Improving force field accuracy by training against condensed phase mixture properties

Simon Boothroyd[1,2,†], **Owen C. Madin**[1,†], **David L. Mobley**[3,4], **Lee-Ping Wang**[5], **John D. Chodera**[6], **Michael R. Shirts**[1]

[1]Department of Chemical & Biological Engineering, University of Colorado Boulder, Boulder, CO, USA 80309; [2]Boothroyd Scientific Consulting Ltd., 71-75 Shelton Street, London, Greater London, United Kingdom, WC2H 9JQ; [3]Department of Pharmaceutical Sciences, University of California, Irvine, California, USA 92617; [4]Department of Chemistry, University of California, Irvine, CA 92617, United States; [5]Department of Chemistry, University of California; 1 Shields Avenue; Davis, CA, USA 95616; [6]Computational & Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA 10065

*For correspondence:
michael.shirts@colorado.edu (MRS)

[†]These authors contributed equally to this work.

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 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 almost universally improves the force field's ability to reproduce both pure and mixture properties, reducing some systematic errors that exist when training vdW interactions against properties of pure systems only.

1 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 set of parameters for these force fields are the parameters of 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 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 potential better than they do individually. We note that while this is the dominant choice, alternatives exist; notably, the GROMOS 53A5/53A6 force fields use enthalpies of hydration and solvation in addition to ρ_l and ΔH_{vap} [8]. 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 [11, 12]. More recently, methods to produce LJ parameters entirely from *ab initio* data, using atom-in-molecule electron density partitioning [13, 14], or the exchange hole dipole model [15] have been proposed. Still, parameterization against small molecule ρ_l and ΔH_{vap} is the dominant paradigm [16, 17].

Training against ΔH_{vap} in particular is problematic in some 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 [18, 19] 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 [19–21]. Methods suggested include calculating an explicit polarization cost [19] or using semi-polarized charges [13, 22], but the issue has not been definitively resolved. Additionally, some compounds, such as acids, can form clusters in the gas phase [23, 24], 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 [25] 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 [26], rigorous provenance [27], or fully computer-readable forms for older measurements. It is difficult to systematically vet the experimental procedures and outputs for large scale, automated, parameter optimizations without access to large amounts of well-formatted and curated data, which is difficult to find for ΔH_{vap}.

For a fixed charge small molecule force field geared towards biomolecular systems in heterogeneous condensed phase, 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, can be sensitive to interactions between functional groups that are not generally present in the pure substances used to train LJ parameters [28, 29]. This is especially important for capturing solute-solvent interactions.
2. Although computing some properties of mixtures may require multiple simulations, most such prop-

erties (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 not completely dissimilar dielectric constants.

3. 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.

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.

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 density properties 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 [30].

Starting with the OpenFF 1.0.0 (Parsley) force field [31], 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.

2 Methods

2.1 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 data points for each system of interest.
2. Optimizing *only* the selected LJ parameters against the training set using ForceBalance [32] in combination with the OpenFF Evaluator framework [33], starting from the OpenFF 1.0.0 (Parsley) [34] force field parameters.
3. Assessing the performance of the trained force field against the test set, 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 to properties of pure systems. Other force field parameters, namely the valence and electrostatic parameters, were not optimized.

2.1.1 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).
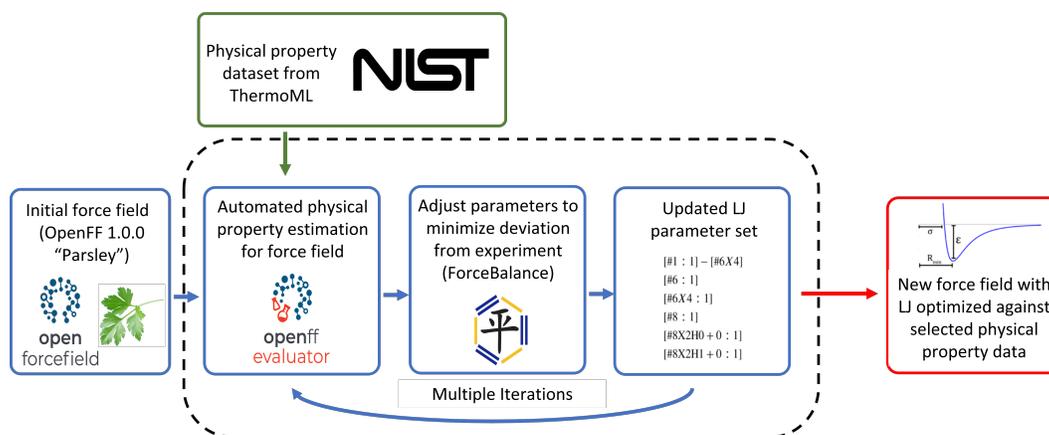
Figure 1. LJ optimization workflow used in this study. A training dataset consisting of physical property measurements for organic molecules is selected from the NIST ThermoML database. Starting with the OpenFF 1.0.0 (Parsley) force field, the physical properties in the training dataset are estimated 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 [32]

	Properties Included			
Training Data set	Pure properties		Mixture properties	
	ρ_l	ΔH_{vap}	$\rho_l(x)$	$\Delta H_{mix}(x)$
"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 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 Figs. 2 and 3. The 4 training sets in this study are labeled based on which of these measurements are included, and are described in Section 2.1.1.

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 systems. 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 predictions 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 [35]) to 35.7 (methanol [36]).

2.2 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, allowing us to focus specifically on the effects of training against mixture data for the molecules chosen. The included molecules exercise a total of 9 LJ types, of which 6 are optimized (Table 2). The three parameters that are not exercised are all hydrogen parameters; an explanation of why they are not optimized is given in Section 2.4.1.

We enforce the criteria that all measurements in the data set contain only the molecules in Fig.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 come from 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.

2.2.1 Pure substance property 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 both long and short chain, branched and unbranched, cyclic and acyclic characteristics where data was available. The ρ_l measurements were sourced from the NIST ThermoML [25] 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 [26], 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 [37–61] and 28 ΔH_{vap} data points) [24, 62–72]. For ΔH_{vap} of acids, measurements were sourced which correspond to an infinitely dilute gas (as computed in [24]), which corresponds to the gas we simulate. This is done because carboxylic acids tend to associate in the gas phase.
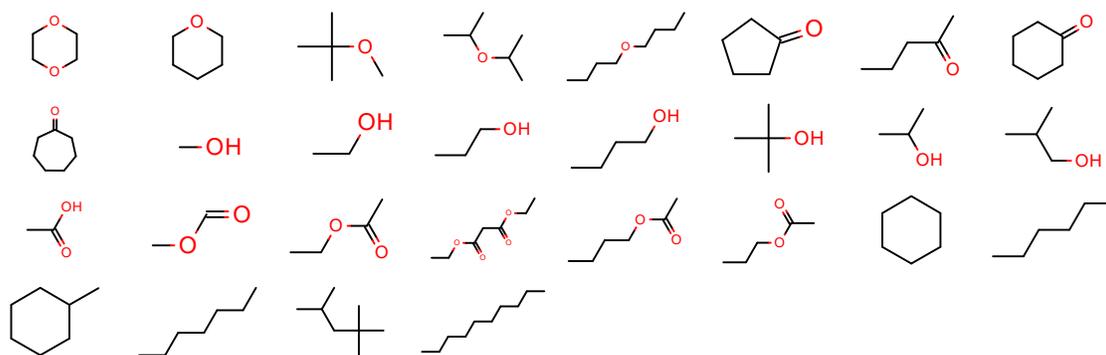

Figure 2. The 28 molecules were included in the "pure only" training set. The pure data used in our training sets contains one ρ_l and one ΔH_{vap} measurement per molecule, measured at close to ambient conditions (P=1 atm, T=298K), yielding a training set of 28 molecules with 56 data points total.

2.2.2 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 [25] archive using the OpenFF Evaluator's built-in data selection tools. For each binary mixture, three $\rho_l(x)$ and three $\Delta H_{mix}(x)$ data points were included, 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 ex-

cess 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 198 measurements. 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 we discovered one $\Delta H_{mix}(x)$ data point (described in supporting information, Section 4) used in the mixture training set that was transcribed into ThermoML incorrectly after training was complete.

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.

2.2.3 Test set

The test set was chosen to include measurements of $\rho_l(x)$, $\Delta H_{mix}(x)$, ρ_l, and ΔH_{vap} data points as in the training set. 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, 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 1.

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 $\Delta H_{mix}(x)$ and 385 $\rho_l(x)$ data points, which was supplemented with a hand-selected test set of 29 ΔH_{vap} and 29 ρ_l pure component measurements.

2.3 Physical property simulations

All estimates of the physical property values were performed using the OpenFF Evaluator [33] package version 0.1.0 [73] using the default estimation workflow schemas, which are outlined in detail in the OpenFF Evaluator documentation [74]. Where possible, simulations are reused to calculate physical properties. For example, mixture densities and mixture enthalpies with a common concentration, temperature, and pressure can be calculated from the same set of simulations.

2.3.1 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 [75], with parameters then assigned using the OpenFF Toolkit version 0.6.0 [76]. 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 [77] 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 [78]. All uncertainties in the average observables were computed by bootstrapping with replacement, and propagated through any further calculations, assuming a Gaussian error model.

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

2.3.2 Enthalpy of Vaporization Calculations

Enthalpies of vaporization require a pure liquid simulation, as described in 2.3.1, 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. This simulation is run for 30 ns instead of the liquid phase 2 ns to converge statistics with only a single molecule.

2.3.3 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 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 1, where the enthalpies of the individual simulated components are multiplied by their mole fractions in the mixture, and then subtracted from the simulated mixture $\Delta H_{mix,x_1,x_2}$.

$$\Delta H_{mix,x_1,x_2} = H_{mix} - x_1 H_1 - x_2 H_2 \qquad (1)$$

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.

2.4 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 $\mathrm{d}O/\mathrm{d}x$ of an observable O with respect to a parameter x is calculated using the central difference method with a relative step size of $h = 10^{-4}$. Values of O at $x - h$ and

$x + h$ are estimated using MBAR reweighting [79], which is accurate for the properties of interest over the step size h. All optimizations were performed using the ForceBalance software package using the built-in OpenFF Evaluator target [32, 33]. Optimizations were run using the Levenberg-Marquardt [80] non-linear least squares algorithm with adaptive trust radius [81, 82] 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 [32, 83]. 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 mathematical parameters of 0.1 for ϵ and 1.0 for σ [32].

$$\chi(\theta) = \sum_n^N \frac{1}{M_n} \sum_m^{M_n} \left(\frac{y_m^{ref} - y_m(\theta)}{d_n} \right)^2 \qquad (2)$$

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, $\sim 50\%$ of the objective function value is due to ρ_l data, and $\sim 50\%$ 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.

2.4.1 Parameters optimized

Both the training and test sets, each containing only molecules composed of carbon, hydrogen, and oxygen, exercise a total of 18 SMIRNOFF parameters (9 different SMIRKS parameter types with one ϵ and σ per SMIRKS). 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 ($\epsilon = 5.27 \times 10^{-5}$) and not reoptimized. This is a slight modification of the AMBER hydroxyl hydrogen parameter [84] (HO, $\epsilon = 0$) to avoid unphysical effects caused by the AMBER parameterization [85]. Here each parameter is uniquely identified by a SMIRKS pattern which encodes the chemical environment to which the parameter will be applied [86]. These parameters, along with brief descriptions, are listed in Table 2.

2.5 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 from section 2.3. Estimated properties for each data point were then compared against the experimental values, with RMSE and average Kendall τ rank correlation [87] values calculated against the test set for each force field.

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	
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 tetravalent carbon attached to 2 N/O/S/Halogen atoms	
`[#1:1]-[#8]`	Hydrogen attached to generic oxygen	
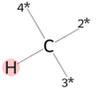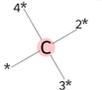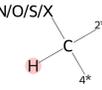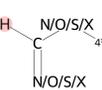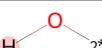
Table 2. All atoms with LJ parameter types exercised by the training and test sets, categorized by whether they are re-optimized in this study. SMIRKS atom types are applied hierarchically, with more specific types superseding less specific types, as described in Mobley et al. [86]. Each of these atom types has both 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.

3 Results & Discussion

3.1 Optimization

3.1.1 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 3.1). 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).
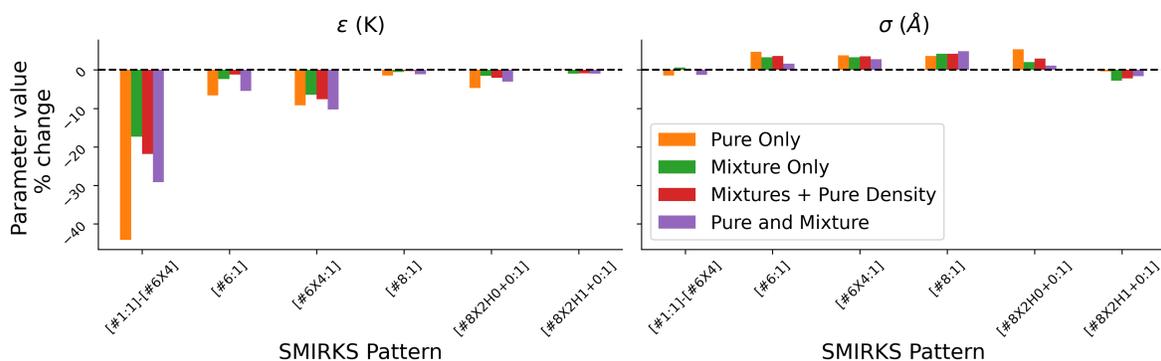
Figure 4. The 4 different training sets generally drive the parameters in the same direction and to similar magnitudes, indicating all data sets encode somewhat similar parameter information. Changes in parameter values for each of the training sets considered in this paper are shown as bar graphs above. 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 nearly unchanged in the "pure only" (orange) set, but reduced by 1.5-3% in the other sets.

3.1.2 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 process sheds light on which parameter changes are driving the specific property improvements that result in an optimized force field. Using the RMSE for each target property measurement and grouping by property and chemical environment as a metric, 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 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 3.4.2,3.5.2).
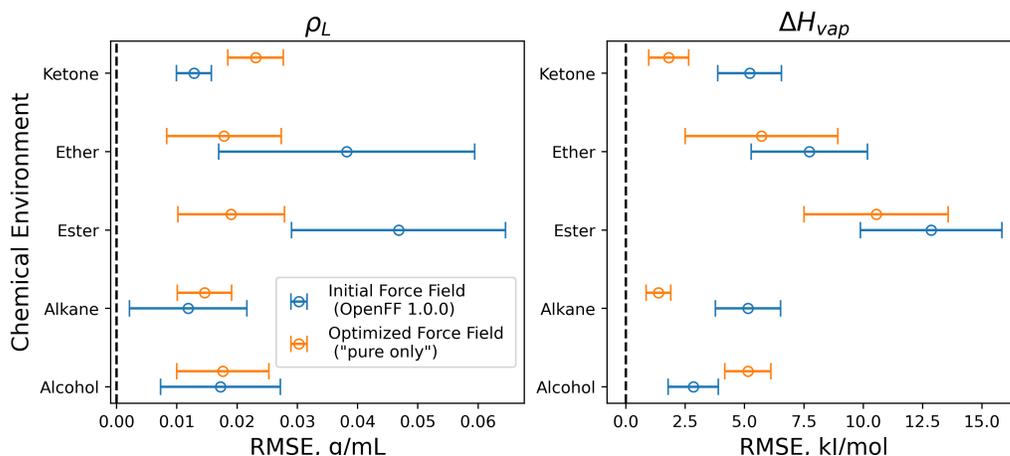
Figure 5. Optimization generally improves RMSEs of pure properties for all training sets. Figure shows categorized RMSE vs. experiment of ρ_l (left panel) and H_{vap} (right panel) measurements in the "pure only" training set, estimated using the initial parameters (OpenFF 1.0.0, blue points) and the 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 3.4.2,3.5.2.
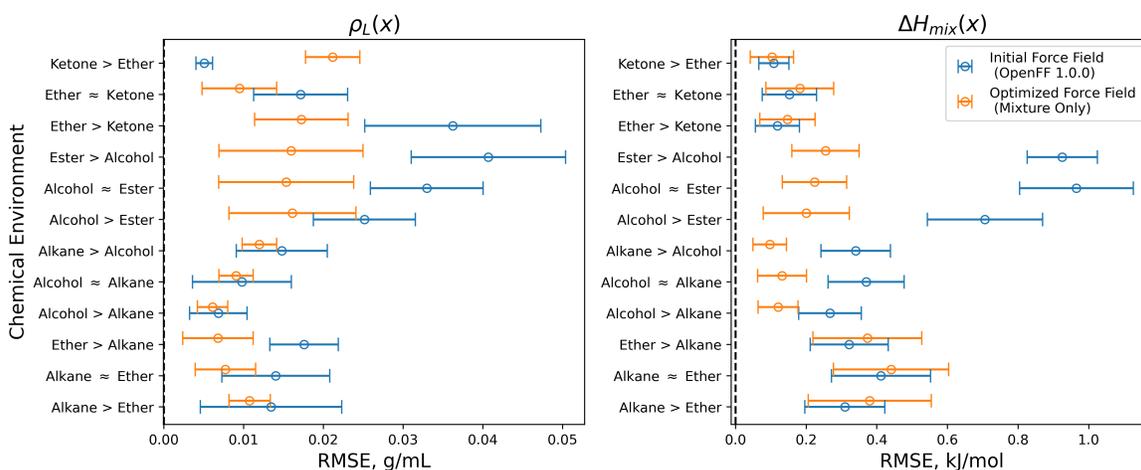
Figure 6. Optimization improves RMSEs of mixture properties for all training sets. Figure shows categorized RMSE vs. experiment of $\rho_l(x)$ (left panel) and $\Delta H_{mix}(x)$ (right panel) measurements in the "mixture only" training set, estimated using the initial parameters (OpenFF 1.0.0, blue 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 3.4.2 and 3.5.2.

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.

Pure Density Gradients by Composition
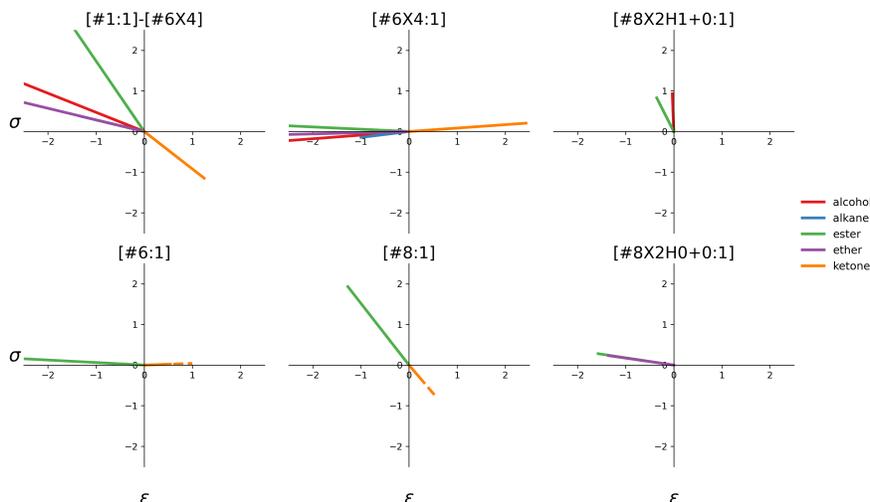
Figure 7. Parameter gradients indicate that ketone measurements drive parameters for hydrogen, carbon, and oxygen in opposite directions as other moieties. The data show the contribution of the first derivatives of the force field parameters to the pure density data portion of the objective function 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.

3.2 Test Set Performance

3.2.1 Overall Results

Each of the retrained force fields and the original OpenFF 1.0.0 force field was assessed against the larger test set. We see that all of the refit force fields improve RMSE against experiment, compared to the base OpenFF 1.0.0 force fields. The Kendall τ and RMSE values for each force field against the test set are shown in Figure 8.
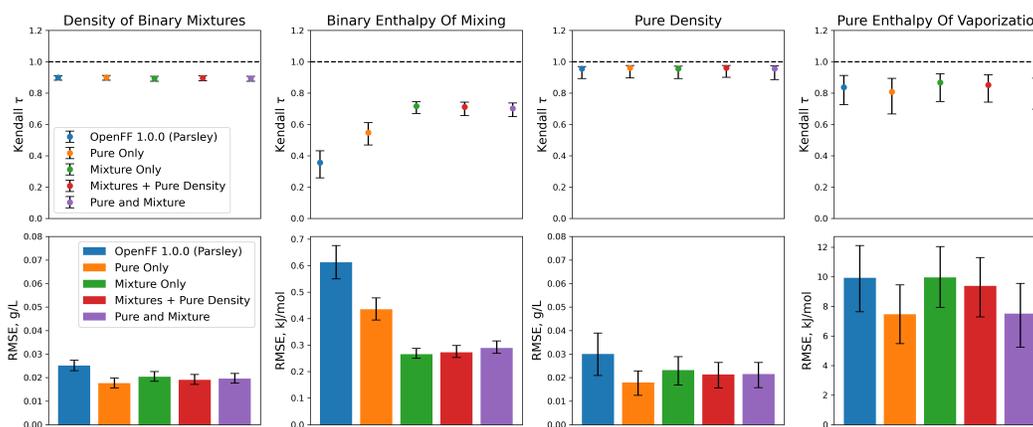
Figure 8. Benchmarking reoptimized force fields against the test set indicates that training against mixture properties yields significantly improved performance compared to the original force field. The figure shows Kendall τ and RMSE for the test set for original and all sets of optimized parameters. Training against data including mixture properties significantly improves enthalpies of mixing without degrading performance in other properties. 95% confidence intervals were computed by bootstrapping with replacement for 1000 iterations.

Force fields reoptimized on all the different sets of observable data have statistically similar performance on ρ_l and $\rho_l(x)$, with highly accurate predictions of densities. Binary mixture densities in particular show significant improvement over the OpenFF 1.0.0 force field, with all refitted force fields achieving an average RMSE <=0.20 g/mL, compared to 0.25 g/mL for OpenFF 1.0.0. Force fields that include mixture data ("mixture only" = green, "mixtures + pure density" = red, "pure and mixture" = purple) performed best on $\Delta H_{mix}(x)$, significantly outperforming the "pure only" (orange) force field. For ΔH_{vap}, the most improved force field was the one trained using the "pure and mixture" training set, which sees similar improvement in most properties compared to the other retrained force fields, but a notable improvement in ΔH_{vap}. However, the utility of improved ΔH_{vap} predictions is questionable for a force field intended to be used for biomolecular systems. We also note that pure and mixture densities are already highly accurate (RMSE $\leq$ 0.03 g/mL) when predicted with the OpenFF 1.0.0 force field, so improvements in enthalpic properties are likely more meaningful given that densities are already well predicted.

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 $\Delta H_{mix}(x)$ against the baseline; however, training directly against the "mixture only" set yields a much larger improvement. Training using mixture properties not only improved test performance against mixture properties, but also performs similarly to training against pure data on ΔH_{vap}. The correlation vs. experimental τ is improved slightly when training against the "mixture only" set, with an τ of 0.85 (95% CI 0.0.74,0.92), vs. the "pure only" set, with an τ of 0.80 (95% CI 0.67,0.89). While the ΔH_{vap} RMSE is higher when trained against the "mixture only" (9.95 kJ/mol, 95% CI (7.92,12.04)) set vs. the "pure only" training set (7.47 kJ/mol, 95% CI (5.47,9.45)); training against the "pure and mixture" set (containing both the pure and mixture data)) performs similarly on RMSE (7.51 kJ/mol, 95% CI (5.24,9.54) for ΔH_{vap}, while maintaining the improvement on $\Delta H_{mix}(x)$. 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.

3.2.2 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 is evident in the parity plots for $\Delta H_{mix}(x)$, where a systematic underprediction of alcohol/ester (green points) and alcohol/ketone (orange points) mixture enthalpies is corrected (Figure 9).

Binary Enthalpy Of Mixing (kJ/mol)
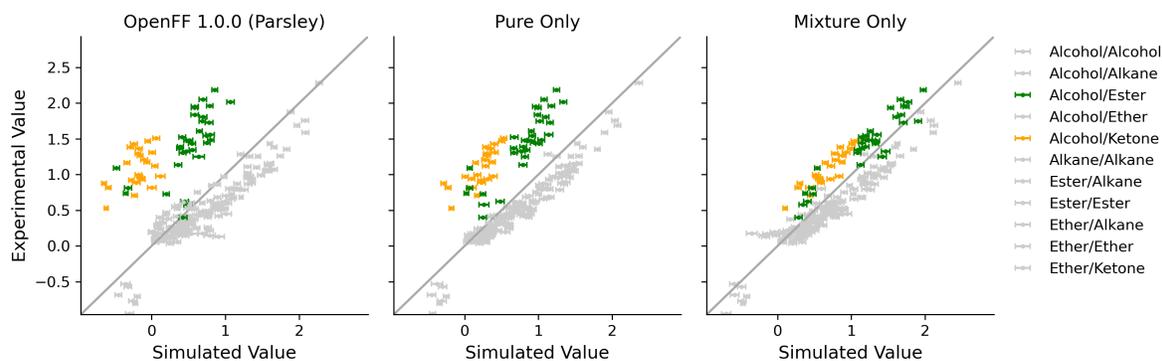
Figure 9. Training against measurements of liquid mixtures corrects systematic error in alcohol/ester and alcohol/ether enthalpies of mixing. This figure shows a comparison of the estimated and experimentally measured $\Delta H_{mix}(x)$ data points for the test set, plotting for force fields optimized 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.

This is particularly significant as alcohol/ester and alcohol/ketone mixture enthalpies have strong positive deviations from ideality. 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 is where mixture properties, and especially their ability to more readily capture complementary interactions, appears to be advantageous over pure properties.

4 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 adjustment of parameter sets, new force fields were produced that all exceeded the performance of the initial force field on the broader 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.

We have shown that 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. We also note that parameterizing against mixture properties alongside pure properties produces parameter sets with improved $\Delta H_{mix}(x)$ *and* ΔH_{vap}. Given that we control for the identity of the molecules in the training set, this demonstrates that mixture properties contains information about LJ interactions that pure component property measurements do not.

While the 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. The same interactions captured in enthalpies of mixing should also be informative for properties of pharmaceutical/biomolecular

interest, such as binding affinities, solvation free energies, and other derivative properties. For this reason, optimization of LJ parameters against mixture property targets is planned to be the standard going forward for our general 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.

5 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

6 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.

7 Disclosures

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.

References

[1] **Bottaro S**, Lindorff-Larsen K. Biophysical Experiments and Biomolecular Simulations: A Perfect Match? Science. 2018 Jul; 361(6400):355–360. doi: 10.1126/science.aat4010.

[2] **Slater JC**, Kirkwood JG. The Van Der Waals Forces in Gases. Physical Review. 1931 Mar; 37(6):682–697. doi: 10.1103/PhysRev.37.682.

[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(2):187–217. doi: 10.1002/jcc.540040211.

[4] **Van Gunsteren WF**, Karplus M. Effect of Constraints on the Dynamics of Macromolecules. Macromolecules. 1982 Nov; 15(6):1528–1544. doi: 10.1021/ma00234a015.

[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 Nov; 118(45):11225–11236. doi: 10.1021/ja9621760.

[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(4):671–690. doi: 10.1002/jcc.21367.

[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(9):1157–1174. doi: 10.1002/jcc.20035.

[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(13):1656–1676. doi: 10.1002/jcc.20090.

[9] **Jorgensen WL**, Madura JD, Swenson CJ. Optimized Intermolecular Potential Functions for Liquid Hydrocarbons. Journal of the American Chemical Society. 1984 Oct; 106(22):6638–6646. doi: 10.1021/ja00334a030.

[10] Monticelli L, Salonen E, editors. Biomolecular Simulations: Methods and Protocols. Methods in Molecular Biology, Humana Press; 2013. https://www.springer.com/us/book/9781627030168.

[11] **Yin D**, MacKerell AD. Combined Ab Initio/Empirical Approach for Optimization of Lennard–Jones Parameters. Journal of Computational Chemistry. 1998; 19(3):334–348. doi: 10.1002/(SICI)1096-987X(199802)19:3<334::AID-JCC7>3.0.CO;2-U.

[12] **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(2):199–213. doi: 10.1002/jcc.1166.

[13] **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 May; 12(5):2312–2323. doi: 10.1021/acs.jctc.6b00027.

[14] **Kantonen SM**, Muddana HS, Schauperl M, Henriksen NM, Wang LP, Gilson MK. Data-Driven Mapping of Gas-Phase Quantum Calculations to General Force Field Lennard-Jones Parameters. Journal of Chemical Theory and Computation. 2020 Feb; 16(2):1115–1127. doi: 10.1021/acs.jctc.9b00713.

[15] **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 Dec; 13(12):6146–6157. doi: 10.1021/acs.jctc.7b00522.

[16] **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 Feb; 33(2):133–203. doi: 10.1007/s10822-018-0111-4.

[17] **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 Feb; 33(2):205–264. doi: 10.1007/s10822-018-0134-x.

[18] **Berendsen HJC**, Grigera JR, Straatsma TP. The Missing Term in Effective Pair Potentials. The Journal of Physical Chemistry. 1987 Nov; 91(24):6269–6271. doi: 10.1021/j100308a038.

[19] **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 Jul; 114(26):8621–8630. doi: 10.1021/jp911699p.

[20] **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 Jul; 114(26):8631–8645. doi: 10.1021/jp911701h.

[21] **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 Mar; 28(3):277–287. doi: 10.1007/s10822-014-9714-6.

[22] **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 Feb; 117(8):2328–2338. doi: 10.1021/jp311851r.

[23] **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 Mar; 25(3):593–596. doi: 10.1016/0584-8539(69)80009-7.

[24] **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. doi: 10.3891/acta.chem.scand.24-2612.

[25] **Frenkel M**, Chiroco RD, Diky V, Dong Q, Marsh KN, Dymond JH, Wakeham WA, Stein SE, Königsberger E, Goodwin ARH. XML-Based IUPAC Standard for Experimental, Predicted, and Critically Evaluated Thermodynamic Property Data Storage and Capture (ThermoML) (IUPAC Recommendations 2006). Pure and Applied Chemistry. 2006 Jan; 78(3):541–612. doi: 10.1351/pac200678030541.

[26] **Majer V**, Svoboda V, Kehiahan H. Enthalpies of Vaporization of Organic Compounds: A Critical Review and Data Compilation, vol. 32. Blackwell Scientific Oxford; 1985.

[27] **Pontolillo J**, Eganhouse RP. 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. U.S. Geological Survey; 2001.

[28] **Fischer J**, Möller D, Chialvo A, Haile JM. The Influence of Unlike Molecule Interaction Parameters on Liquid Mixture Excess Properties. Fluid Phase Equilibria. 1989 Sep; 48:161–176. doi: 10.1016/0378-3812(89)80199-2.

[29] **Dai J**, Li X, Zhao L, Sun H. Enthalpies of Mixing Predicted Using Molecular Dynamics Simulations and OPLS Force Field. Fluid Phase Equilibria. 2010 Mar; 289(2):156–165. doi: 10.1016/j.fluid.2009.11.028.

[30] **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 Dec; 13(12):6201–6212. doi: 10.1021/acs.jctc.7b00800.

[31] **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 LP. Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. Journal of Chemical Theory and Computation. 2021 Sep; doi: 10.1021/acs.jctc.1c00571.

[32] **Wang LP**, Martinez TJ, Pande VS. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. The Journal of Physical Chemistry Letters. 2014 Jun; 5(11):1885–1891. doi: 10.1021/jz500737m.

[33] **Boothroyd S**, Wang LP, 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 Aug; doi: 10.33774/chemrxiv-2021-jz3zq.

[34] **Qiu Y**, Smith DGA, 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 LP, Openforcefield/Openforcefields: Version 1.0.0 "Parsley"; 2019. doi: 10.5281/zenodo.3483227. Zenodo.

[35] **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(4):287–292. doi: 10.6028/jres.071A.035.

[36] **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 Apr; 33(4):433–440. doi: 10.1006/jcht.2000.0746.

[37] **Giner B**, Villares A, Martín S, Lafuente C, Royo FM. Isothermal Vapour–Liquid Equilibrium for Cyclic Ethers with 1-Chloropentane. Fluid Phase Equilibria. 2007 Jan; 251(1):8–16. doi: 10.1016/j.fluid.2006.10.024.

[38] **Alcalde R**, Aparicio S, Dávila MJ, García B, Leal JM. Liquid–Liquid Equilibria of Lactam Containing Binary Systems. Fluid Phase Equilibria. 2008 Apr; 266(1):90–100. doi: 10.1016/j.fluid.2008.02.008.

[39] **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 Oct; 271(1):76–81. doi: 10.1016/j.fluid.2008.07.001.

[40] **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 Apr; 291(1):8–12. doi: 10.1016/j.fluid.2009.11.030.

[41] **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 Jan; 361:23–29. doi: 10.1016/j.fluid.2013.10.034.

[42] **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 Jan; 385:92–104. doi: 10.1016/j.fluid.2014.11.005.

[43] **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 Jun; 395:1–8. doi: 10.1016/j.fluid.2015.03.023.

[44] **Requejo PF**, Calvar N, Domínguez 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 Jun; 417:137–143. doi: 10.1016/j.fluid.2016.02.028.

[45] **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 Jan; 39(1):128–141. doi: 10.1016/j.jct.2006.05.004.

[46] **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 Nov; 39(11):1452–1457. doi: 10.1016/j.jct.2007.03.008.

[47] **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 Mar; 40(3):442–459. doi: 10.1016/j.jct.2007.09.007.

[48] **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 May; 41(5):591–597. doi: 10.1016/j.jct.2009.01.008.

[49] **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 Sep; 41(9):1025–1029. doi: 10.1016/j.jct.2009.04.006.

[50] **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 Oct; 53:52–59. doi: 10.1016/j.jct.2012.04.011.

[51] **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 Jul; 98:21–32. doi: 10.1016/j.jct.2016.02.016.

[52] **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 Sep; 543:66–73. doi: 10.1016/j.tca.2012.05.004.

[53] **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 Feb; 62(2):623–632. doi: 10.1021/acs.jced.6b00606.

[54] **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 Sep; 62(9):2944–2952. doi: 10.1021/acs.jced.7b00185.

[55] **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 Jan; 48(1):195–197. doi: 10.1021/je0256028.

[56] **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 Jul; 49(4):1116–1126. doi: 10.1021/je034022n.

[57] **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 May; 51(3):1031–1034. doi: 10.1021/je050514j.

[58] **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 Oct; 56(10):3949–3954. doi: 10.1021/je200350a.

[59] **Dohnal V**, Řehák K. Thermal and Volumetric Properties of Four Aqueous Aroma Compounds at Infinite Dilution. Journal of Chemical & Engineering Data. 2012 Jun; 57(6):1822–1828. doi: 10.1021/je300280s.

[60] **Zorębski 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 Feb; 53(2):591–595. doi: 10.1021/je700580g.

[61] **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 May; 54(5):1575–1579. doi: 10.1021/je800915d.

[62] **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):1–6. doi: 10.1135/cccc19760001.

[63] **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 Apr; 12(4):387–391. doi: 10.1016/0021-9614(80)90152-4.

[64] **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(3):637–651. doi: 10.1135/cccc19790637.

[65] **Snelson A**, Skinner HA. Heats of Combustion: Sec-Propanol, 1,4-Dioxan, 1,3-Dioxan and Tetrahydropyran. Transactions of the Faraday Society. 1961 Jan; 57(0):2125–2131. doi: 10.1039/TF9615702125.

[66] **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(12):3233–3240. doi: 10.1135/cccc19803233.

[67] **Majer V**, Svoboda V, Uchytilová V, Finke M. Enthalpies of Vaporization of Aliphatic C5 and C6 Alcohols. Fluid Phase Equilibria. 1985 Jan; 20:111–118. doi: 10.1016/0378-3812(85)90026-3.

[68] **Uchytilová V**, Majer V, Svoboda V, Hynek V. Enthalpies of Vaporization and Cohesive Energies for Seven Aliphatic Ketones. The Journal of Chemical Thermodynamics. 1983 Sep; 15(9):853–858. doi: 10.1016/0021-9614(83)90091-5.

[69] **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 Jan; (5):565–569. doi: 10.1039/P29820000565.

[70] **Wolf G**. Thermochemische Untersuchungen an Cyclischen Ketonen. Helvetica Chimica Acta. 1972; 55(5):1446–1459. doi: 10.1002/hlca.19720550510.

[71] **Wadsö I**, Murto ML, 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. doi: 10.3891/acta.chem.scand.20-0536.

[72] **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 Apr; 56(4):800–810. doi: 10.1021/je100231g.

[73] **Boothroyd S**, Madin O, Wagner J, Setiadi J, Thompson M, Rodríguez-Guerra J, Openforcefield/Openff-Evaluator: 0.1.0 OpenFF Evaluator; 2020. doi: 10.5281/zenodo.3839928. Zenodo.

[74] Common Workflows — OpenFF Evaluator Documentation;. https://openff-evaluator.readthedocs.io/en/stable/properties/commonworkflows.html#simulation-layer.

[75] **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(13):2157–2164. doi: 10.1002/jcc.21224.

[76] **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; 2019. doi: 10.5281/zenodo.3541134. Zenodo.

[77] **Leimkuhler B**, Matthews C. Rational Construction of Stochastic Numerical Methods for Molecular Sampling. Applied Mathematics Research eXpress. 2013 Jan; 2013(1):34–56. doi: 10.1093/amrx/abs010.

[78] **Chodera JD**. A Simple Method for Automated Equilibration Detection in Molecular Simulations. Journal of Chemical Theory and Computation. 2016 Apr; 12(4):1799–1805. doi: 10.1021/acs.jctc.5b00784.

[79] **Shirts MR**, Chodera JD. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. The Journal of Chemical Physics. 2008 Sep; 129(12):124105. doi: 10.1063/1.2978177.

[80] **Levenberg K**. A Method for the Solution of Certain Non-Linear Problems in Least Squares. Quarterly of Applied Mathematics. 1944; 2(2):164–168. doi: 10.1090/qam/10666.

[81] **Moré JJ**, Sorensen DC. Computing a Trust Region Step. SIAM Journal on Scientific and Statistical Computing. 1983 Sep; 4(3):553–572. doi: 10.1137/0904038.

[82] **Dennis JE**, Gay DM, Walsh RE. An Adaptive Nonlinear Least-Squares Algorithm. ACM Transactions on Mathematical Software. 1981 Sep; 7(3):348–368. doi: 10.1145/355958.355965.

[83] **Wang LP**, McKiernan KA, Gomes J, Beauchamp KA, Head-Gordon T, Rice JE, Swope WC, Martínez 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 Apr; 121(16):4023–4039. doi: 10.1021/acs.jpcb.7b02320.

[84] **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 May; 117(19):5179–5197. doi: 10.1021/ja00124a002.

[85] **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 Jul; p. 286542. doi: 10.1101/286542.

[86] **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 Nov; 14(11):6076–6092. doi: 10.1021/acs.jctc.8b00640.

[87] **Kendall MG**. A New Measure of Rank Correlation. Biometrika. 1938 Jun; 30(1-2):81–93. doi: 10.1093/biomet/30.1-2.81.