Non-bonded force field parameters from MBIS partitioning of the molecular electron density improve thermophysical properties prediction of organic liquids

Jorge Pulido,[†,‡] Luis Macaya,[†,‡] and Esteban Vöhringer-Martinez[*,†]

†*Departamento de Físico-Química, Facultad de Ciencias Químicas, Universidad de Concepción, 4070386 Concepción, Chile*

‡*Contributed equally to this work*

E-mail: evohringer@udec.cl

Phone: +56 41 2204986

Abstract

The accuracy of predicting thermophysical properties through molecular dynamics simulations is constrained by the precision of models used to describe molecular interactions. The Open Force Field initiative has established a computational structure to develop new models and introduced two non-polarizable force fields, Parsley and Sage. Sage version 2.0.0 focused on refining Lennard-Jones parameters to reflect thermophysical properties accurately. In this context, we evaluate the ability of our introduced D-MBIS non-bonded force field parameters to replicate liquid densities and enthalpies of evaporation of 49 neutral compounds from the ThermoML database using the openff-evaluator package. Our findings confirm that our ab-initio derived non-bonded force field parameters accurately mirror both thermophysical properties with a high degree of precision.

Introduction

Molecular dynamics (MD) simulation serves as a powerful tool for investigating the thermodynamic properties of molecules, an aspect crucial to physics, chemistry, and biology.[1] Its wide application in condensed-phase systems stems particularly from a beneficial balance between model accuracy and computational cost. While classical atomistic models only approximate the accurate quantum-mechanical description of molecules, they can still offer a realistic picture of the structure and dynamics of molecular systems, offering spatial and temporal resolutions of 0.1 nm and 1-2 fs respectively. Moreover, their computational costs allow exploring system sizes of more 100 nm[2,3] and time scales of several microseconds.[4]

Nonetheless, the accuracy of MD simulations in replicating thermodynamic properties of a molecular system is reliant on the quality of the models, referred to as force fields.[5,6] Force fields for condensed-phases like CHARMM,[7,8] AMBER,[9,10] OPLS,[11,12] and GROMOS[13] primarily focus on the depiction of liquids and solutions, with solvated biomolecules being a significant special case. They possess a comparatively straightforward functional form combined with the widespread application of parameter combination and transferability assumptions. Typically, they encompass a vast number of atom types that account for atoms in diverse chemical environments. Their parameters were primarily calibrated against experimental thermodynamic and spectroscopic data related to liquids and molecules in solution.[9,11,14] As their main objective is to provide an accurate representation of condensed-phase properties the focus during parameterization is on the correct description of non-bonded interactions.[15]

Five years ago the Open Force Field (OpenFF) initiative started an open and collaborative approach for better force fields. Its first version[16] (Parsley) was created for treating small molecule parameters. It offered substantial coverage of chemical space through a novel direct chemical perception method while maintaining the precision of other small molecule force fields. Instead of atom typing used in previous force fields, direct chemical perception employs parameter assignment based on substructure.[17] This unites intricate chemistry into one physically discernible chemical classification. SMIRKS patterns are utilized to establish

these groups, and the relevant parameters can be applied to any substructure match in any molecule, thereby enhancing its universal applicability. This straightforward chemical perception method significantly cuts down the quantity of empirical force field parameters, enabling swift readjustment of parameters to enhance chemical precision.[17] The earliest versions of Parsley largely adopted the initial non-bonded parameters from SMIRNOFF99Frosst,[18] which is an AMBER compatible small molecule force field. The following Parsley versions experienced significant enhancement of the valence parameters by aligning them with a broad dataset of quantum chemical calculations.[19]

The OpenFF Sage 2.0.0 small molecule force field[19] was recently launched, building on the foundation of the Open Force Field Parsley generation of force fields. This was the first occasion in which the Lennard-Jones (LJ) parameters were recalibrated to mirror thermophysical attributes. This became possible through openff-evaluator,[20] a computational package developed to predict thermophysical properties and train LJ parameters. In Sage 2.0.0, these parameters were adjusted to optimize the thermophysical properties of the condensed phase, including the mixing enthalpy and densities of pure or binary mixtures.[14,19] Sage, such as Parsley, describes molecules that encompass the chemical space of elements such as C, H, O, N, P, S, F, Cl, Br, and I. All OpenFF variants depend on the AM1-BCC model[21,22] for the assignment of partial charges. This method is extensively employed for organic compounds.

The refined LJ parameters in Sage or the AM1-BCC atomic charges, from a physical chemistry viewpoint, should accurately represent the four kinds of intermolecular interactions: electrostatic, induction, dispersion, and exchange repulsion. Our research lab has proposed a novel method for acquiring atomic properties from the polarized electron density in varied configurations of condensed phases.[23–25] The D-MBIS method uses the Minimal Basis Iterative Stockholder (MBIS) method[26,27] to partition the molecular electron density of various conformations in atomic contributions. We have shown that partial charges derived from the average electron density of an atom in a molecule successfully decreased the

errors in predicting hydration free energies[23] or binding affinity in host-guest complexes.[24,25]

Lately, along with the atomic charges, we have also developed new force field parameters for the dispersion and exchange-repulsion interaction.[24] Earlier methods employ atoms-in-molecules partition of the electron density for a molecule's single conformation in vacuum.[28] Consequently, these methods overlook the changes in molecular structure in condensed phases and the impact of solvent molecules on the spread of electron density. Our latest computational framework aids in extracting non-bonded force field parameters from the polarized electron density of molecules.[24] Using Federov *et al.*'s[29] and Tkatchenko's *et al.*'s[30,31] methodologies, we calculate the van-der-Waals radii and C_6 dispersion coefficients of the LJ potential respectively.

Here, we combine our D-MBIS non-bonded force field parameters with the optimized valence parameters in OpenFF 2.0.0 and test their ability to reproduce liquid densities and enthalpies of evaporation of pure neutral substances under standard conditions.

Methods

Filtering the ThermoML Database

In order to assess the quality of the new parameters, physical properties of diverse small organic molecules were predicted and compared with predictions of other force fields and with experimental data.

The small organic molecules were extracted from ThermoML database.[32,33] This database was filtered for pure, liquid, neutral molecules which had both their densities or heats of evaporation reported at 298 K and 1 atm with the openff-evaluator[20] functionality (see notebook in the SI). Thereafter their Lennard-Jones parameters and atomic charges were derived using the D-MBIS partitioning method[24,25] described below.

Non-bonded force field parameters from the polarized electron density

To derive non-bonded force field parameters for each evaluated molecule, we improved our workflow to obtain D-MBIS atomic charges[25] and Lennard-Jones (LJ) parameters[24] including support to OpenFF force field. This workflow combines molecular dynamics simulations carried out with OpenMM[34] and the QM/MM methodology applied to several configurations of the trajectory to obtain polarized electron densities. With these densities, we applied the Minimal Basis Iterative Stockholder (MBIS)[26] partitioning method to calculate atomic charges and LJ parameters and update these in an iterative manner.

The main difference compared to our previous workflow[24,25] involves atomic charges and LJ parameters replacement in every parameter's update iteration. Respect to software, we add the OpenFF's ecosystem for small molecule force field treatment, we change to ORCA 5.0.3 version[35] that include support to OpenMM combined with correct treatment of large PDB files and RDKit to treat chemically equivalent atoms (e.g. hydrogen atoms in methyl groups) for averaged parameter's assignment.

For this work, we rely on OpenFF 2.0.0 force field for bonded parameters. AM1-BCC atomic charges and OpenFF LJ parameters (ϵ, $r_{min}/2$) were replaced with the derived non-bonded parameters. We maintain the same workflow's parameters and convergence criteria from our previous studies.[24,25] Initially, a SMILES string for the analyzed molecule and a system's PDB file were required for the non-bonded parameter's derivation workflow. This PDB file was obtained from openff-evaluator (see below). As output, an OpenFF XML file and a CSV file (.dat) were generated with the new LJ parameters and atomic charges. These files were used to modify OpenFF 2.0.0 LJ parameters and AM1-BCC charges respectively.

Calculation of liquid densities and heats of evaporation

To obtain the liquid densities and heats of evaporation we used the openff-evaluator package. The Python scripts are provided in the Supporting Information.

The standard protocol for the condensed phase simulations generates a cubic box with 1000 molecules with an approximated density of $0.95\frac{g}{mL}$ by PACKMOL software,[36] then molecular interaction parameters are assigned according to the specified force field, and energy minimization for the simulation box was applied by means of OpenMM.[34] The system is equilibrated in a NPT ensemble with a Langevin integrator and 100 000 steps, integration time-step of 2 fs and a MonteCarlo barostat with a pressure correction every 25 steps. The production simulation considers 2 ns in a isothermal-isobaric ensemble at 298 K and 1 atm.

Simulations for the gas phase used to obtain the enthalpy of evaporation considers a cubic box without periodic boundary conditions and one molecule as ideal gas, force field parameters are applied and the box is equilibrated in a NVT ensemble, then the production simulation is carried out in another NVT ensemble with 15 000 000 steps and a time-step of 2 fs. Density was calculated by the average of the box's mass and volume:

$$\rho = \langle \frac{M}{V} \rangle \tag{1}$$

Enthalpy of evaporation was calculated assuming that $\bar{V}_{gas} >> \bar{V}_{liquid}$ and ideal gas behavior,

$$\Delta H_{vap} = \Delta\langle E \rangle + P\Delta\langle V \rangle \tag{2}$$

$$\Delta H_{vap} = \langle U_{gas} \rangle - \langle U_{liquid} \rangle + RT \tag{3}$$

where U is the potential energy, R the universal gas constant and T the temperature. To include the polarization effects in the system,[37,38] we add a correction term (i.e E^{pol}) to the enthalpy of evaporation calculated with D-MBIS parameters. This term accounts for the potential energy cost associated with the molecular electron density distortion in a particular

chemical environment (e.g in a homogeneous condensed phase) respect to vacuum.

$$E^{\text{pol}} = \left\langle \Psi | \hat{H}_{vac} | \Psi \right\rangle - \left\langle \Psi_{vac} | \hat{H}_{vac} | \Psi_{vac} \right\rangle \tag{4}$$

An ensemble-averaged value for the polarization energy ($\left\langle E^{\text{pol}} \right\rangle$) was obtained in parallel from the D-MBIS non-bonded parameter derivation. Finally, the expression of polarization-corrected enthalpy of vaporization is represented in Eq.5:

$$\Delta H_{vap} = \langle U_{gas} \rangle - \langle U_{liquid} \rangle + RT - \left\langle E^{\text{pol}} \right\rangle \tag{5}$$

Results & Discussion

To validate our derived non-bonded force field parameters we calculated thermophysical properties of diverse small molecules extracted from the ThermoML database using openff-evaluator.[20] The ThermoML database contains more than 100000 thermophysical properties of substances as liquid densities, enthalpy of mixing and enthalpy of evaporation at different thermodynamic states reported in literature.[33] We filtered the database for neutral molecules with a reported liquid density or enthalpy of evaporation at 101.325 kPa and 298 K. Figure 1 shows the chemical structure of the 49 substances colored by the thermophysical property available in ThermoML. 42 substances possess a reported liquid density at 298K and 1 atm and 28 substances an enthalpy of evaporation.

The selected substances evidence various functional groups formed by carbon, nitrogen, oxygen, phosphor, and halogens. Some of them are linear alkanes or have an aromatic core combined with one functional group, although others may possess more than one and quite complex structures (see substance 16). While IUPAC names and SMILES string of the 49 substances are provided for download together with their force field parameters, for compact notation and further analysis we numbered them from 1 to 49 as shown in Figure 1.

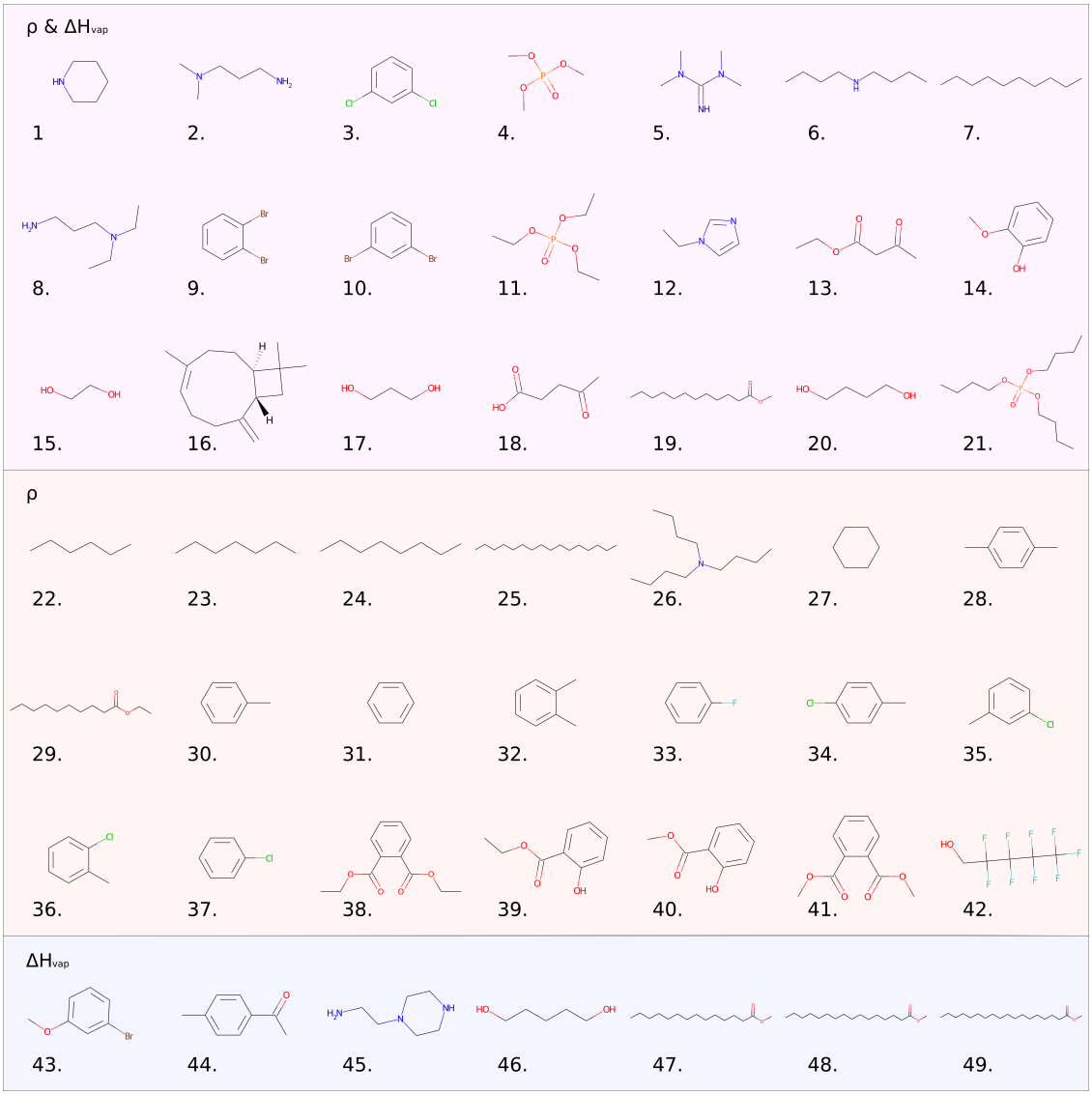

Figure 1: Chemical structure of all substances filtered from the ThermoML database colored with the thermophysical property used to test the performance of the force field parameters.

D-MBIS non-bonded force field parameters predict liquid densities accurately

We chose the Open Force Field 2.0.0 (OpenFF 2.0.0) as a benchmark to evaluate our non-bonded force field parameters. Contrary to prior force fields, OpenFF does not depend on atom names. Instead, it assigns parameters according to the chemical environment of each atom, as depicted in SMIRKS patterns.[17] The same chemical surroundings also modulates the local electron density, from which we obtain the non-bonded force field parameters.

Consequently, we anticipate that atoms, which are assembled in accordance with SMIRKS patterns, would possess similar non-bonded force field parameters. The initial version of OpenFF fine-tuned the bonded parameters that describe covalent interactions, utilizing an extensive database of quantum chemical computations.[16] Thus, OpenFF predominantly depends on electronic structure calculations, similar to our non-bonded force field parameters which we conjectured would enhance compatibility. The OpenFF 2.0.0 version has updated the LJ parameters, which depict the atomic repulsion and dispersion interactions.[19] However, these non-bonded force field parameters were calibrated to reproduce the liquid densities and enthalpy of mixing for substances in the ThermoML database rather than relying on molecular properties. To ensure that microscopic attributes remain the key factors in defining atomic properties, we replaced them with our first-principle LJ parameters. At the same time, we also switched out the AM1-BCC atomic charges for our D-MBIS atomic charges that are directly derived from the polarized molecular electron density. Then, we assessed their ability to replicate the liquid densities of 42 different substances to validate our D-MBIS non-bonded force field parameters.

Figure 2 contrasts liquid densities determined with the original OpenFF 2.0.0 on the left (A) with those resulting from our non-bonded D-MBIS Lennard-Jones and atomic charges (LJ-Q) on the right (B). The densities produced by OpenFF 2.0.0 have a root mean square error (RMSE) of 0.019 g/mL, accompanied by a slightly negative mean error (-0.006 g/mL) for the 42 unique substances. The ability to replicate experimental densities with these parameters is confirmed by a correlation coefficient of $R^2 = 0.996$. Nonetheless, these outcomes have been anticipated given that some of the 42 substances were included in the training data set utilized to optimize LJ parameters in OpenFF version 2.0.0. In Figure 2A, beneath the parity plot, we show the chemical structure of substances that deviate by more than one RMSE from the experimental values. These are marked as reference within the plot. Interestingly, it is worth noting that many of these structures possess more than one functional group or include hetero atoms such as phosphor or halogens, which were not included in the

original OpenFF 2.0.0 training set. Just one alkane, hexadecane (25), substantially diverges from the empirical value. We hypothesized that this variation stems from the small volumes of the periodic simulation boxes used in the simulations. The small volume in relation to the molecule's size hinders the sampling of all rotational conformers of the single carbon-carbon bonds.

Figure 2B displays for comparison densities for the 42 pure substances calculated with our D-MBIS non-bonded force field parameters (Lennard-Jones and atomic charges, LJ-Q). The RMSE in comparison to experimental values rises to 0.045 g/mL, while the average error turns positive (0.022 g/mL). The value of the correlation coefficient R^2 decreases slightly to 0.984. While these statistical descriptors appear to imply a greater variation, it's important to acknowledge that our parameters weren't fine-tuned for thermophysical properties. Instead, they were derived ab-initio from the molecular polarized electron density. Given that their derivation did not employ any thermophysical property, their ability to accurately reproduce experimental liquid densities is remarkably praiseworthy. Compounds with deviations exceeding the RMSE shown at the bottom of Figure 2B also feature numerous functional groups. Cooperative effects or specific halogen or hydrogen bonds may become more prevalent in these compounds, which our non-polarizable force field may not accurately depict. The absence of these interactions might be the reason why the densities of all substances with numerous functional groups are underestimated. Unlike the others, hexadecane (25) exhibits a positive divergence from the empirical data. This pattern has also been noted with the original OpenFF parameters. This validates the hypothesis that the estimation of long linear alkanes' density using small system sizes is limited in reaching the thermodynamic limit due to entropic barriers that hinder sampling conformations with alternating carbon dihedral angles and possibly requires more sampling time or replicates to achieve a better prediction.[39]

To summarize, the D-MBIS non-bonded force field parameters, which are extracted from the molecular density, accurately mirror liquid densities. This is accomplished within the

limitations set by the character of interactions represented in the non-polarizable Open Force Field.

D-MBIS non-bonded force field parameters improve enthalpy of evaporation prediction compared to OpenFF 2.0.0

After examining the densities of various liquids, we evaluated the evaporation enthalpy of neutral substances at a temperature of 298.15 K and an atmospheric pressure of 1.015 atm. We found experimental values for the 28 substances illustrated by the chemical structure in Figure 1. To corroborate the quality of the D-MBIS Lennard Jones parameters and atomic charges, we utilized OpenFF 2.0.0 with AM1-BCC atomic charges as our reference. Figure 3 compares the enthalpies of evaporation derived from both sets of parameters, as illustrated in panels A and B respectively.

In Figure 3A, the parity plot for the OpenFF reference indicates a significant overestimation of the enthalpies of evaporation, as evidenced by a substantial mean error of 10.5 kJ mol^{-1}. With a RMSE of 14.9 kJ mol^{-1} and a correlation coefficient of $R^2 = 0.79$, there's a notable divergence from the experimental results. Molecules exhibiting discrepancies greater than one Root Mean Square Error (RMSE), as shown at the bottom and referred to in the graph, are either phosphoesters, possess additional functional groups, or are esters accompanied by lengthy alkane chains. The latter might encounter similar sampling issues as those noted for hexadecane in liquid densities. The other molecules exhibit pronounced polarizability, indicating a variance in charge distribution between the condensed and gas phases. This varying charge distribution is not accounted for when employing AM1-BCC atomic charges for both corresponding phases.

The D-MBIS non-bonded force field parameters shown in Figure 3B replicate the enthalpy of evaporation more accurately. The parameters for these non-bonded force field are determined by separately deriving atomic charges for each phase, and incorporating a polarization cost when calculating the enthalpy of evaporation (refer to Methods). Their enhanced ability

to replicate this thermophysical property is indicated by a very minimal average error of 0.2 kJ mol^{-1} and an improved R^2 value of 0.86. The Root Mean Square Error (RMSE) of 8.4 kJ mol^{-1} is relatively minor when contrasted with the enthalpy of evaporation, which typically fluctuates between around 50 and 100 kJ mol^{-1}. Interestingly, certain phosphoesters that have previously shown significant deviations also display errors exceeding one RMSE. This could potentially illustrate the boundaries that can be reached with a non-polarizable force field for these specific functional groups. Among other molecules exhibiting significant deviation, we also identified two esters with lengthy alkyl chains that were affected by the previously discussed sampling issue. The remaining ones display chemical structures with numerous functional groups that could potentially favor their aggregation in the gas phase. These gas phase clusters, however, aren't considered in our simulation method which is based solely on a single molecule.

Conclusions

We evaluated the ability of non-bonded force field parameters to accurately replicate standard condition liquid density and enthalpy of evaporation of neutral substances from the ThermoML database. Utilizing the valence parameters of OpenFF 2.0.0 and the openff-evaluator package to predict thermophysical properties, we demonstrate that D-MBIS LJ parameters along with atomic charges accurately emulate liquid densities. Our parameters notably diminish the enthalpy of evaporation error compared to the initial force field.

Acknowledgement

This research was supported by FONDECYT 1200369 and ANID scholarship "Beca Doctorado Nacional" Nº 21210159.

Supporting Information Available

Jupyter-notebooks used for the simulation and analysis of the results as well as the force field parameters can be downloaded from DOI:10.5281/zenodo.10822060.

Data and Software Availability

OpenMM and the openff-evaluator package are open source software and can be used free of charge. The simulation workflow written in Python can be obtained from the following Github repository DOI:https://github.com/QCMM/ffparaim. ORCA 5.0.3 is free for academic use open request. All figures were created with matplotlib.

References

(1) Hollingsworth, S. A.; Dror, R. O. Molecular dynamics simulation for all. *Neuron* **2018**, *99*, 1129–1143.

(2) Jung, J.; Nishima, W.; Daniels, M.; Bascom, G.; Kobayashi, C.; Adedoyin, A.; Wall, M.; Lappala, A.; Phillips, D.; Fischer, W.; others Scaling molecular dynamics beyond 100,000 processor cores for large-scale biophysical simulations. *J. Comp. Chem.* **2019**, *40*, 1919–1930.

(3) Casalino, L.; Seitz, C.; Lederhofer, J.; Tsybovsky, Y.; Wilson, I. A.; Kanekiyo, M.; Amaro, R. E. Breathing and Tilting: Mesoscale simulations illuminate influenza glycoprotein vulnerabilities. *ACS Cent. Sci.* **2022**, *8*, 1646–1663.

(4) Herrera-Nieto, P.; Pérez, A.; De Fabritiis, G. Characterization of partially ordered states in the intrinsically disordered N-terminal domain of p53 using millisecond molecular dynamics simulations. *Sci. Rep.* **2020**, *10*, 12402.

(5) Dauber-Osguthorpe, P.; Hagler, A. T. Biomolecular force fields: where have we been, where are we now, where do we need to go and how do we get there? *J. Comput. Aided Mol. Des.* **2019**, *33*, 133–203.

(6) Riniker, S. Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. *J. Chem. Inf. Model.* **2018**, *58*, 565–578.

(7) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73.

(8) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; others CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comp. Chem.* **2010**, *31*, 671–690.

(9) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(10) Tian, C.; Kasavajhala, K.; Belfon, K. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; others ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **2019**, *16*, 528–552.

(11) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc* **1996**, *118*, 11225–11236.

(12) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; others OPLS4: Improving force field accuracy on challenging regimes of chemical space. *J. Chem. Theory Comput.* **2021**, *17*, 4291–4300.

(13) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; Van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856.

(14) Boothroyd, S.; Madin, O. C.; Mobley, D. L.; Wang, L.-P.; Chodera, J. D.; Shirts, M. R. Improving Force Field Accuracy by Training against Condensed-Phase Mixture Properties. *J. Chem. Theory Comput.* **2022**, *18*, 3577–3592.

(15) Messerly, R. A.; Razavi, S. M.; Shirts, M. R. Configuration-sampling-based surrogate models for rapid parameterization of non-bonded interactions. *J. Chem. Theory Comput.* **2018**, *14*, 3144–3162.

(16) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. *J. Chem. Theory Comput.* **2021**, *17*, 6262–6280.

(17) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Escaping Atom Types in Force Fields Using Direct Chemical Perception. *J. Chem. Theory Comput.* **2018**, *14*, 6076–6092.

(18) Mobley, D.; Bannan, C.; Wagner, J.; Rizzi, A.; Lim, N.; Henry, M. openforcefield/smirnoff99Frosst: Version 1.1. 0. 2019. *Publisher Full Text*

(19) Boothroyd, S.; Behara, P. K.; Madin, O. C.; Hahn, D. F.; Jang, H.; Gapsys, V.; Wagner, J. R.; Horton, J. T.; Dotson, D. L.; Thompson, M. W.; Maat, J.; Gokey, T.; Wang, L.-P.; Cole, D. J.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Shirts, M. R.; Mobley, D. L. Development and Benchmarking of Open Force Field 2.0.0: The Sage Small Molecule Force Field. *J. Chem. Theory Comput.* **2023**, *19*, 3251–3275.

(20) Boothroyd, S.; Wang, L.-P.; Mobley, D. L.; Chodera, J. D.; Shirts, M. R. Open Force Field Evaluator: An Automated, Efficient, and Scalable Framework for the Estimation of Physical Properties from Molecular Simulation. *J. Chem. Theory Comput.* **2022**, *18*, 3566–3576.

(21) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.

(22) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(23) Riquelme, M.; Lara, A.; Mobley, D. L.; Matamala, A. R.; Vöhringer-Martinez, E. Hydration Free Energies in the FreeSolv Database Calculated with Polarized Iterative Hirshfeld Charges. *J. Chem. Inf. Model.* **2018**, *58*, 1779–1797.

(24) González, D.; Macaya, L.; Castillo-Orellana, C.; Verstraelen, T.; Vogt-Geisse, S.; Vohringer-Martinez, E. Nonbonded Force Field Parameters from Minimal Basis Iterative Stockholder Partitioning of the Molecular Electron Density Improve CB7 Host–Guest Affinity Predictions. *J. Chem. Inf. Model.* **2022**, *62*, 4162–4174.

(25) González, D.; Macaya, L.; Vöhringer-Martinez, E. Molecular Environment-Specific Atomic Charges Improve Binding Affinity Predictions of SAMPL5 Host–Guest Systems. *J. Chem. Inf. Model.* **2021**, *61*, 4462–4474.

(26) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development. *J. Chem. Theory Comput.* **2016**, *12*, 3894–3912.

(27) Heidar-Zadeh, F.; Ayers, P. W.; Verstraelen, T.; Vinogradov, I.; Vöhringer-Martinez, E.; Bultinck, P. Information-Theoretic Approaches to Atoms-in-Molecules: Hirshfeld Family of Partitioning Schemes. *J. Phys. Chem. A* **2018**, *122*, 4219–4245.

(28) Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning. *J. Chem. Theory Comput.* **2016**, *12*, 2312–2323.

(29) Fedorov, D. V.; Sadhukhan, M.; Stöhr, M.; Tkatchenko, A. Quantum-Mechanical Relation between Atomic Dipole Polarizability and the van der Waals Radius. *Phys. Rev. Lett.* **2018**, *121*.

(30) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*.

(31) Tkatchenko, A.; Fedorov, D. V.; Gori, M. Fine-Structure Constant Connects Electronic Polarizability and Geometric van-der-Waals Radius of Atoms. *J. Phys. Chem. Lett.* **2021**, *12*, 9488–9492.

(32) Riccardi, D.; Bazyleva, A.; Paulechka, E.; Diky, V.; Magee, J. W.; Kazakov, A. F.; Townsend, S. A.; Muzny, C. D. ThermoML/Data Archive. https://data.nist.gov/od/id/mds2-2422.

(33) Riccardi, D.; Trautt, Z.; Bazyleva, A.; Paulechka, E.; Diky, V.; Magee, J. W.; Kazakov, A. F.; Townsend, S. A.; Muzny, C. D. Towards improved FAIRness of the ThermoML Archive. *J. Comput. Chem.* **2022**, *43*, 879–887.

(34) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.

(35) Neese, F. Software update: The ORCA program system—Version 5.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2022**, *12*, e1606.

(36) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.

(37) Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *J. Phys.: Condens. Matter* **2009**, *21*, 333102.

(38) Swope, W. C.; Horn, H. W.; Rice, J. E. Accounting for Polarization Cost When Using Fixed Charge Force Fields. I. Method for Computing Energy. *J. Phys. Chem. B* **2010**, *114*, 8621–8630.

(39) Gapsys, V.; de Groot, B. L. On the importance of statistics in molecular simulations for thermodynamics, kinetics and simulation box size. *Elife* **2020**, *9*, e57589.

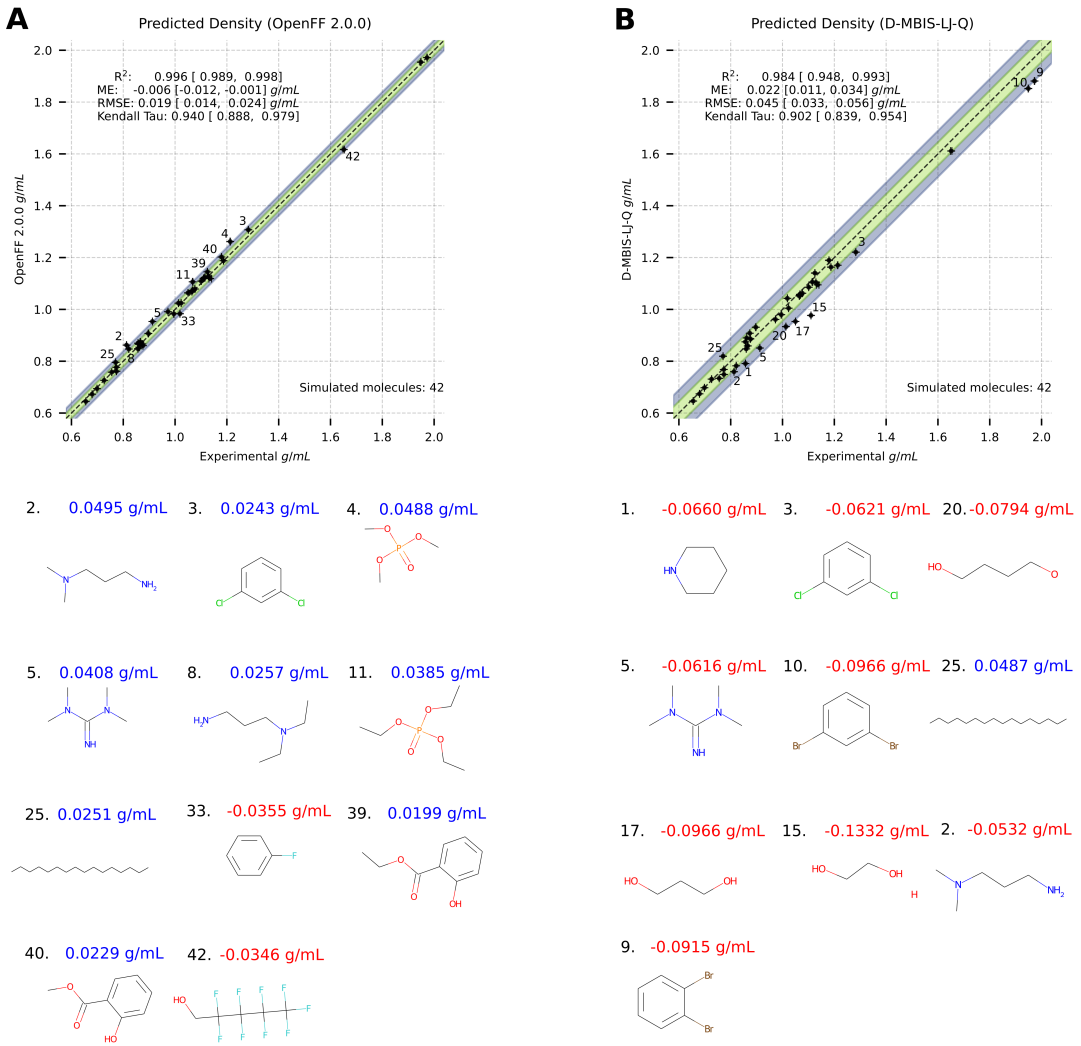

Figure 2: Parity plot between the calculated and experimental densities from 42 substances represented by their chemical structure in Figure 1. Densities obtained from simulations using the original OpenFF 2.0.0 with AM1-BCC atomic charges in **A** are compared to the results obtained with D-MBIS atomic charges and Lennard-Jones parameters in **B**. Statistical descriptors are shown as insets and substances with deviations larger than the root mean square error (RMSE) are labeled in the plot and their chemical structure displayed below.

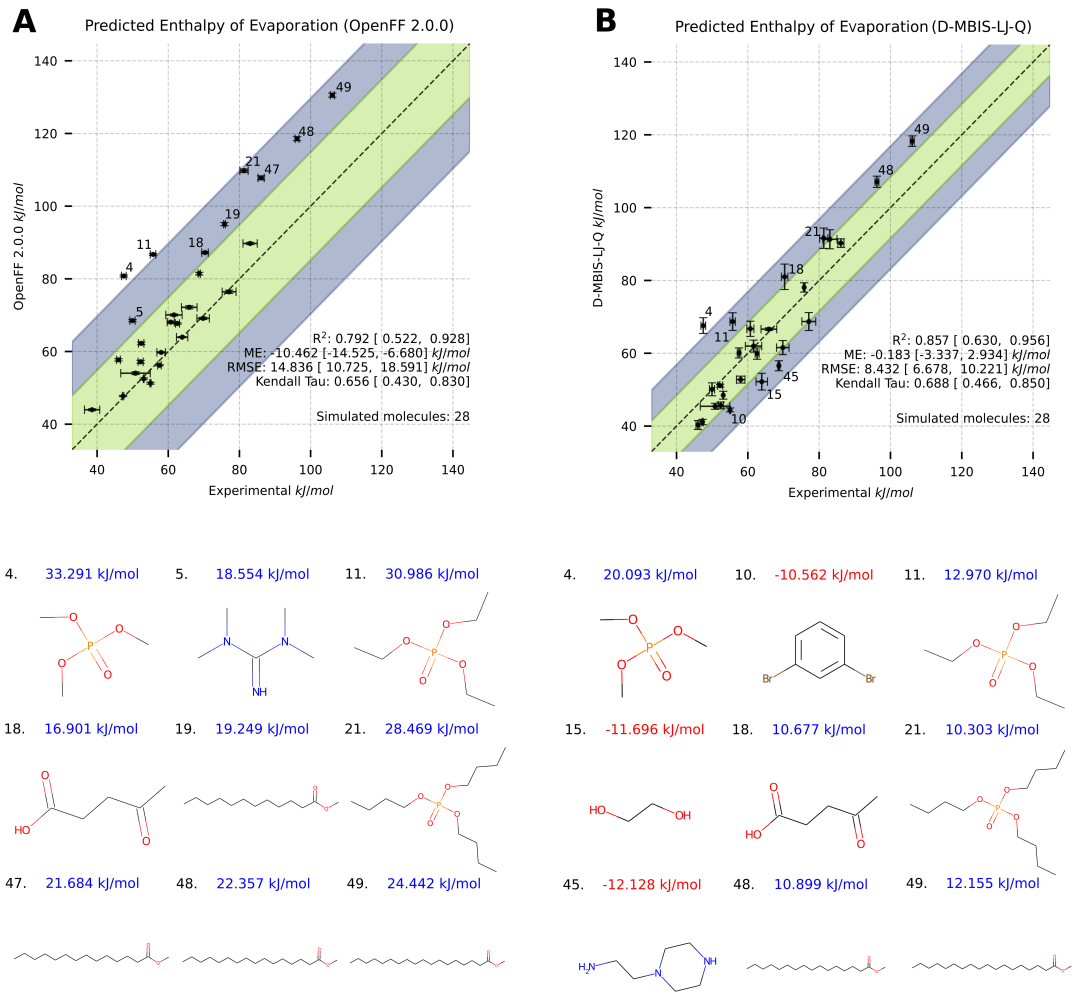

Figure 3: Parity plot between the calculated and experimental enthalpy of evaporation from 28 substances represented by their chemical structure in Figure 1. Enthalpy of evaporation at 298 K obtained from simulations using the original OpenFF 2.0.0 with AM1-BCC atomic charges in **A** are compared to the results obtained with D-MBIS atomic charges and Lennard-Jones parameters in **B**. Statistical descriptors are shown as insets and substances with deviations larger than the root mean square error (RMSE) are labeled in the plot and their chemical structure displayed below.