

# UC Irvine

## UC Irvine Previously Published Works

### Title

Benchmark assessment of molecular geometries and energies from small molecule force fields

### Permalink

<https://escholarship.org/uc/item/8p3657zh>

### Authors

Lim, Victoria T

Hahn, David F

Tresadern, Gary

et al.

### Publication Date

2020

### DOI

10.12688/f1000research.27141.1

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



## RESEARCH ARTICLE

# Benchmark assessment of molecular geometries and energies from small molecule force fields [version 1; peer review: 2 approved]

Victoria T. Lim <sup>1</sup>, David F. Hahn <sup>2</sup>, Gary Tresadern <sup>2</sup>, Christopher I. Bayly<sup>3</sup>, David L. Mobley <sup>1,4</sup>

<sup>1</sup>Department of Chemistry, University of California, Irvine, CA, 92697, USA

<sup>2</sup>Computational Chemistry, Janssen Research & Development, Beerse, B-2340, Belgium

<sup>3</sup>OpenEye Scientific Software, Santa Fe, NM, 87507, USA

<sup>4</sup>Department of Pharmaceutical Sciences, University of California, Irvine, CA, 92697, USA

**V1** First published: 03 Dec 2020, 9(CHEM INF SCI):1390  
<https://doi.org/10.12688/f1000research.27141.1>

Latest published: 03 Dec 2020, 9(CHEM INF SCI):1390  
<https://doi.org/10.12688/f1000research.27141.1>

## Abstract

**Background:** Force fields are used in a wide variety of contexts for classical molecular simulation, including studies on protein-ligand binding, membrane permeation, and thermophysical property prediction. The quality of these studies relies on the quality of the force fields used to represent the systems.

**Methods:** Focusing on small molecules of fewer than 50 heavy atoms, our aim in this work is to compare nine force fields: GAFF, GAFF2, MMFF94, MMFF94S, OPLS3e, SMIRNOFF99Frosst, and the Open Force Field Parsley, versions 1.0, 1.1, and 1.2. On a dataset comprising 22,675 molecular structures of 3,271 molecules, we analyzed force field-optimized geometries and conformer energies compared to reference quantum mechanical (QM) data.

**Results:** We show that while OPLS3e performs best, the latest Open Force Field Parsley release is approaching a comparable level of accuracy in reproducing QM geometries and energetics for this set of molecules. Meanwhile, the performance of established force fields such as MMFF94S and GAFF2 is generally somewhat worse. We also find that the series of recent Open Force Field versions provide significant increases in accuracy.

**Conclusions:** This study provides an extensive test of the performance of different molecular mechanics force fields on a diverse molecule set, and highlights two (OPLS3e and OpenFF 1.2) that perform better than the others tested on the present comparison. Our molecule set and results are available for other researchers to use in testing.

## Open Peer Review

Reviewer Status

Invited Reviewers

1

2

version 1

03 Dec 2020



report



report

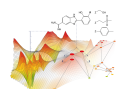
1. **Sereina Riniker** , ETH Zürich, Zürich, Switzerland

2. **Sreyoshi Sur** , Moderna, Cambridge, USA

Any reports and responses or comments on the article can be found at the end of the article.

**Keywords**

force field, molecular modeling, OpenFF, OPLS, molecular mechanics, molecular dynamics, quantum mechanics



This article is included in the **Chemical Information Science** gateway.

**Corresponding author:** David L. Mobley ([dmobley@mobleylab.org](mailto:dmobley@mobleylab.org))

**Author roles:** **Lim VT:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Hahn DF:** Data Curation, Resources, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Tresadern G:** Project Administration, Resources, Software, Supervision, Visualization, Writing – Review & Editing; **Bayly CI:** Conceptualization, Methodology, Writing – Review & Editing; **Mobley DL:** Conceptualization, Data Curation, Funding Acquisition, Methodology, Project Administration, Resources, Software, Supervision, Validation, Writing – Review & Editing

**Competing interests:** DLM serves on the scientific advisory board for OpenEye Scientific Software and is an Open Science Fellow with Silicon Therapeutics.

**Grant information:** We acknowledge funding from NIH grants GM098973 and (to DLM) 1R01GM108889-01 and 1R01GM124270-01A1. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2020 Lim VT *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Lim VT, Hahn DF, Tresadern G *et al.* **Benchmark assessment of molecular geometries and energies from small molecule force fields [version 1; peer review: 2 approved]** F1000Research 2020, 9(Chem Inf Sci):1390 <https://doi.org/10.12688/f1000research.27141.1>

**First published:** 03 Dec 2020, 9(Chem Inf Sci):1390 <https://doi.org/10.12688/f1000research.27141.1>

## Introduction

The study of chemical and biological systems relies on an accurate assessment of the energetics and geometries of the systems. Many computational methods serve to help investigate these systems, ranging from more accurate, higher cost quantum mechanical techniques to more approximate methods which compromise accuracy in favor of increased efficiency. Classical mechanics-based calculations fall into the latter group, and have an advantage over more theoretically rigorous calculations of being able to model larger systems over longer time scales<sup>1-4</sup>.

The modeling and simulation of molecular systems in classical mechanical calculations typically requires a force field, a set of energy functions and associated parameters comprising the potential energy function. This potential energy function defines interactions between components in the system based on the coordinates of its particles<sup>5,6</sup>.

Force fields have a long history of development. Strategies for force field development vary in terms of the chemical space covered, the types of data used for training, and the approach to optimize parameters given a set of input data<sup>7-10</sup>. The training data used to develop a force field usually includes input data from both experimental and reference quantum mechanical (QM) calculations. This finite amount of input data is carefully chosen to be representative of the systems for which the force field is designed. The limit of accuracy for some force field is measured by its ability to reproduce experimental observables, such as hydration free energies. When experimental evidence is unavailable, the force field can be assessed with respect to quantum mechanical data, for instance its ability to reproduce QM geometries and relative energies. Given the complexity of force field development, including multidimensional input data, various functional forms, and approaches to chemical perception<sup>11</sup>, force fields vary in how accurately they can compute properties of interest. Indeed, many examples serve to highlight the limitations of force fields<sup>12-16</sup>.

Our focus in this work is on force fields for small molecules, which are instrumental in drug discovery; for instance, evaluating binding free energies and modeling ligand binding poses. Relatively few literature studies evaluate force field accuracy on general small drug-like molecules, in contrast to force fields for proteins<sup>17-25</sup>, nucleic acids<sup>26-28</sup>, carbohydrates<sup>29-32</sup>, and other specific chemical systems<sup>33-45</sup>. On small molecules, these studies comprise predictions of solvation free energies<sup>46,47</sup>, strain energies<sup>48</sup>, experimental osmotic coefficients<sup>49</sup>, partition coefficients<sup>50,51</sup>, conformer energies<sup>52-56</sup>, conformer geometries<sup>55,57</sup>, and robustness of parameterization<sup>57</sup>. Most of these studies assess four or fewer force fields on molecule sets up to several hundreds of molecules. We present a broader assessment of general small molecule force fields on a large, diverse library of drug-like compounds and evaluate how accurately these force fields perform. We use QM data as a valuable source of information for force field assessment and to explore chemical space relatively quickly and easily.

In this work, we benchmarked small molecule force fields with respect to quantum mechanical results. We assessed nine force fields belonging to four families: the General Amber Force Field, first and second generations (GAFF<sup>58</sup> and GAFF2<sup>59</sup>); the Merck Molecular Force Field, initial and "static" versions (MMFF94<sup>60-64</sup> and MMFF4S<sup>56,65</sup>); the third extended version of the Optimized Potentials for Liquid Simulations Force Field (OPLS3e<sup>66</sup>); and the SMIRKS-based force fields from the Open Force Field Initiative (SMIRNOFF99Frosst<sup>67</sup> and its successor OpenFF "Parsley"<sup>68</sup>, versions 1.0, 1.1, and 1.2). For a dataset of 22,675 molecular structures of 3,271 small molecules, we conducted molecular mechanics (MM) energy minimizations using force fields and evaluated optimized geometries and energies, compared with reference to quantum mechanical data. We also identified particular chemical groups that represent systematic outliers in the force field-optimized geometries and energies. This work provides a general understanding of the strengths of different small molecule force fields and identifies areas of improvement for future force field development.

## Methods

### We acquired reference geometries and energies of molecules from QCArchive and grouped them by connectivity

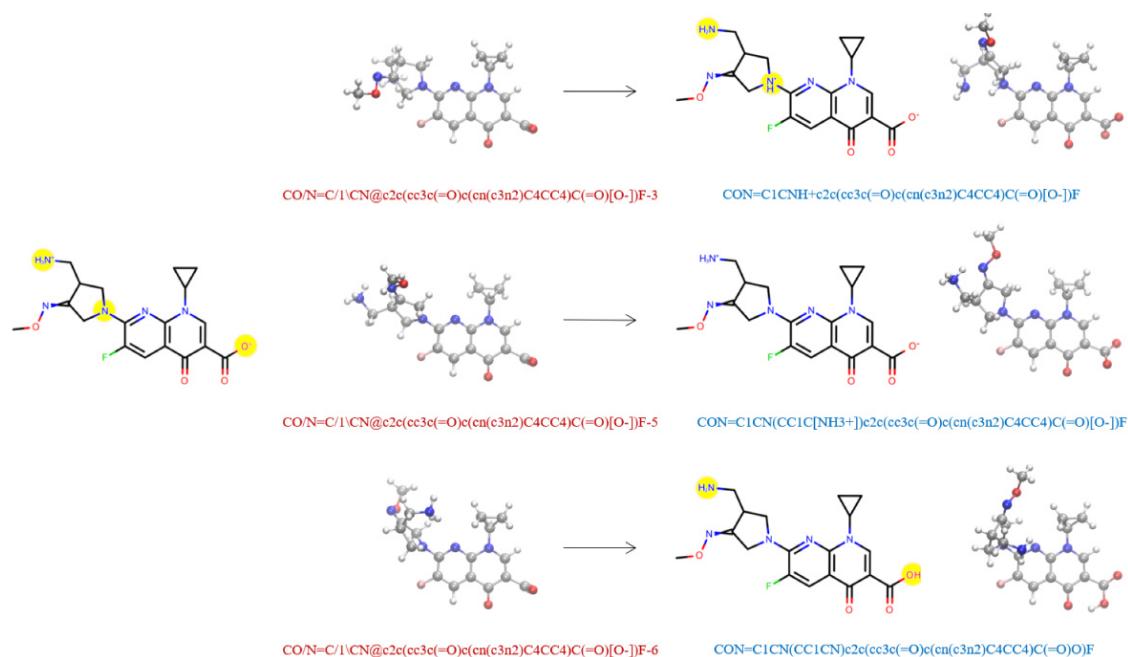
We obtained the molecule set in this work from QCArchive<sup>69</sup> from the dataset labeled OpenFF Full Optimization Benchmark 1 (accessed November 11, 2019), which was created for the purpose of benchmarking OpenFF-1.0<sup>70,71</sup>. An initial preprint of this work was posted after benchmarking OpenFF-1.0, but subsequently we were able to include OPLS3e results and added benchmarking of OpenFF-1.1 and 1.2. It is important to note, then, that this dataset was not curated to present any force field in a particular light; it was selected for benchmarking OpenFF-1.0 and has been retained as-is for the present comparison. However, OpenFF-1.2 marked a substantial refit and used an expanded training set of molecules, selection of which was at least partially informed by

benchmarking of OpenFF-1.0<sup>72</sup>. These training set changes meant we had to remove some structures from our benchmark set to ensure there was no overlap between training and test sets. Particularly, we removed 2398 structures from 419 molecules which were used for training the more recent OpenFF-1.2.

Overall, the benchmark set was chosen to include a broad range of drug-like compounds<sup>71,73</sup>. This QCArchive dataset contains QM geometry-optimized structures and energies at the B3LYP-D3BJ/DZVP level of theory<sup>74-78</sup>. This method and basis set were chosen by the Open Force Field initiative to provide reasonably accurate conformational energies and geometries at moderate computational cost<sup>52,53</sup>.

In our dataset, we organized molecular structures such that conformers of the same molecule were grouped together if they have the same absolute (non-isomeric) graph. Importantly, we do not use the SMILES string listed in the QCArchive DataFrame to represent the molecule itself, because the identity of the molecule may change during QM geometry optimization due to changes in bonding/tautomerization, such as shown in Figure 1. Molecules with different tautomerization states, which have different chemical connectivity, are treated as distinct molecules in our study. While two molecular structures may start QM optimizations from the same connectivity, we only use their final geometries to identify and distinguish molecules based on their connectivity. We grouped together all structures in the dataset whose final geometries yielded the same canonical isomeric SMILES string, as evaluated by `OEMolToSmiles` from the OpenEye OEChem Python toolkit<sup>79</sup>. The structures were then organized into conformer sets as perceived by OEChem's `OEAbsCanonicalConfTest`. This dataset organization procedure takes into account any molecular identity changes during QM optimization, such as if two molecules no longer had the same tautomerization state after QM optimization or if two different molecules ended up in the same tautomerization state. We ensured that what we identified as a molecule, and all of its given conformers, contained the same chemical connectivity.

The resulting QM geometries were used as input structures for gas phase energy minimizations using the following small molecule force fields: GAFF<sup>58</sup>, GAFF2<sup>59</sup>, MMFF94<sup>60-63</sup>, MMFF4S<sup>56,65</sup>, OPLS3e<sup>66</sup>,



**Figure 1. Molecules may change connectivity during QM optimization leading to different tautomers.** On the left hand side, we show the Lewis structure and three associated conformers of an example molecule from the QCArchive OpenFF Full Optimization Benchmark 1 dataset. Yellow circles highlight the regions of potential tautomerization changes. The QCArchive SMILES labels are colored in red. The right hand side shows the structures after QM optimization. The canonical isomeric SMILES labels representing the optimized molecules are colored in blue. Only the middle structure retains the original tautomeric identity. In our dataset, the geometries on the right hand side would be analyzed as distinct molecules.

SMIRNOFF99Frosst<sup>67</sup>, and Parsley<sup>68</sup>. The SMIRNOFF99Frosst version used here is SMIRNOFF99Frosst-1.1.0.offxml. SMIRNOFF99Frosst is a SMIRKS Native Open Force Field (SMIRNOFF) and descends from the AMBER parm99 force field as well as Merck-Frosst's parm@frosst. Its successor is the OpenFF Parsley force field, for which three versions (1.0, 1.1, and 1.2) were evaluated, specifically openff\_unconstrained-1.0.0-RC2.offxml (OpenFF-1.0), openff\_unconstrained-1.1.1.offxml (OpenFF-1.1) and openff\_unconstrained-1.2.0.offxml (OpenFF-1.2).

### We assigned FF parameters then energy minimized all molecules

Each structure was assigned AM1 Mulliken-type partial charges with bond-charge corrections (AM1-BCC charges)<sup>80,81</sup> from the electrostatically least-interacting functional group technique. The partial charges were generated using the openmoltools wrapper<sup>82</sup> to OpenEye's oequacpac charging engine<sup>79</sup> calling OEAM1BCCEL-F10Charges.

To assign force field parameters to each molecule, we used antechamber and tleap<sup>59</sup> via openmoltools<sup>82</sup> for the GAFF2 force fields. Parameter assignment as well as energy minimization for the MMFF94S force fields were handled using OpenEye oeszybki<sup>79</sup>. The custom OPLS3e charge and parameter assignment was performed in two steps using Schrodinger Maestro (v. 2020-1<sup>83</sup>). First, ligprep<sup>84</sup> was used to convert ligands to Maestro format with settings to avoid modifying protonation or tautomeric states. Then ffbuilder was used to check for missing parameters and launch torsional drives with constrained minimization at the B3LYP/6-31G\* level followed by single-point M06-2X/cc-pVTZ(-f) calculations. New OPLS3e parameters were derived for 1096 dihedrals, at a computational cost of about 100 CPU cores for 2 weeks to run high level DFT torsion fitting. This made the OPLS3e calculations substantially more costly; this may be in part because some of the benchmark set originates from eMolecules and consists of rather diverse and in some cases unusual chemistry which is not well captured by OPLS3e without additional parameterization.

We used the Open Force Field toolkit for SMIRNOFF99Frosst and Parsley, in all cases applying pre-assigned charges as described above. For the minimizations with OPLS3e, Schrodinger's macromodel was used with the PRCG algorithm with a gradient tolerance of 0.05 kcal/mol. All other energy minimizations were completed in OpenMM<sup>85</sup> using the LBFGS algorithm with an energy tolerance of 5.0e-9 kJ/mol and 1500 maximum number of iterations.

We removed any molecular structure that was not successfully parameterized with all force fields. This set consisted of 721 structures that were unable to be parameterized by GAFF or GAFF2, 522 structures that raised an error during OpenMM setup through the Open Force Field toolkit, and 50 which had various OpenEye charging or stereochemical perception errors. Our pruned set going into energy minimization contained 22,675 structures from 3,271 molecules with unique chemical connectivity. Corresponding files containing QM geometries and energies, SMILES strings and depictions are deposited on GitHub, benchmarkff/molecules/set\_v03\_non\_redundant/<sup>86</sup>. The repository also contains the structures removed due to parameterization or setup errors (in the benchmarkff/molecules/issues directory) and the structures removed due to overlap with the OpenFF-1.2 training set (in the benchmarkff/molecules/set\_overlapping directory).

### We evaluated relative energies and geometric agreement with optimized QM geometries

We compared the energy-minimized geometries and energies for each force field with respect to the QM reference data by computing the following metrics: relative energy difference (ddE), root-mean-square deviation of atomic positions (RMSD), and torsion fingerprint deviation (TFD)<sup>87-89</sup>. The relative energy difference (ddE) between the FF and QM energy for the *i*th conformer of a specific molecule was computed using the following expression:

$$ddE_i = dE_{FF,i} - dE_{QM,i} \quad (1)$$

$$= [\text{FF energy}(i) - \text{FF energy}(0)] - [\text{QM energy}(i) - \text{QM energy}(0)], \quad (2)$$

where the 0th conformer is defined as the conformer with the lowest QM energy for the given molecule.

Molecules may change conformation after energy minimization, which may lead to lower agreement between FF and QM energies for minimizations beginning from a particular conformer. To address this potential issue, we performed a conformer matching process for each FF structure which considered the final optimized geometries and energy differences. We ensured that every MM conformer was within 1.0 Å RMSD of a QM reference

structure. The QM reference conformer was removed from analysis if there were no FF conformers that matched it within 1.0 Å RMSD. Furthermore, if a molecule ends up with two of the “same” FF-minimized conformers compared to a QM reference structure, we only keep the FF conformer with the lowest RMSD score while any redundant conformers are removed from analysis. For this reason, the number of total molecular structures for each force field will likely differ after conformer matching as the intricate conformational energy landscapes are represented differently by various QM methods and force fields. Then, the mean signed deviation (MSD) was computed over all  $N$  conformers of each molecule with Equation 3, iterating over the relative energy  $dE$  of each conformer  $i$ . The reference conformer with  $dE = 0$  was removed from the MSD calculation. The molecule MSDs were then represented in violin plots to compare among all force fields.

$$MSD = \frac{1}{N-1} \sum_{i=1}^{N-1} dE_{FF,i} - dE_{QM,i} \quad (3)$$

To compare FF geometries with QM geometries, we used RMSD and TFD scores. The RMSD values, calculated with OpenEye OERMSD, took into account hydrogen atoms, symmetry-related transformations, and overlaid structures to yield the lowest possible RMSD. TFDs were computed using the RDKit Python library. We evaluated each of these three metrics individually and looked for potential correlations between energies and structures in terms of agreement with reference QM data.

We identified specific OpenFF-1.2 parameters which were overrepresented in high TFD regions. First, we collected all molecules having TFD scores above a visually determined cutoff of 0.12. We considered only molecules with distinct chemical connectivity. For each molecule in the high TFD group, we took the unique set of all parameters applied to the molecule. Thus, while a parameter may be applied multiple times to a single molecule, it would count only one time for that molecule. A parameter may be included multiple times when considering the entire TFD subset if it is applied to more than one molecule in the subset. For each parameter  $i$  in the force field, we computed its representation ratio as the fraction of molecules which apply that parameter:

$$\text{representation ratio}(i) = \frac{\text{number of molecules which apply parameter } i \text{ at least once}}{\text{total number of molecules in set}} \quad (4)$$

This ratio was calculated for the high TFD subset as well as for the full set of molecules. To identify whether a parameter would be found more likely in the high TFD subset than the full molecule set, we compared the two representation ratios between the subset and full set using the one-sample Z-test for proportions. The population proportion for some parameter was designated as its representation ratio in the full molecule set, and the sample proportion was assigned to be the parameter's representation ratio in the high TFD subset. We took the 95% confidence intervals from this Z-test to be the error bars for the representation ratios of the high TFD subset. Parameters having 20 or fewer molecules in the high TFD subset were excluded from further analysis and plotting due to inconclusive results from small sample sizes.

The complete Python code used for the setup, FF minimizations, and analysis of this work is open sourced and available on Github at <https://github.com/MobleyLab/benchmarkff<sup>90</sup>>. An earlier version of this article can be found on chemRxiv (doi: <https://doi.org/10.26434/chemrxiv.12551867.v2>).

## Results and discussion

Here, we present and discuss our results comparing several general small molecule force fields against reference QM data. We are interested in two major categories of comparison – energetic agreement and geometric agreement. Particularly, an ideal force field will yield the same energy minima or optimized geometries as the QM energy landscape, with no additional minima, and the relative energies of those minima will agree between QM and MM. Thus, to assess performance in these two categories, we computed relative conformer energies and compared these between MM and QM, as well as assessed geometric agreement of MM optimized geometries with those from QM. We also identified specific parameters for the improvement of future versions of the OpenFF small molecule force field.

Our study relies on the assumption that force field accuracy can be evaluated using gas phase energies and geometries. One of the greater goals of force field science, such as that of the Open Force Field Initiative, is building force fields that will work well in the condensed phase (e.g., small molecules in solution or binding to biomolecules). That being said, we make our assumption based on two key observations. First, force fields—especially

those in the AMBER family—are usually fitted to reproduce gas phase conformational energies and geometries<sup>58</sup>. This means that we are testing these force fields on properties they are fitted to reproduce. Second, bonded parameters are not expected to change significantly on transfer to the condensed phase. Rather, non-bonded interactions are particularly important in condensed phase simulations. Of the non-bonded interactions, electrostatics models are often polarized beyond what would be expected in the gas phase in order to reproduce condensed-phase properties, and Lennard-Jones parameters can be tuned to reproduce condensed phase properties (as has been a particular focus of the OPLS force fields<sup>91,92</sup>). Even when these are done, force fields retain bonded terms parameterized to reproduce QM geometries and energetics, further emphasizing the importance of testing in such a context. We therefore believe our assumption is reasonable and that this work warrants investigation.

We start our force field benchmark analysis by comparing FF energies to QM energies. Here, since our choice of reference energy for MM is arbitrary, we choose to compare relative conformer energies. For any given molecule, an ideal force field would have relative energies for different conformers in MM that agree with those for the same conformers in QM. For the differences in relative conformer energies that we computed—that is, the difference between the MM relative conformer energies and the QM relative conformer energies—a FF with greater agreement to QM should have more values around or at 0 kcal/mol, and a FF with lower agreement with QM would exhibit a broader distribution of values that are further away from 0 kcal/mol.

The relative conformer energies of all molecular structures in our dataset with the nine force fields were generally within  $\pm 50$  kcal/mol of the energies of the most favorable QM conformers (Table 1), and 95% of the relative conformer energies were within 11 kcal/mol. However, GAFF had outlying energies that were several orders of magnitude beyond this range (row 1 of Table 1). These energies were traced back to six molecules (62 conformers thereof) shown in Figure 2. These molecules all contain a polar hydrogen atom which, after geometry optimization, overlaps with its parent atom. The spurious overlap of these hydrogen atoms, and associated energy extremes, is due to a missing van der Waals parameter in GAFF. In GAFF2 (and SMIRNOFF99Frosst and subsequent OpenFF force fields<sup>11,67,68</sup>), hydroxyl hydrogens no longer have zero Lennard-Jones parameters, which seems to eliminate the problem for these molecules. Similar collapse of hydroxyl groups in close proximity has been observed previously in force fields with zero LJ parameters for hydroxyl hydrogens<sup>11</sup>.

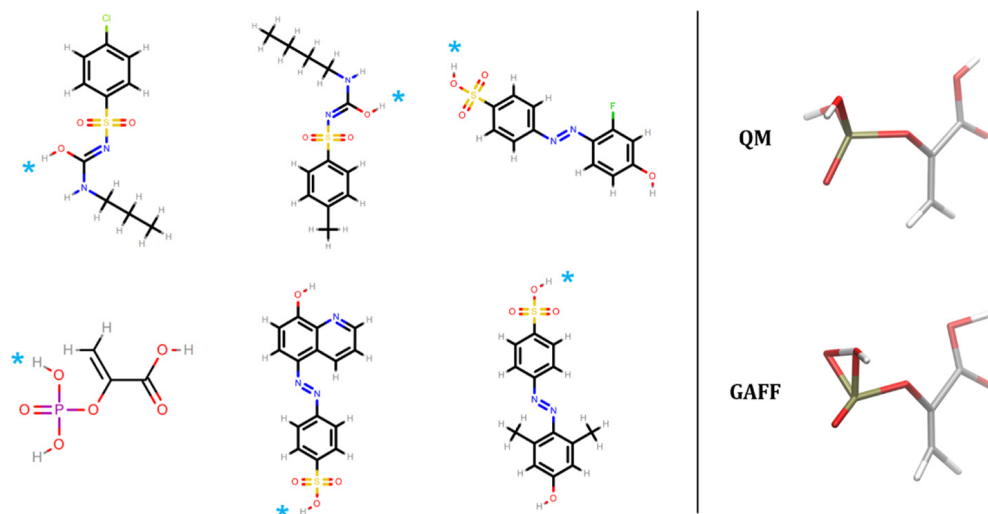
After excluding the 62 GAFF outliers, the ddE energies are histogrammed in Figure 3 and *Extended data*, Figure S.1<sup>93</sup>. The difference between MM relative conformer energies and QM relative conformer energies exhibit very similar distributions for all force fields. All distributions appear asymmetric, having a skew towards more negative ddE values than positive ones, indicating that the conformer energy differences may be underpredicted by MM compared to QM. Force fields of the same family tend to be more consistent with each other (GAFF and GAFF2,

**Table 1. Minimum and maximum ddE values as computed in Equation 2 for all structures of each force field.** Energy units are in kcal/mol.

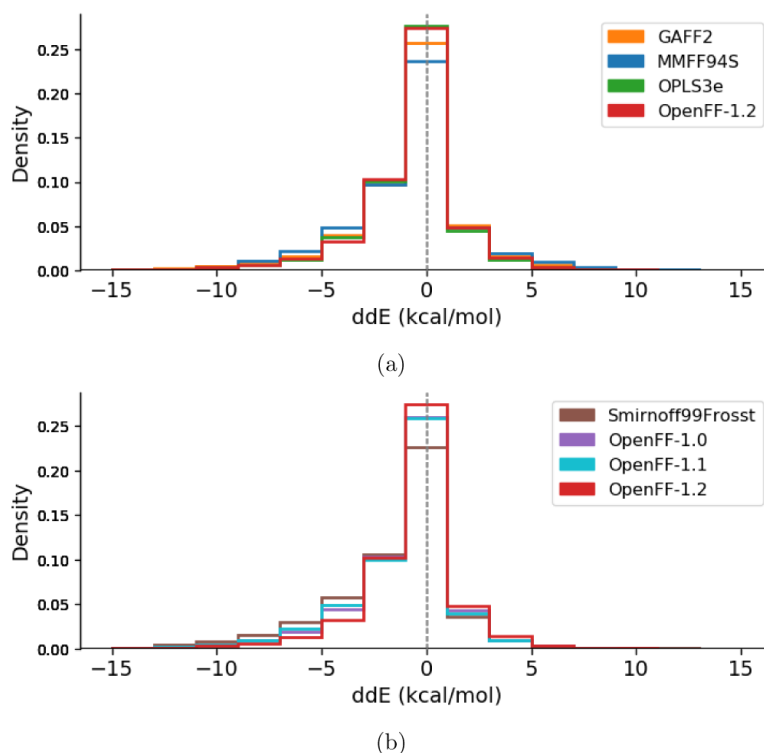
Force field	min ddE	max ddE
GAFF	-35002325.4	5549.7
GAFF <sup>a</sup>	-44.1	14.8
GAFF2	-43.7	15.6
MMFF94	-52.1	29.8
MMFF94S	-49.5	25.1
SMIRNOFF99Frosst	-42.8	18.8
OpenFF-1.0	-38.4	19.3
OpenFF-1.1	-38.6	18.3
OpenFF-1.2	-37.9	15.4
OPLS3e	-30.4	9.6

<sup>a</sup> With outliers removed.





**Figure 2. Molecules with extreme relative conformer energies for GAFF.** The right hand side depicts the QM and FF geometries for phosphoenolpyruvic acid. The GAFF structure shows a representative overlap of a polar hydrogen atom with its connected parent atom due to a missing van der Waals parameter. On the left hand side, the overlapping hydrogen for the six molecules are denoted by cyan asterisks.

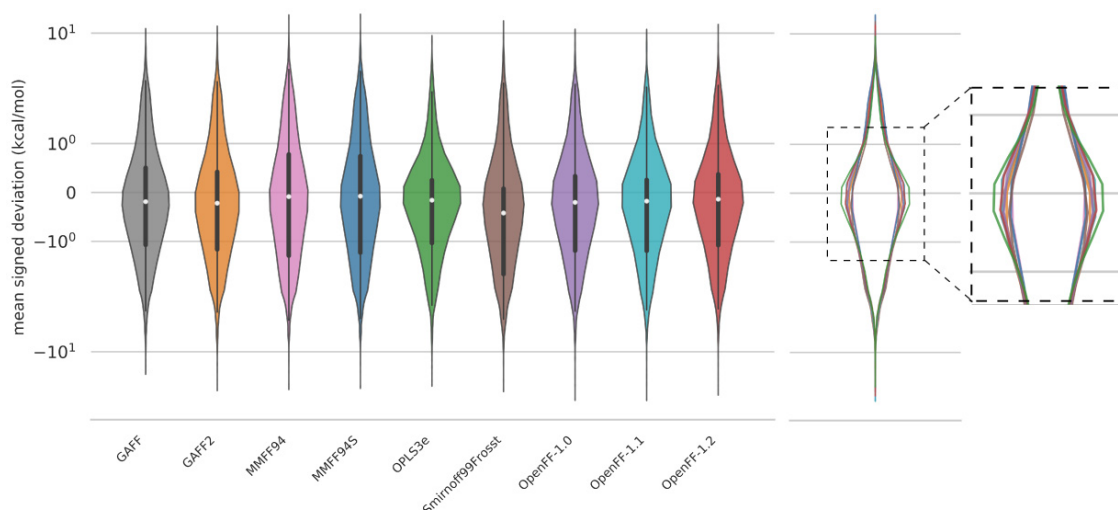


**Figure 3. Histograms of the relative conformer energy differences as computed in Equation 2 for each force field relative to QM.** Each molecular structure, including different conformers of the same molecule, is counted separately. Since the global minimum molecular structures were set to zero deliberately and add a constant offset to the central bin, they are removed from the counts. A force field having higher agreement with QM would have a higher bin centered at  $ddE = 0$  kcal/mol. (a) compares the latest release of all four force field families. (b) shows the four histograms belonging to the OpenFF family of force fields. OpenFF-1.0 (purple) and OpenFF-1.1 (light blue) overlap in the central bin. The corresponding graph comparing histograms of all calculated force fields can be found in the *Extended data*, Figure S.1<sup>93</sup>.

MMFF94 and MMFF94S), see *Extended data*, Figure S.1<sup>93</sup>. From these results, the qualitative ordering of force fields from lowest to highest agreement with QM energies goes as SMIRNOFF99Frosst < MMFF94 ~ MMFF94S < GAFF ~ GAFF2 ~ OpenFF-1.0 ~ OpenFF-1.1 < OpenFF-1.2 < OPLS3e. In other words, the peak size around  $\text{ddE} = 0$  kcal/mol (the fraction of molecules described particularly well) is greatest for OPLS3e, closely followed by OpenFF-1.2. OPLS3e predicts  $55.3 \pm 0.3\%$  of conformers within 1 kcal/mol of QM, with OpenFF-1.2, GAFF2, and MMFF94S identifying  $54.8 \pm 0.3\%$ ,  $51.3 \pm 0.3\%$ , and  $47.0 \pm 0.3\%$  respectively. By this metric, OPLS3e and OpenFF-1.2 seem to exhibit roughly similar performance, with the other force fields performing somewhat worse. **Figure 3b** illustrates the progress made within the OpenFF family of force fields. The predecessor SMIRNOFF99Frosst performs worst of all investigated force fields and is improved upon by the first releases OpenFF-1.0 and OpenFF-1.1, which show intermediate performance. Finally, the most recent release OpenFF-1.2 indicates further improvement.

Given that two conformers starting from the same geometry may optimize to two distinct conformers after FF minimization, we took another approach to analyzing energy distributions, only considering the FF conformers that correspond to a QM counterpart. A FF conformer is deemed to have a “match” with a QM conformer if its RMSD is less than or equal to 1 Å (see more details in Methods). The number of matched conformers for each force field are: 20,815 (GAFF), 20,836 (GAFF2), 20,674 (MMFF94), 20,684 (MMFF94S), 21,961 (OPLS3e), 16,177 (SMIRNOFF99Frosst), 19,103 (OpenFF-1.0), 17,965 (OpenFF-1.1), 21,428 (OpenFF-1.2). The mean signed deviation of the matched conformer energies are shown as violin plots in **Figure 4**. The violin plots are scaled such that each violin has the same area. This figure shows that the mean signed deviation of relative conformer energies is also fairly consistent between different force fields as seen in **Figure 3**. Upon closer inspection, the violins for OPLS3e and OpenFF-1.2 are slightly wider around 0 kcal/mol (and narrower elsewhere), signifying marginally higher agreement with QM energies. Equivalent results for an RMSD threshold of 0.3 Å to the QM structure is shown in the *Extended data*, Figure S.2<sup>93</sup>. With this lower RMSD criteria, the number of structures within the cutoff is roughly halved compared to a threshold of 1 Å while the ranking of force fields remains unaltered. Note that this conformer filtering step was only used for analyzing the energies in the violin plots, and other results throughout this work do not rely on matched conformers.

We next examine agreement between FF-optimized geometries and those from QM, as calculated by each molecule’s RMSD and TFD scores with reference to the parent QM-optimized geometries. While RMSD is the more common metric, it may depend on the molecule size, complicating interpretation of geometric agreement<sup>94,95</sup>. In contrast, TFD was designed to be more independent of molecule size in order to compare molecular conformations

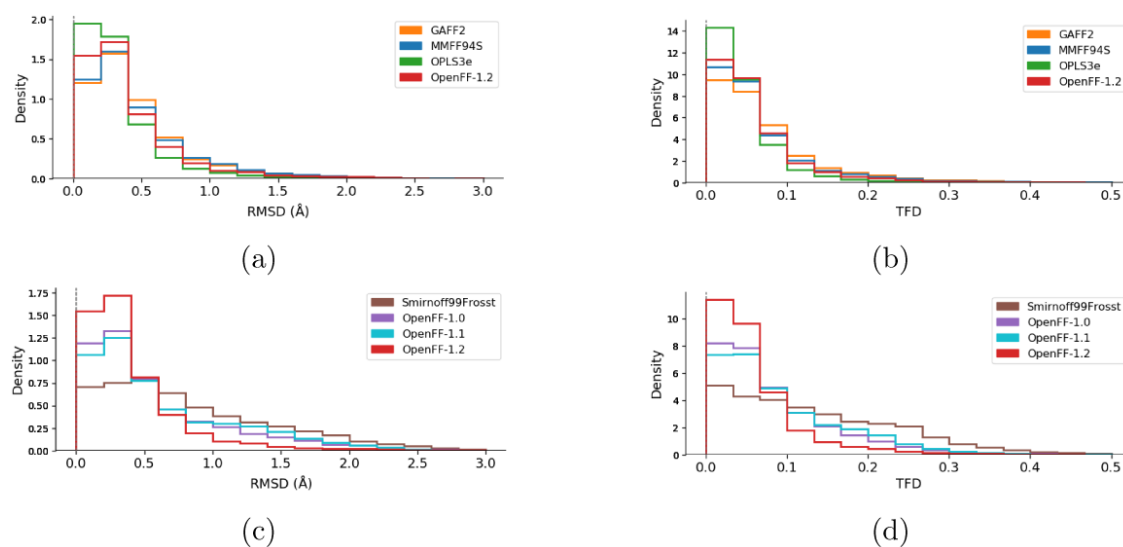


**Figure 4.** Violin plots of the mean signed deviations of MM conformer energies relative to QM conformer energies as computed in **Equation 3**. The energy MSDs only take into account structures matched within 1 Å of the QM reference structure, so there are minor differences in the amount of data used to plot each violin (see text). To correct for this, each plot was scaled to the same area. The vertical axis is shown on a logarithmic scale. An overlay of the violin plots on the right panel better shows the subtle distinctions between the force fields in the most populated region, near zero error. An equivalent graph with an RMSD threshold of 0.3 Å is shown in the *Extended data*, Figure S.2<sup>93</sup>.

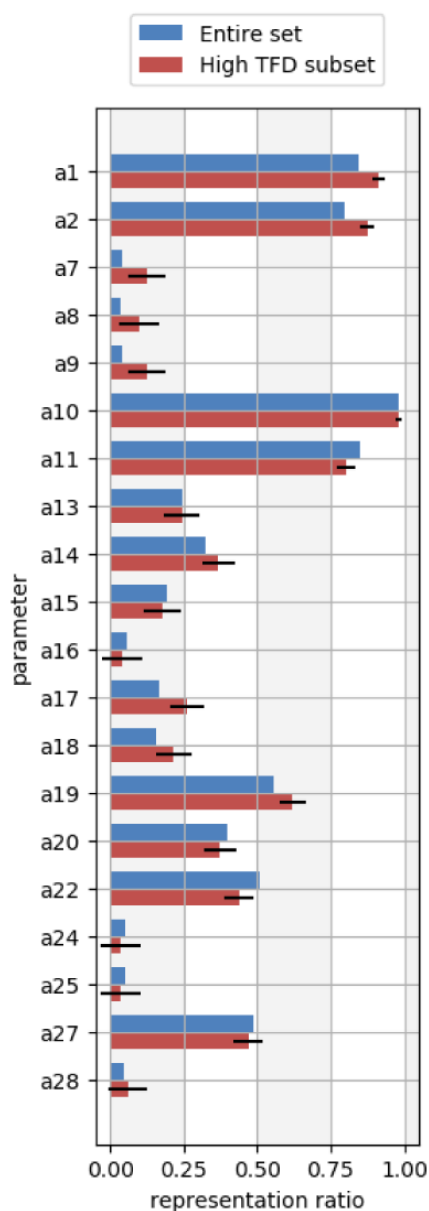
more meaningfully<sup>87</sup>. This can help offset issues with RMSD where larger, more flexible molecules can contribute the most to RMSD. The TFD score between two molecular structures is evaluated by computing, normalizing, and Gaussian weighting the (pseudo)torsion deviation for each bond and ring system. While TFD is normalized from 0 to 1, RMSD is unbounded. Both RMSD and TFD are similar in that a higher value signifies lower agreement between the geometries of the two molecules. A FF which yields optimized geometries closer to those of QM would have generally smaller RMSD/TFD values. We calculated RMSD and TFD scores for all MM optimized geometries with respect to QM geometries. We plotted this data in histograms in [Figure 5](#).

In terms of geometry agreement, we observed similar results between the RMSD and TFD plots. The ranking of the force fields is mostly the same as with the *ddE* rankings above, with OPLS3e performing best followed by the latest open force field release, OpenFF-1.2. One major difference is the ranking of MMFF94S over GAFF2, while the latter had better agreement with QM in terms of *ddE*. The OpenFF force fields show clear improvement with newer versions by having higher densities close to zero and also by having tails successively reduced. Although SMIRNOFF99Frosst had a non-negligible density at RMSD > 2 Å, virtually all structures optimized with OpenFF-1.2 agree with the QM structures with RMSD < 2 Å ([Figure 5c](#)). Both TFD and RMSD distributions show qualitatively the same ranking of force fields, whereas the quantitative differences appear to be of different magnitudes. For example, MMFF94S is very close to GAFF2 in terms of RMSD ([Figure 5a](#)). According to TFD, MMFF94S appears to be closer to OpenFF-1.2, with GAFF2 having less agreement with QM ([Figure 5b](#)).

From the histograms, we can identify areas for force field refinement of molecular geometries by analyzing molecules with significant conformational differences from the QM reference (molecules with TFD values > 0.12), and in particular by focusing on parameters which occur more frequently than expected in such molecules. Parameters which are overrepresented in molecules with significant deviations are more likely to be responsible for such deviations. To assess this, we computed the representation ratio ([Equation 4](#)) for each OpenFF-1.2 force field parameter in both the high TFD molecule subset as well as in the full set of molecules. We estimated whether each parameter was applied more frequently in the high TFD subset compared to the full set by computing the one-sample Z-test for proportions. [Figure 6](#) shows the results for a subset of the OpenFF-1.2 force field parameters, wherein the parameters of interest have a statistically significantly higher representation ratio in the high TFD subset within a 95% confidence interval. These parameters are listed in [Table 2](#) for the complete OpenFF-1.2 force field, and likely warrant further investigation as a possible cause of deviations from the QM reference. The complete set of OpenFF-1.2 representation ratio plots are placed in the *Extended data*, [Figures S.4-S.6](#)<sup>93</sup>.



**Figure 5.** Histograms of the RMSD (**a, c**) and TFD (**b, d**) values between force field structures as compared to QM structures. Values closer to zero indicate higher geometric similarity for both RMSD and TFD. Panels (**a**) and (**b**) compare the families of force fields (GAFF2, MMFF94s, OPLS3e, and OpenFF-1.2). Panels (**c**) and (**d**) compare the force fields of the OpenFF family (Smirnoff99Frosst, OpenFF-1.0, OpenFF-1.1, and OpenFF-1.2). The corresponding graphs with histograms of all force fields are shown in the *Extended data*, [Figure S.3](#)<sup>93</sup>.



**Figure 6. Prevalence of a subset of angle parameters used in the OpenFF-1.2 force field as calculated by Equation 4.** The blue bars represent the parameter ratios from the full molecule set, and the red bars represent the parameter ratios from only the set of molecules with TFD values greater than 0.12. Error bars denote the 95% confidence interval determined from the one-sample Z-test for proportions. Parameters which are estimated to be overrepresented in molecules with high TFDs have statistically significant differences between the full set and high TFD set of parameter ratios (also see Table 2). Parameters with statistically significant differences in this plot are a1, a2, a7, a8, a9, a17, a18, and a19.

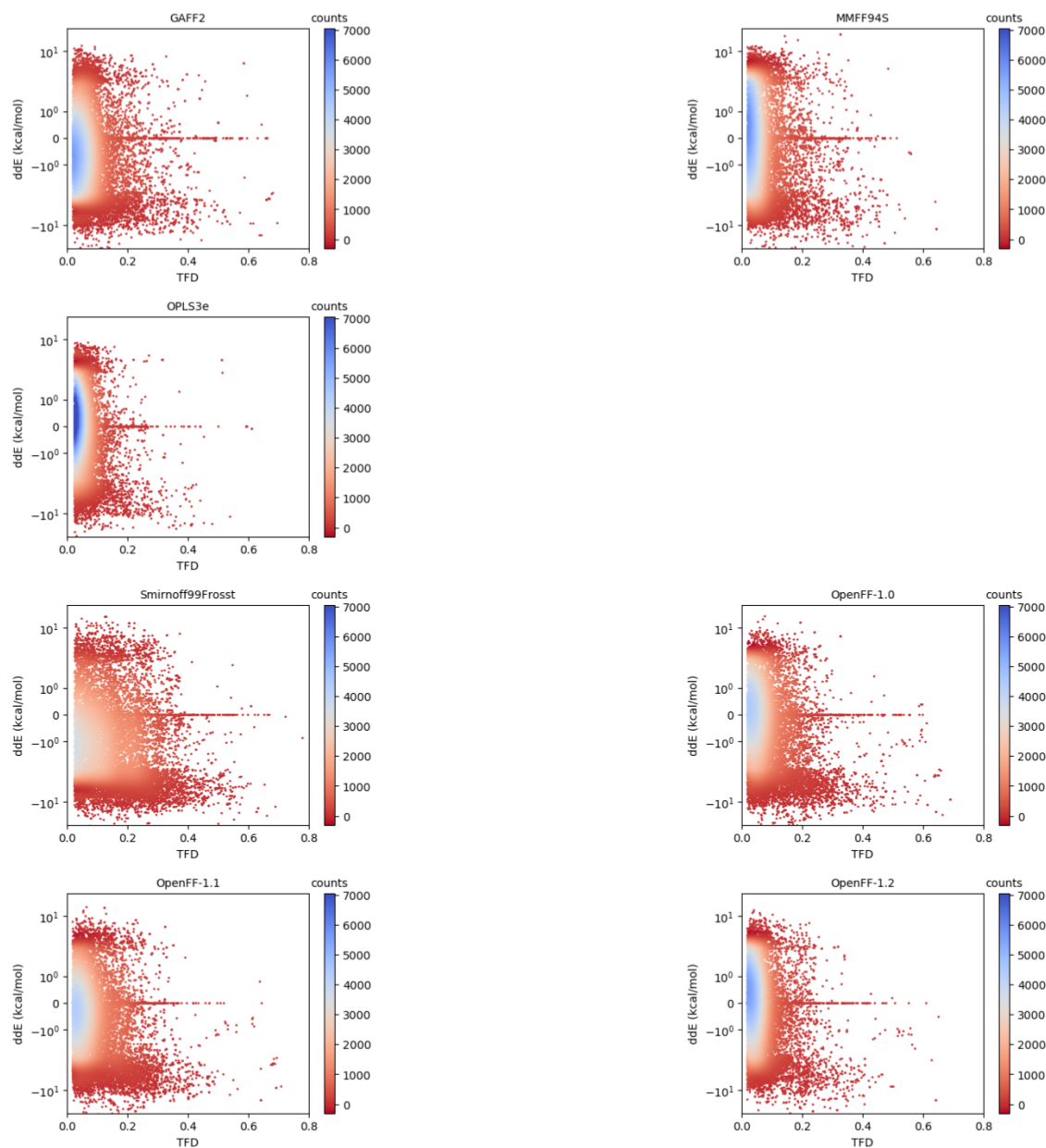
We then sought to determine if there was a dependence between the relative energies and geometries. Scatter plots of ddE versus RMSD/TFD are shown for all force fields in Figure 7. Each structure in our dataset is plotted as a single point. The ddE values are plotted on a logarithmic scale. We include in the *Extended data*, Figure S.7 analogous plots with ddE represented on a linear scale<sup>93</sup>. Given tens of thousands of points on each plot leading to many overlapping points, we applied a color gradient from red to blue to represent regions from low to high density, respectively. Similar to the data represented as one-dimensional histograms (Figure 3 and Figure 5), a higher density of points at the origin indicates results in better agreement with the reference QM data. There seems

**Table 2. OpenFF-1.2 force field parameters identified to be overrepresented in high TFD molecules.** These parameters show statistically significant differences ( $p < 0.05$ ) in representation ratios of the high TFD molecules compared to ratios of the full molecule set. Refinement of these parameters may address conformational differences in MM-optimized molecular geometries compared to QM-optimized geometries.

Angles	Bonds	Improper dihedrals	Van der Waals	Proper dihedrals
a1	b1	i3	n2	t1
a2	b2		n3	t2
a7	b3		n16	t3
a8	b7		n20	t4
a9	b9			t17
a17	b10			t18
a18	b83			t20
a19				t22
				t23
				t51
				t52
				t59
				t61
				t62
				t68

to be no general correlation between the energies and geometries. However, using this visualization we identified particular chemical moieties that represent outlying energies or geometries (vide infra).

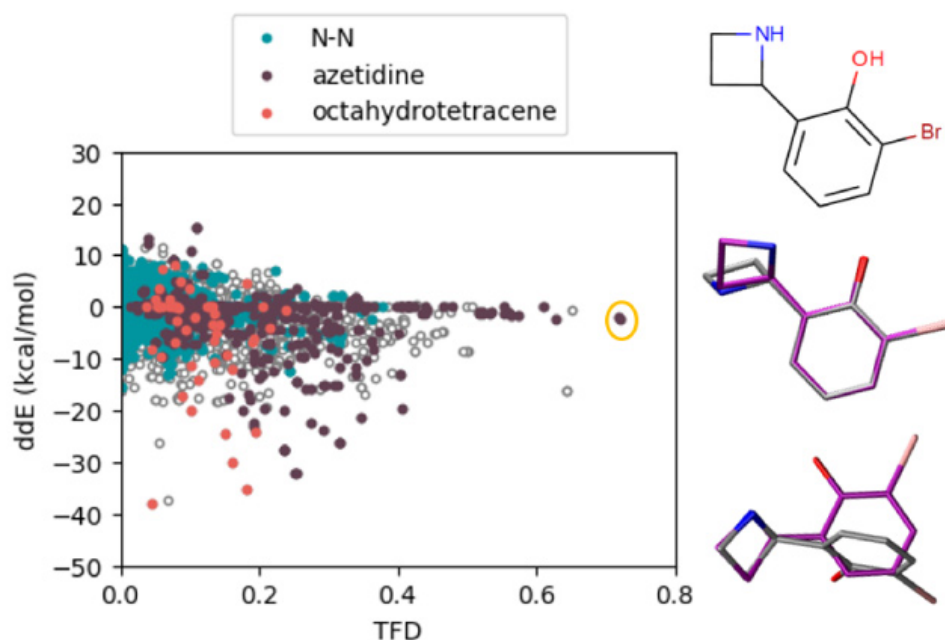
In this diverse set of molecules, we point out three particular moieties, those containing an N-N single bond (3824 structures), those containing an azetidine ring (543 structures), and a highly substituted octahydrotetracene (50 structures). These subsets are highlighted for OpenFF-1.2 in [Figure 8](#) (see *Extended data*, S.9 for other force field results<sup>93</sup>). Molecules containing an N-N single bond have a wide spread of energies with several ddE outliers between -10 to -20 kcal/mol. Structures with azetidine revealed both energy and geometry outliers in the Amber and OpenFF force field families. Lastly, the substituted octahydrotetracene scaffold was found to be challenging to all force fields in reproducing QM energies (an example is presented in the *Extended data*, [Figure S.8](#)<sup>93</sup>). These moieties represent systematic outliers that can be used in future studies investigating particular shortcomings of force fields or improving future versions of force fields. Indeed, some of these issues have been a focus of fitting of the OpenFF 1.1 and 1.2 force fields<sup>96</sup>. We have calculated the average and standard deviation statistics of ddE and TFD for the whole set of structures and the subsets containing these moieties. The results are listed as *Extended data*, [Table 1](#) and [Figure S.10](#)<sup>93</sup>. Both the spread and average of the distributions of the subset are generally larger than the ones of the whole set, emphasizing that these moieties are challenging to be parameterized. For the OpenFF family of force fields, a clear improvement in these statistics can be seen for the newer versions, especially for the N-N moiety (both TFD and ddE) and the octahydrotetracene (in terms of ddE).



**Figure 7. Scatter plots of relative conformer energies (ddE) versus TFD scores.** The points are colored by the interpolated density of points in a certain area. Blue indicates region of high density, that is, high compactness of points in that area. A force field having better agreement in both relative energies and geometries with the QM reference would have more points around the origin (ddE = 0, TFD = 0), though it is presumably possible for a force field to improve along one axis without improving along the other. The vertical axis is represented on logarithmic scale; the same plots with linear scaling can be found in the *Extended data*, Figure S.7<sup>93</sup>.

## Conclusions

In this work, we presented a large-scale analysis of nine small molecule force fields in terms of their relative conformer energies and geometries compared to reputable QM data. Amongst the force fields (GAFF, GAFF2, MMFF94, MMFF94S, OPLS3e, SMIRNOFF99Frosst, OpenFF-1.0, OpenFF-1.1, and OpenFF-1.2), OPLS3e performed best in terms of reproducing QM conformer energies and geometries. However, it is worth noting the higher computational cost of the high level DFT torsion fitting for generating the optimal OPLS3e parameters (likely in part due to the diversity of the present molecule set), whereas with the other force fields this step was rapid.



**Figure 8. Scatter plot of the OpenFF-1.2 force field of relative conformer energies versus TFD scores.** Colors highlight particular chemical groups that appear to be systematic outliers in energies or geometries. On the right hand side, we show a figure with high TFD and low ddE as circled in the scatter plot. The QM structure is in purple, and the force field structure is colored in silver. Analogous plots for all other force fields are shown in the *Extended data*, Figure S.9<sup>93</sup>.

The OpenFF versions showed improvements in both metrics with each new version, and the latest OpenFF-1.2 appears to be approaching the degree of accuracy of OPLS3e, at least on this dataset. This is despite the extra dihedral parameter fitting with OPLS3e. Thus OpenFF-1.2 seems to be positioned as the best open source/free small molecule force field in this study, as OPLS3e is proprietary.

Other aspects of interest included the ability of MMFF94 and MMFF94S to capture QM geometries better than several other force fields, but still not as well as OPLS3e or OpenFF-1.2, especially when using a more size-independent geometry measure. Finally, we identified particular chemical moieties that were systematic outliers in terms of relative energies or geometries. These N-N, azetidine, and octahydrotetracene-like compounds represent potential areas for improvement in future force field development.

Our work also highlights the progress the Open Force Field Initiative has made towards its goal of producing high quality public, open force fields built with infrastructure which enables rapid parameterization. Particularly, the series of OpenFF force fields presented here demonstrate marked improvements in accuracy over a relatively short time, and these improved force fields are available to everyone. One key challenge going forward will be to continue improving treatment of problematic areas of chemical space and expanding coverage. In parallel, future OpenFF updates will include improved treatment of torsions (via Wiberg bond order-based parameter interpolation<sup>97</sup> which was recently implemented in our toolkit) and better handling of trivalent nitrogen geometries<sup>98</sup> which we hope will boost performance further.

Beyond these specific conclusions, we believe the general strategies employed here for benchmarking force field performance will be useful far more broadly than this specific study. Particularly, comparing performance by both geometric and energetic measures is particularly important, as the analysis we have done demonstrates. Additionally, the availability of a large amount of public data in QCArchive facilitates straightforward large scale benchmarking in a way it has not been done previously.

We share our Python code comprising the setup, minimization, and analysis of this research on Github, available at: <https://github.com/MobleyLab/benchmarkff90>.

## Data availability

### Underlying data

Zenodo: Molecular geometries and energies from quantum mechanical calculations and small molecule force field evaluations. <https://dx.doi.org/10.5281/zenodo.4247859><sup>99</sup>.

### Extended data

Zenodo: Supporting Information: Molecular geometries and energies from quantum mechanical calculations and small molecule force field evaluations. <http://dx.doi.org/10.5281/zenodo.4299200><sup>93</sup>

This project contains the following extended data:

- Histograms for all force fields regarding energies of conformers, RMSD and TFD relative to QM reference data for all force fields investigated in this work
- Plots similar to those in [Figure 7](#) with linear scaling of the vertical axis
- Plots in the same manner of [Figure 8](#) for all force fields in this work
- Average and standard deviation statistics of relative energies and TFDs for different (sub)sets of structures
- An example of one of the octahydrotetracene-based structures having high deviation in dde

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

## Code availability

Source code used in conducting the modeling, analysis and plots is available on GitHub, with the specific version used here archived on Zenodo.

- Source code available from: <https://github.com/mobleylab/benchmarkff>
- Archived source code at time of publication: <https://dx.doi.org/10.5281/zenodo.4252694><sup>90</sup>
- License: [MIT](#)

---

## Acknowledgements

The authors thank the Open Force Field Initiative researchers for their contributions in project planning and feedback, especially Caitlin Bannan, Jordan Ehrman, Jessica Maat, Lee-Ping Wang, and Hyesu Jang. VTL appreciates conversations on statistics with Ethan Jain-Washburn. We also thank Jeffrey R. Wagner and Daniel G. A. Smith for code review and improvements.

---

## References

1. González MA: **Force Fields and Molecular Dynamics Simulations.** *JDN.* 2011; **12**: 169–200.  
[Publisher Full Text](#)
2. Cole DJ, Vilseck JZ, Tirado-Rives J, *et al.*: **Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning.** *J Chem Theory Comput.* 2016; **12**(5): 2312–2323.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Lane TJ, Shukla D, Beauchamp KA, *et al.*: **To milliseconds and beyond: challenges in the simulation of protein folding.** *Curr Opin Struct Biol.* 2013; **23**(1): 58–65.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Lange OF, vanderSpoel D, deGroot BL: **Scrutinizing Molecular Mechanics Force Fields on the Submicro second Time scale with NMR Data.** *Biophys J.* 2010; **99**(2): 647–655.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
5. Riniker S: **Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview.** *J Chem Inf Model.* 2018; **58**(3): 565–578.  
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Ponder JW, Case DA: **Advances in Protein Chemistry.** Protein Simulations; Aca-demic Press, 2003; **66**: 27–85.  
[Reference Source](#)
7. Nerenberg PS, Head-Gordon T: **New developments in force fields for biomolecular simulations.** *Curr Opin Struct Biol.* 2018; **49**: 129–138.  
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Monticelli L, Tieleman DP: **In Biomolecular Simulations: Methods and Protocols.** Monticelli, L., Salonen, E., Eds, *Methods in Molecular Biology* Humana Press:To-towa, NJ, 2013; 197–213.  
[Reference Source](#)



9. Hagler AT: **Force field development phase II: Relaxation of physics-based criteria... or inclusion of more rigorous physics into the representation of molecular energetics.** *J Comput Aided Mol Des.* 2019; **33**(2): 205–264.  
[PubMed Abstract](#) | [Publisher Full Text](#)
10. 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?** *J Comput Aided Mol Des.* 2019; **33**(2):133–203.  
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Mobley DL, Bannan CC, Rizzi A, et al.: **Escaping Atom Types in Force Fields Using Direct Chemical Perception.** *J Chem Theory Comput.* 2018; **14**(11): 6076–6092.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Cailliez F, Pernot P: **Statistical approaches to forcefield calibration and prediction uncertainty in molecular simulation.** *J Chem Phys.* 2011; **134**(5): 054124.  
[PubMed Abstract](#) | [Publisher Full Text](#)
13. Geballe MT, Guthrie JP: **The SAMPL3 blind prediction challenge: transfer energy overview.** *J Comput Aided Mol Des.* 2012; **26**(5): 489–496.  
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Hopkins CW, Roitberg AE: **Fitting of dihedral terms in classical force fields as an analytic linear least-squares problem.** *J Chem Inf Model.* 2014; **54**(7): 1978–1986.  
[PubMed Abstract](#) | [Publisher Full Text](#)
15. Köster A, Spura T, Rutkai G, et al.: **Assessing the accuracy of improved force-matched water models derived from Ab initio molecular dynamics simulations.** *J Comput Chem.* 2016; **37**(19): 1828–1838.  
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Mishra SK, Calabró G, Loeffler HH, et al.: **Evaluation of Selected Classical Force Fields for Alchemical Binding Free Energy Calculations of Protein-Carbohydrate Complexes.** *J Chem Theory Comput.* 2015; **11**(7): 3333–3345.  
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Guvench O, MacKerell AD: **In Molecular Modeling of Proteins.** Kukol A, Ed *Methods Molecular Biology™*; Humana Press: Totowa, NJ, 2008; 63–88.  
[Reference Source](#)
18. Martín-García F, Papaleo E, Gomez-Puertas P, et al.: **Comparing molecular dynamics force fields in the essential subspace.** *PLoS One.* 2015; **10**(3): e0121114.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Henriques J, Cragnell C, Skepö M: **Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment.** *J Chem Theory Comput.* 2015; **11**(7): 3420–3431.  
[PubMed Abstract](#) | [Publisher Full Text](#)
20. Cino EA, Choy WY, Karttunen M: **Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations.** *J Chem Theory Comput.* 2012; **8**(8): 2725–2740.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Lindorff-Larsen K, Maragakis P, Piana S, et al.: **Systematic validation of protein force fields against experimental data.** *PLoS One.* 2012; **7**(2): e32131.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Hornak V, Abel R, Okur A, et al.: **Comparison of multiple Amber force fields and development of improved protein backbone parameters.** *Proteins.* 2006; **65**(3): 712–725.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Shirts MR, Pitner JW, Swopes WC, et al.: **Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins.** *J Chem Phys.* 2003; **119**(11): 5740–5761.  
[Publisher Full Text](#)
24. Robustelli P, Piana S, Shaw DE: **Developing a molecular dynamics force field for both folded and disordered protein states.** *Proc Natl Acad Sci U S A.* 2018; **115**(21): E4758–E4766.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Matthes D, deGroot BL: **Secondary structure propensities in peptide folding simulations: a systematic comparison of molecular mechanics interaction schemes.** *Biophys J.* 2009; **97**(2): 599–608.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Ricci CG, deAndrade ASC, Mottin M: **Molecular dynamics of DNA: comparison of force fields and terminal nucleotide definitions.** *J Phys Chem B.* 2010; **114**(30): 9882–9893.  
[PubMed Abstract](#) | [Publisher Full Text](#)
27. Reddy SY, Leclerc F, Karplus M: **DNA polymorphism: a comparison of force fields for nucleic acids.** *Biophys J.* 2003; **84**(3): 1421–1449.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Feig M, Pettitt BM: **Structural equilibrium of DNA represented with different force fields.** *Biophys J.* 1998; **75**(1): 134–149.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
29. Matthews JF, Beckham GT, Bergensträhle-Wohlert M, et al.: **Comparison of Cellulose I $\beta$  Simulations with Three Carbohydrate Force Fields.** *J Chem Theory Comput.* 2012; **8**(2): 735–748.  
[PubMed Abstract](#) | [Publisher Full Text](#)
30. Pérez S, Imbertya A, Engelsen SB, et al.: **A comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates.** *Carbohydrate Research.* 1998; **314**(3–4): 141–155.  
[Publisher Full Text](#)
31. Asensio JL, Martín-Pastor M, Jimenez-Barbero J: **The use of CVFF and CFF91 force fields in conformational analysis of carbohydrate molecules. Comparison with AMBER molecular mechanics and dynamics calculations for methyl  $\alpha$ -lactoside.** *International Journal of Biological Macromolecules.* 1995; **17**(3–4): 137–148.  
[PubMed Abstract](#) | [Publisher Full Text](#)
32. Stortz CA, Johnson GP, French AD, et al.: **Comparison of Different Force Fields for the Study of Disaccharides.** *Carbohydr Res.* 2009; **344**(16): 2217–2228.  
[PubMed Abstract](#) | [Publisher Full Text](#)
33. Wang R, Bi S, Presser V, et al.: **Systematic Comparison of Force Fields for Molecular Dynamic Simulation of Au(111)/Ionic Liquid Interfaces.** *Fluid Phase Equilib.* 2018; **463**: 106–113.  
[Publisher Full Text](#)
34. Odegard G, Clancy T, Gates, T: **46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference.** American Institute of Aeronautics and Astronautics, eprint: [Reference Source](#)
35. Mijaković M, Polok KD, Kežić B, et al.: **A Comparison of Force Fields for Ethanol-Water Mixtures.** *Mol Simul.* 2015; **41**: 699–712.  
[Publisher Full Text](#)
36. Pisarev VV, Zakharov SA: **Comparison of Force fields for Molecular Dynamics Simulations of Hydrocarbon Phase Diagrams.** *J Phys Conf Ser.* 2018; **946**: 012100.  
[Publisher Full Text](#)
37. Ewen JP, Gattinoni C, Thakkar FM, et al.: **A Comparison of Classical Force-Fields for Molecular Dynamics Simulations of Lubricants.** *Materials (Basel).* 2016; **9**(8): 651.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
38. Skaf MS: **Molecular Dynamics Simulations of Dielectric Properties of Dimethyl Sulfoxide: Comparison between Available Potentials.** *J Chem Phys.* 1997; **107**: 7996–8003.  
[Publisher Full Text](#)
39. Pluhackova K, Kirsch SA, Han J, et al.: **A Critical Comparison of Biomembrane Force Fields: Structure and Dynamics of Model DMPC, POPC, and POPE Bilayers.** *J Phys Chem B.* 2016; **120**(16): 3888–3903.  
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Patra M, Karttunen M: **Systematic Comparison of Force Fields for Microscopic Simulations of NaCl in Aqueous Solutions: Diffusion, Free Energy of Hydration, and Structural Properties.** *J Comput Chem.* 2004; **25**(5): 678–689.  
[PubMed Abstract](#) | [Publisher Full Text](#)
41. Baucom J, Transue T, Fuentes-Cabrera M, et al.: **Molecular dynamics simulations of the d(CCAACGTTGG)(2) decamer in crystal environment: comparison of atomic point-charge, extra-point, and polarizable force fields.** *J Chem Phys.* 2004; **121**(14): 6998–7008.  
[PubMed Abstract](#) | [Publisher Full Text](#)
42. Mu Kosov DS, Stock G: **Conformational Dynamics of Trialanine in Water. 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS Force Fields to NMR and Infrared Experiments.** *J Phys Chem B.* 2003; **107**: 5064–5073.  
[Publisher Full Text](#)
43. Kouwijzer MLCE, Eijck BPV, Kroes SJ, et al.: **Comparison of two force fields by molecular dynamics simulations of glucose crystals: Effect of using ewald sums.** *J Comput Chem.* 1993; **14**: 1281–1289.  
[Publisher Full Text](#)
44. Hagler AT, Lifson S, Dauber P: **Consistent Force Field Studies of Intermolecular Forces in Hydrogen-Bonded Crystals. 2. A Bench mark for the Objective Comparison of Alternative Force Fields.** *J Am Chem Soc.* 1979; **101**: 5122–5130.  
[Publisher Full Text](#)
45. Skelton AA, Fenter P, Kubicki JD, et al.: **Simulations of the Quartz(1011)/Water Interface: A Comparison of Classical Force Fields Ab Initio Molecular Dynamics, and X-Ray Reflectivity Experiments.** *J Phys Chem C.* 2011; **115**: 2076–2088.  
[Publisher Full Text](#)
46. Vassetti D, Pagliai M, Proccaci P: **Assessment of GAFF2 and**

- OPLS-AA General Force Fields in Combination with the Water Models TIP3P, SPCE, and OPC3 for the Solvation Free Energy of Druglike Organic Molecules. *J Chem Theory Comput.* 2019; **15**(3): 1983–1995.  
[PubMed Abstract](#) | [Publisher Full Text](#)
47. Mobley DL, Dumont E, Chodera JD: **Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent.** *J Phys Chem B.* 2007; **111**(9): 2242–2254.  
[PubMed Abstract](#) | [Publisher Full Text](#)
48. Sellers BD, James NC, Gobbi AA: **Comparison of Quantum and Molecular Mechanical Methods to Estimate Strain Energy in Druglike Fragments.** *J Chem Inf Model.* 2017; **57**(6): 1265–1275.  
[PubMed Abstract](#) | [Publisher Full Text](#)
49. Zhu S: **Validation of the Generalized Force Fields GAFF, CGenFF, OPLS-AA, and PRODRGFF by Testing Against Experimental Osmotic Coefficient Data for Small Drug-Like Molecules.** *J Chem Inf Model.* 2019; **59**(10): 4239–4247.  
[PubMed Abstract](#) | [Publisher Full Text](#)
50. Kamath G, Kurnikov I, Fain B, et al.: **Prediction of cyclohexane-water distribution coefficient for SAMPL5 drug-like compounds with the QMPFF3 and ARROW polarizable force fields.** *J Comput Aided Mol Des.* 2016; **30**(11): 977–988.  
[PubMed Abstract](#) | [Publisher Full Text](#)
51. Fan S, Iorga BI, Beckstein O: **Prediction of octanol-water partition coefficients for the SAMPL6-[Formula: see text] molecules using molecular dynamics simulations with OPLS-AA, AMBER and CHARMM force fields.** *J Comput Aided Mol Des.* 2020; **34**(5): 543–560.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
52. Řezáč J, Bím D, Gutten O, et al.: **Toward Accurate Conformational Energies of Smaller Peptides and Medium-Sized Macrocycles: MCONF196 Benchmark Energy Data Set.** *J Chem Theory Comput.* 2018; **14**(3): 1254–1266.  
[PubMed Abstract](#) | [Publisher Full Text](#)
53. Kesharwani MK, Karton A, Martin JML: **Benchmark *ab Initio* Conformational Energies for the Proteinogenic Amino Acids through Explicitly Correlated Methods. Assessment of Density Functional Methods.** *J Chem Theory Comput.* 2016; **12**(1): 444–454.  
[PubMed Abstract](#) | [Publisher Full Text](#)
54. Kanal IY, Keith JA, Hutchison GR: **A Sobering Assessment of Small-Molecule Force Field Methods for Low Energy Conformer Predictions.** *Int J Quantum Chem.* 2018; **118**(5): e25512.  
[Publisher Full Text](#)
55. White BR, Wagner CR, Truhlar DG, et al.: **Molecular Modeling of Geometries, Charge Distributions, and Binding Energies of Small, Drug-Like Molecules Containing Nitrogen Heterocycles and Exocyclic Amino Groups in the Gas Phase and Aqueous Solution.** *J Chem Theory Comput.* 2008; **4**(10): 1718–1732.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
56. Halgren TA: **MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries.** *J Comput Chem.* 1999; **20**(5): 730–748.  
[Publisher Full Text](#)
57. Hornak V: **Comparison of robustness and quality of three small molecule force-fields.** Proceedings of CUP XI, Santa Fe, NM, USA. 2010.  
[Reference Source](#)
58. Wang J, Wolf RM, Caldwell JW, et al.: **Development and Testing of a General Amber Force Field.** *J Comput Chem.* 2004; **25**(9): 1157–1174.  
[PubMed Abstract](#) | [Publisher Full Text](#)
59. Case DA, Betz RM, Cerutti DS, et al.: **Kollman, AMBER 2016.** University of California, San Francisco. 2016.
60. Halgren TA: **Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94.** *J Comput Chem.* 1996; **17**(5–6): 490–519.  
[Publisher Full Text](#)
61. Halgren TA: **Merck Molecular Force Field. II. MMFF94 van Der Waals and Electrostatic Parameters for Intermolecular Interactions.** *J Comput Chem.* 1996; **17**(5–6): 520–522.  
[Publisher Full Text](#)
62. Halgren TA: **Merck Molecular Force Field. III. Molecular Geometries and Vibrational Frequencies for MMFF94.** *J Comput Chem.* 1996; **17**(5–6): 553–586.  
[Publisher Full Text](#)
63. Halgren TA, Nachbar RB: **Merck Molecular Force Field. IV. Conformational Energies and Geometries for MMFF94.** *J Comput Chem.* 1996; **17**: 587–615.  
[Publisher Full Text](#)
64. Halgren TA: **Merck Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, Additional Computational Data, and Empirical Rules.** *J Comput Chem.* 1996; **17**: 616–641.  
[Publisher Full Text](#)
65. Halgren TA: **MMFF VI. MMFF94s Option for Energy Minimization Studies.** *J Comput Chem.* 1999; **20**(7): 720–729.  
[Publisher Full Text](#)
66. Roos K, Wu C, Damm W, et al.: **OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules.** *J Chem Theory Comput.* 2019; **15**(3): 1863–1874.  
[PubMed Abstract](#) | [Publisher Full Text](#)
67. Mobley DL, Bannan CC, Wagner JR, et al.: **openforcefield/smirnoff99Frosst: Version 1.1.0.** 2019.  
[Publisher Full Text](#)
68. Wagner JR: **openforcefield/openforcefields: Version 1.0.0 “Parsley”.** 2019.  
[Publisher Full Text](#)
69. Smith D, Altarawy D, Burns L, et al.: **The MolSSI QCArchive Project: An Open-Source Platform to Compute, Organize, and Share Quantum Chemistry Data.** 2020.  
[Publisher Full Text](#)
70. Open Force Field Initiative: **The Open Force Field 1.0 small molecule force field, our first optimized force field (codename ‘Parsley’).**  
[Reference Source](#)
71. Qiu Y, Smith DGA, Boothroyd S, et al.: **Introducing the First Optimized Open Force Field 1.0.0 (Codename ‘Parsley’).** 2019.  
[Publisher Full Text](#)
72. Maat J: **Training Dataset Selection.** 2020.
73. Mobley D: **Constructing benchmark/test sets for OpenFF quantum chemistry benchmarks.** 2019.  
[Reference Source](#)
74. Becke AD: **Density-functional Thermochemistry. III. The Role of Exact Exchange.** *J Chem Phys.* 1993; **98**: 5648–5652.  
[Publisher Full Text](#)
75. Lee C, Yang W, Parr RG: **Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density.** *Phys Rev B Condens Matter.* 1988; **37**(2): 785–789.  
[PubMed Abstract](#) | [Publisher Full Text](#)
76. Vosko SH, Wilk L, Nusair M: **Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis.** *Can J Phys.* 1980; **58**: 1200–1211.  
[Publisher Full Text](#)
77. Stephens PJ, Devlin FJ, Chabalowski CF, et al.: **Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields.** *J Phys Chem.* 1994; **98**: 11623–11627.  
[Publisher Full Text](#)
78. Godbout N, Salahub DR, Andzelm J, et al.: **Optimization of Gaussian-Type Basis Sets for Local Spin Density Functional Calculations. Part I. Boron through Neon, Optimization Technique and Validation.** *Can J Chem.* 1992; **70**(2): 560–571.  
[Publisher Full Text](#)
79. OpenEye Python Toolkits: **OpenEye Scientific Software Inc.: Santa Fe NM, USA.**
80. Jakalian A, Bush BL, Jack DB, et al.: **Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method.** *J Comput Chem.* 2000; **21**(2): 132–146.  
[Publisher Full Text](#)
81. Jakalian A, Jack DB, Bayly CI: **Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation.** *J Comput Chem.* 2002; **23**(16): 1623–41.  
[PubMed Abstract](#) | [Publisher Full Text](#)
82. Beauchamp K, Rustenburg A, Rizzi A, et al.: **OpenMolTools.**
83. Schrödinger: **Schrödinger Release 2020-1: Maestro.** 2020.  
[Reference Source](#)
84. Madhavi Sastry G, Adzhigirey M, Day T, et al.: **Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments.** *J Comput Aided Mol Des.* 2013; **27**(3): 221–34.  
[PubMed Abstract](#) | [Publisher Full Text](#)
85. Eastman P, Swails J, Chodera JD, et al.: **OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics.** *PLoS Comput Biol.* 2017; **13**(7): e1005659.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
86. Lim VT: **BenchmarkFF.** 2020.  
[Reference Source](#)
87. Schulz-Gasch T, Schäfer, Guba W, et al.: **TFD: Torsion Fingerprints as a new measure to compare small molecule conformations.** *J Chem Inf Model.* 2012; **52**(6): 1499–1512.  
[PubMed Abstract](#) | [Publisher Full Text](#)
88. Ehrman JN, Bannan CC, Lim VT, et al.: **Improving Force Fields**

- by Identifying and Characterizing Small Molecules with Parameter Inconsistencies. 2019.  
[Publisher Full Text](#)
89. Ehrman J, Lim VT, Bannan CC, *et al.*: **Improving Small Molecule Force Fields by Identifying and Characterizing Small Molecules with Inconsistent Parameters.** *ChemRxiv*. 2020.  
[Reference Source](#)
90. Lim VT, Mobley DL, Hahn DF: **MobleyLab/benchmarkff: Version 1.0.0: Used in paper.** 2020.  
<http://www.doi.org/10.5281/zenodo.4252694>
91. 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.** *J Am Chem Soc*. 1996; **118**: 11225–11236.  
[Publisher Full Text](#)
92. Dodda LS, Cabeza de Vaca I, Tirado-Rives J, *et al.*: **LigParGen Web Server: An Automatic OPLS-AA Parameter Generator for Organic Ligands.** *Nucleic Acids Res*. 2017; **45**(W1): W331–W336.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
93. Lim VT, Hahn D, Tresadern G, *et al.*: **Supporting Information: Molecular geometries and energies from quantum mechanical calculations and small molecule force field evaluations (Version 1.0.1).** *Zenodo*. 2020.  
<http://www.doi.org/10.5281/zenodo.4299200>
94. Reva BA, Finkelstein AV, Skolnick J: **What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å?** *Fold Des*. 1998; **3**(2): 141–147.  
[PubMed Abstract](#) | [Publisher Full Text](#)
95. Sargsyan K, Grauffel C, Lim C: **How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations.** *J Chem Theory Comput*. 2017; **13**(4): 1518–1524.  
[PubMed Abstract](#) | [Publisher Full Text](#)
96. Jang H: **Update on Parsley Minor Releases (Openff-1.1.0, 1.2.0).** 2020.  
[Publisher Full Text](#)
97. Stern C: **Capturing Non-Local through-Bond Effects When Fragmenting Molecules for QC Torsion Scans.** 2020.  
[Publisher Full Text](#)
98. Mobley DL: **Current Status of OpenFF and Our Near-Term Roadmap.** 2020.  
[Publisher Full Text](#)
99. Lim VT, Hahn DF, Tresadern G, *et al.*: **Molecular geometries and energies from quantum mechanical calculations and small molecule force field evaluations.** 2020.  
<http://www.doi.org/10.5281/zenodo.4247859>

# Open Peer Review

Current Peer Review Status:  

---

## Version 1

Reviewer Report 04 February 2021

<https://doi.org/10.5256/f1000research.29983.r77769>

© 2021 Sur S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Sreyoshi Sur** 

Moderna, Cambridge, MA, USA

In this paper, the authors have looked at how closely do the molecular mechanics and quantum mechanics optimized data match for small molecules in gas phase. In order to achieve this they used two parameters:

- Conformer relative energies, calculated using different forcefields and the corresponding quantum mechanically calculated energies.
- Energy minimized geometric structures evaluated using the forcefields and the QM calculations.

They provided a list of 3271 molecules with 22675 different structures. Molecules were selected based on unique chemical connectivity and whether all the molecules can be parametrized by GAFF, GAFF2, MMFF94, MMFF4S, OPLS3e, SMIRNOFF99Frosst, and OpenFF Parsley (1.1,1.2,1.3). Molecules that were part of training set for the forcefield OpenFF Parsley was removed for evaluation too. For every evaluated structure, first the authors compared each molecules conformers QM calculated and the MM energies. Their results show that for a greater number of molecules OPLS 3e evaluated energies were closest to the QM energies, followed by OpenFF 1.2. They also evaluated MM energy for the conformers which most closely matched the energy minimized QM structures. In this case also OPLS3e and OpenFF 1.2 performed better than the rest of the molecules. To compare the 3D structures, two metrics were used, RMSD and Torsion fingerprint deviation. Using these two metrics very slight changes were observed in terms of the relative energies and the two forcefields which match closest to QM remain the same (OPLS 3e and OpenFF 1.2).

Comments: The paper addresses most of the concerns in terms of comparing small molecule's structure and energies w.r.t different MM forcefields. The authors have provided their GitHub repository consisting of all the test data and their code which was used to obtain the results. This enables the reader to reproduce the results of the paper and also have a workflow to run some of these QM vs MM comparisons with completely new sets of molecules.

Questions/Minor revisions for the authors:

1. For Figure 7, I would suggest rearranging the plots so that they are next to each other and thus easy to compare.
2. For Figure 7, the authors can also show a difference plot between OPLS3e and OpenFF 1.2 and between OPLS3e and MMFF945, then the subtle changes between OpenFF1.2 and MMFF945 around  $\Delta E=0$  compared to OPLS3e will be more prominent.
3. For the 3271 molecules chosen, what is the overall chemical diversity of these molecules?
4. These small molecules are usually used as ligands to proteins, can the authors comment on how these improvements in the OpenFF 1.2 will impact binding affinity calculations?

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Computational Chemistry and Computational Biophysics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 22 December 2020

<https://doi.org/10.5256/f1000research.29983.r75766>

© 2020 Riniker S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Sereina Riniker 

Laboratory of Physical Chemistry, Department of Chemistry & Applied Biosciences, ETH Zürich, Zürich, Switzerland

**Summary:**

The authors describe the benchmarking of different general force fields for small organic molecules against QM reference data in terms of gas-phase geometries and energies. A large dataset of 3271 molecules with 22675 conformers was assembled and is freely available. Although the force fields are typically aimed for condensed-phase systems and the dataset provides gas-phase data, such a benchmarking set and the accompanying analysis is highly valuable for the force-field community. Python code is also available on Github. From the results, it is encouraging to see that systematic improvement of force fields is possible. Minor revisions are suggested below to further improve the clarity and quality of the article.

**Suggested Revisions:**

- Dataset: The authors had to remove the 419 molecules used for training of OpenFF-1.2 from the QCArchive dataset to ensure an unbiased test set. Did this lead to a certain enrichment or depletion of particular functional groups in the remaining test set?
- QM reference data: How much variation in the QM ddE would you expect for different functionals/basis sets? In other words, what "error" (or uncertainty) is associated with the QM calculations? An indication of the variation between different QM methods is given in Faraday Discuss., 195, 497-520 (2016).
- The authors compare the ddE between the QM and MD approaches. However, can we expect the deviations to be systematic?
- Table 1: It would be interesting to see the min and max ddE for GAFF after the removal of the six molecules with collapsing hydroxyl groups (maybe in parentheses after the original values?).
- Fig. 3 (and Fig. S1): I think there should be a sufficiently large number of data points to choose a smaller bin size for the histograms.
- Fig. 7: It appears that there is a straight line around ddE = 0, i.e. conformers with a large variety in TFD all have a ddE of zero. Is there an explanation for this?
- The TFD does not give equal weights to all torsions in a molecule: (1) all bonds in rings are combined to a single value in the torsion fingerprint, and (2) torsions in the centre of the molecule are weighted more than terminal torsion angles. In the calculation of ddE, on the other hand, all deviations contribute similarly to the energy difference. Could this be one of the explanations that there is basically no correlation between higher ddE values and higher TFD values? I would suggest to consider a torsional-angle RMSD as an alternative measure for the structural differences.
- How important is the targeted torsion fitting in OPLS3e for its performance? In other words, would you expect the performance of e.g. OpenFF-1.2 to improve significantly with custom torsions?

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** I am on the Initiative Advisory Board of the Open Force Field Initiative.

**Reviewer Expertise:** Computational chemistry

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**