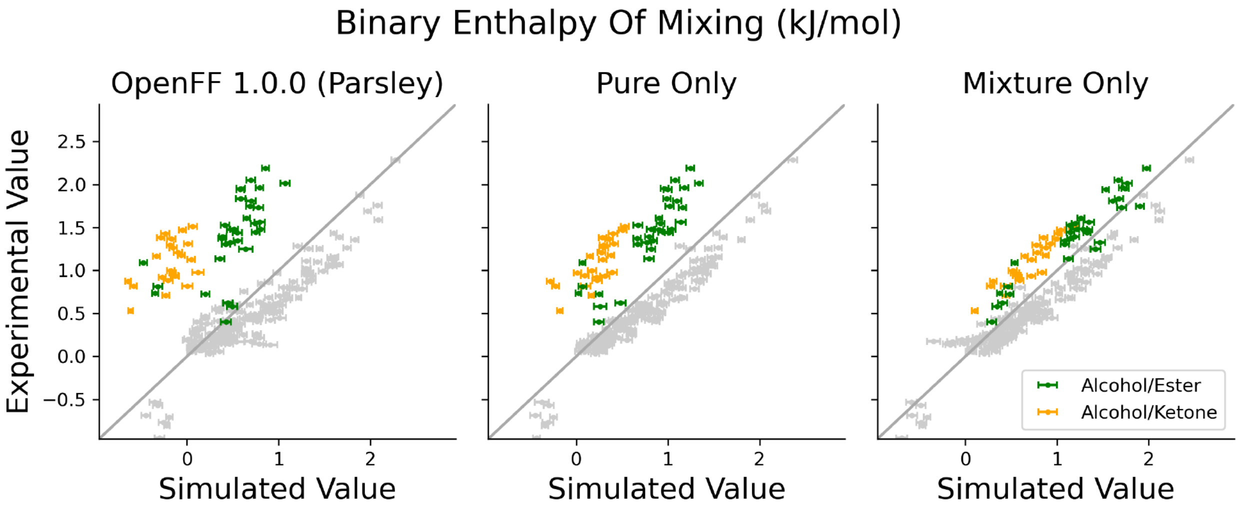


Figure 10: Training against measurements of liquid mixtures corrects systematic error in alcohol/ester and alcohol/ketone enthalpies of mixing.

This figure shows a comparison of the estimated and experimentally measured $\Delta H_{mix}(x)$ data points for the test set, plotting for force fields optimized against the "mixture only" and "pure only" training sets, as well as the baseline OpenFF 1.0.0 (Parsley) force field. The systematic error in alcohol/ester and alcohol/ketone mixtures (highlighted green and orange points) is significantly reduced when training against the properties of mixture, but not when training against properties of pure systems.

Table 1:

Four training sets containing different combinations of pure and mixture data were considered in this study.

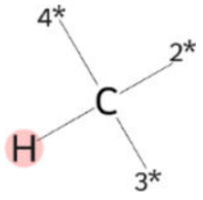
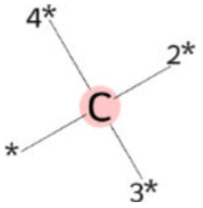
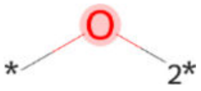
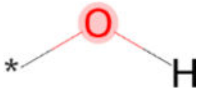
These training sets are composed of measurements of pure-component liquid density (ρ_l), pure-component enthalpy of vaporization (ΔH_{vap}), binary mixture densities ($\rho_l(x)$), and binary enthalpies of mixing ($\Delta H_{mix}(x)$). These measurements cover a set of alcohols, esters, ethers, ketones, acids and alkanes, which is further described in Figures 2 and 3. The 4 training sets in this study are labeled based on which of these measurements are included.

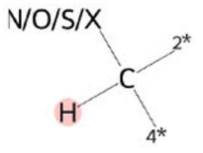
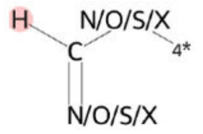
	Properties Included			
Training Data set	**Pure properties**		**Mixture properties**	
	$\boldsymbol{\rho_l}$	$\boldsymbol{\Delta H_{vap}}$	$\boldsymbol{\rho_l(x)}$	$\boldsymbol{\Delta H_{mix}(x)}$
“pure only”	**Yes**	**Yes**	No	No
“mixture only”	No	No	**Yes**	**Yes**
“mixtures + pure density”	**Yes**	No	**Yes**	**Yes**
“pure and mixture”	**Yes**	**Yes**	**Yes**	**Yes**

Table 2:

All atoms with LJ parameter types exercised by the training and test sets, categorized by whether they are re-optimized in this study.

SMIRKS atom types are applied hierarchically, with more specific types superseding less specific types, as described in Mobley et al.[99] Each of these atom types has a σ and ε parameter that describe the Lennard-Jones interactions; with 6 SMIRKS types included in the optimization, 12 Lennard-Jones parameters were optimized. In the "illustration" figures, any atomic index including a '*' is a wildcard, representing any atom, or group of atoms.

SMIRKS Pattern	Description	Illustration
Atoms with Optimized Parameters		
`[#1:1]-[#6X4]`	Hydrogen attached to tetravalent carbon	
`[#6:1]`	Generic carbon	
`[#6X4:1]`	Tetravalent Carbon	
`[#8:1]`	Generic oxygen	
`[#8X2H0+0:1]`	Divalent oxygen attached to zero hydrogens	
`[#8X2H1+0:1]`	Divalent oxygen attached to one hydrogen	

SMIRKS Pattern	Description	Illustration
Atoms with Fixed Parameters		
[#1:1]-[#6X4] -[#7,#8,#9,#16,#17,#35]	Hydrogen attached to tetravalent carbon attached to N/O/S/Halogen	
[#1:1]-[#6X3] (~[#7,#8,#9,#16,#17,#35]) ~[#7,#8,#9,#16,#17,#35]	Hydrogen attached to trivalent carbon attached to 2 N/O/S/Halogen atoms	
[#1:1]-[#8]	Hydrogen attached to generic oxygen