open
forcefield

@openforcefield

www.openforcefield.org

# 4th open force field workshop

June 11, 2021 | Virtual meeting

# Speakers

Simon Boothroyd

Jeff Wagner

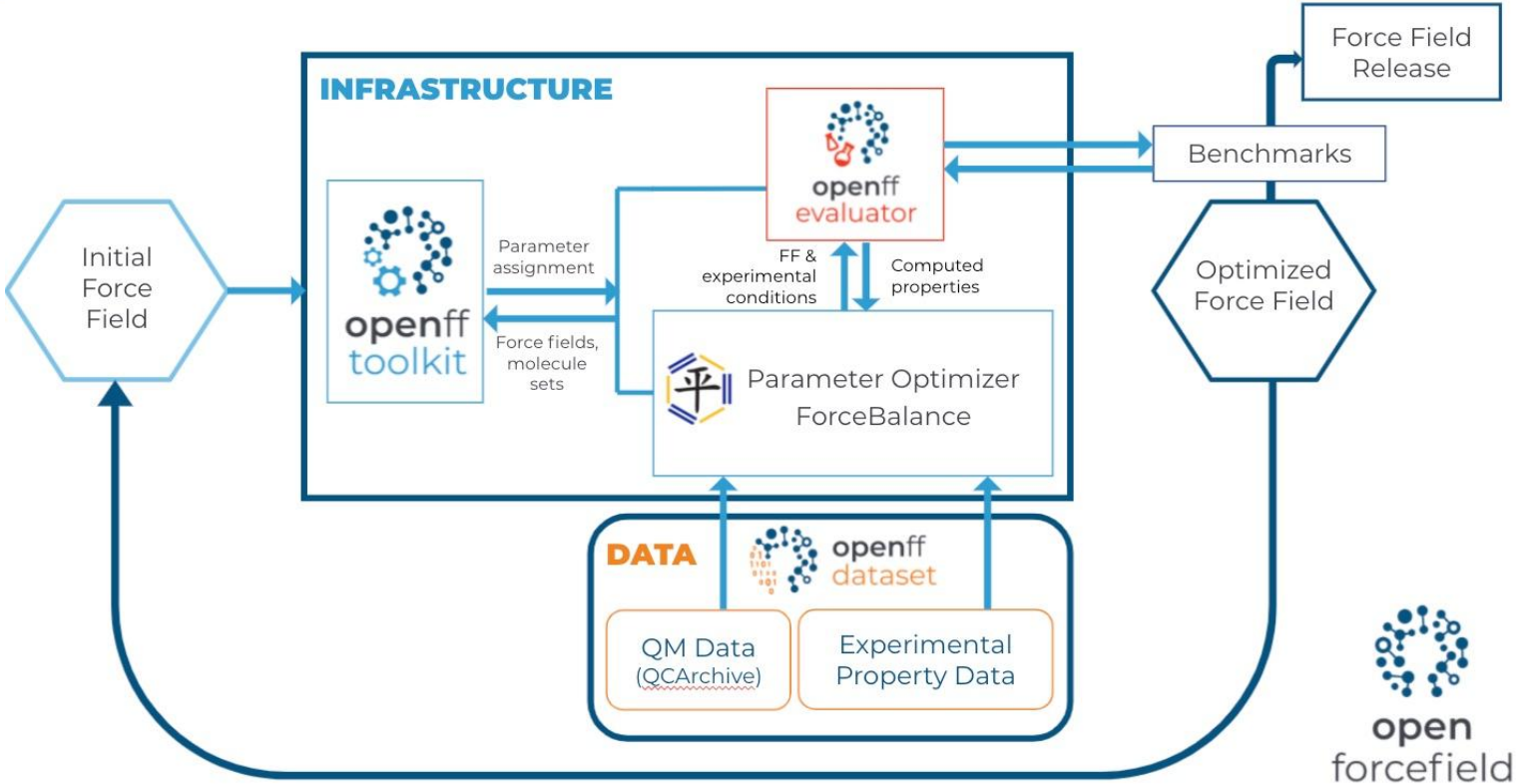
David Mobley

# Workshop Format

## Prioritizing your time

- 1 hour keynote and 1 hour Q&A today
- As-desired focus sessions in 2-3 weeks
    - Will poll for topics and times during Q&A

# We've come a long way since our late 2018 start



# Since Parsley (OpenFF 1.0) we've had a string of releases

## A series of fixes and improvements leading up to Sage (2.0)

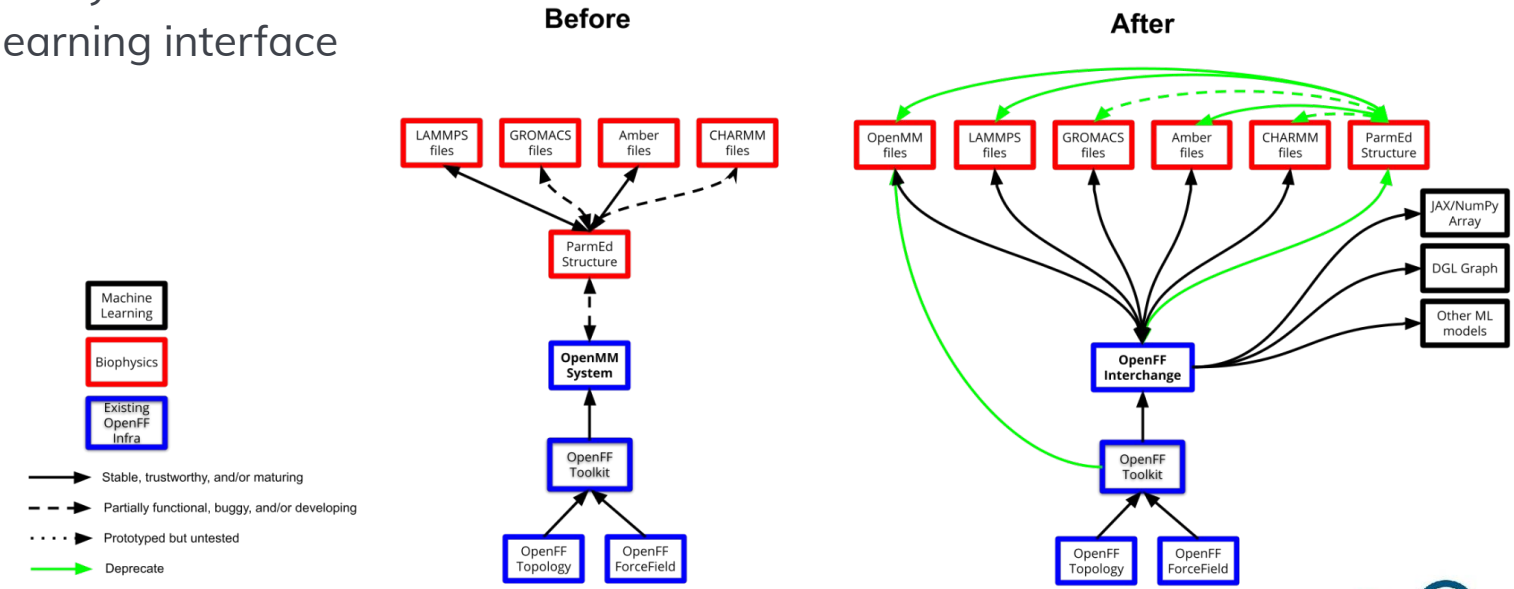
- 1.1: More valence parameter refits and some fixes
- 1.2: Expanded/redesigned QM dataset dramatically improved accuracy
- 1.2.1: Bugfix for propynes/HMR
- 1.3.0: Addresses some amide issues
- 1.3.1: Bugfix for sulfonamide geometries

**In parallel, tons of new science** -- tens to hundreds of fitting experiments to test out a wide range of ideas, e.g. effect of vibrational frequency fitting

# Infrastructure remains a major focus

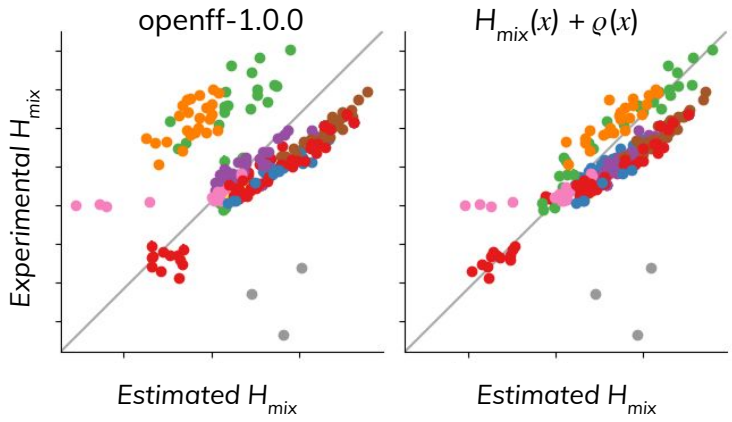
## Our new interchange object should be online soon

- Interoperability
- Machine learning interface



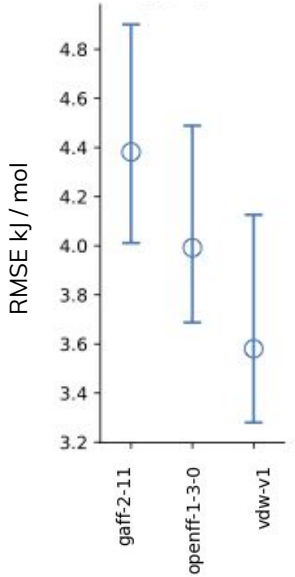
# Our Sage release candidate does better on mixture and condensed phase properties

## PREVIOUS STUDY SHOWS FITTING TO $H_{MIX}$ RESOLVES SYSTEMATIC ERRORS

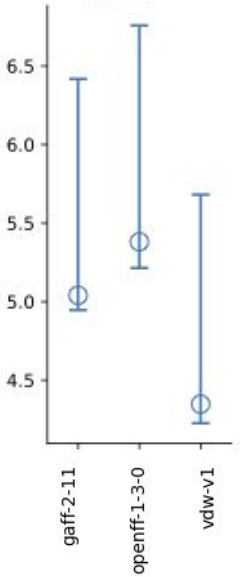



openff evaluator

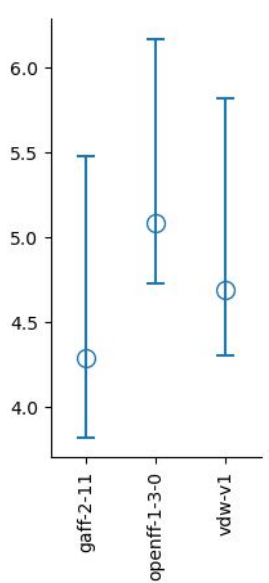
Non- aq. $\Delta G_{solv}$




aq. $\Delta G_{solv}$

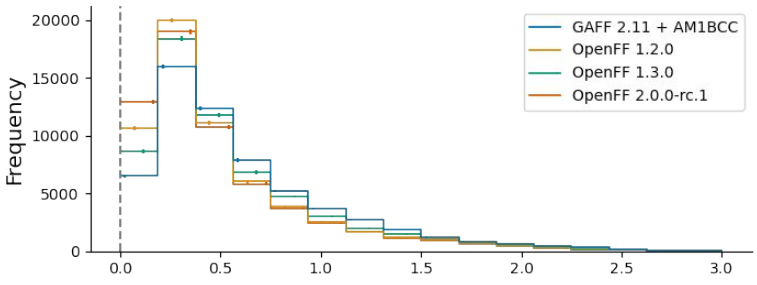


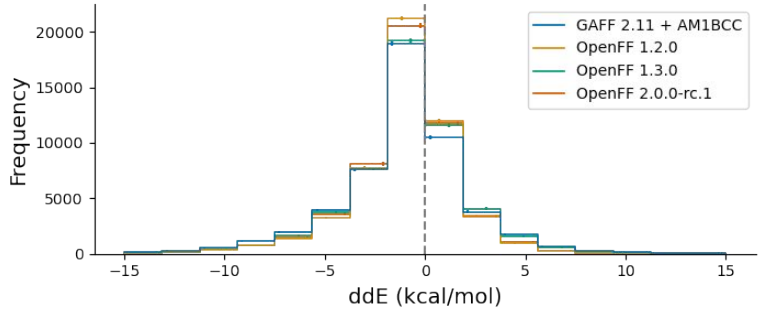

Non-aq. -> aq. $\Delta G_{trans}$

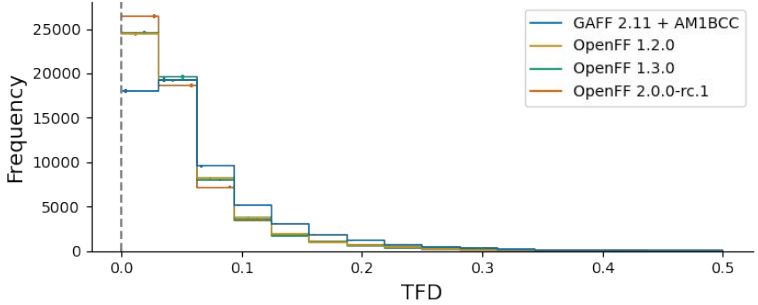


# Automated benchmarking with industry has been great, indicates progress relative to QM

- **SAGE 2.0.0-rc.1** showed excellent performance when benchmarked against the **Public** OpenFF Industry Benchmark Season 1 v1.0



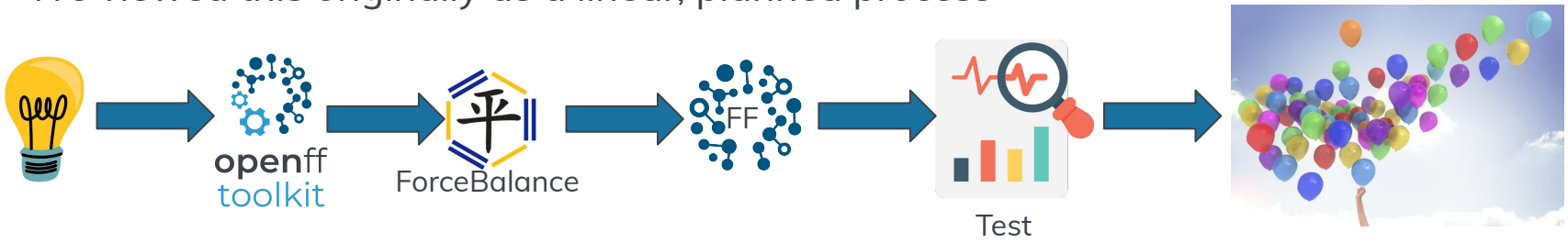



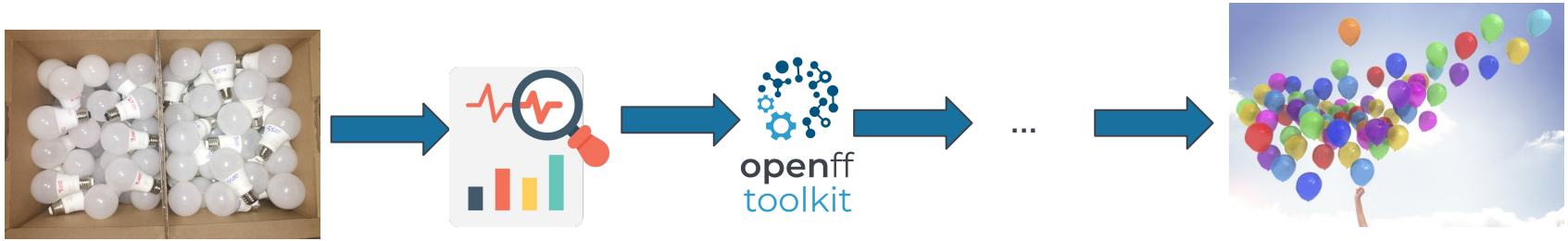

* 95% bootstrapped confidence intervals

# The process works, and it's changing how we think about building force fields

- We viewed this originally as a linear, planned process




- We're learning it works better as a parallel process where the best ideas may be unexpected and apparent only later



Making it easy to USE force fields

Making it easy to COMPARE force fields

Making it easy to CREATE force fields

## Making it easy to USE force fields

Making it easy to COMPARE force fields

Making it easy to CREATE force fields

# Infrastructure

Making it easy to **USE** force fields

## OpenFF-Toolkit

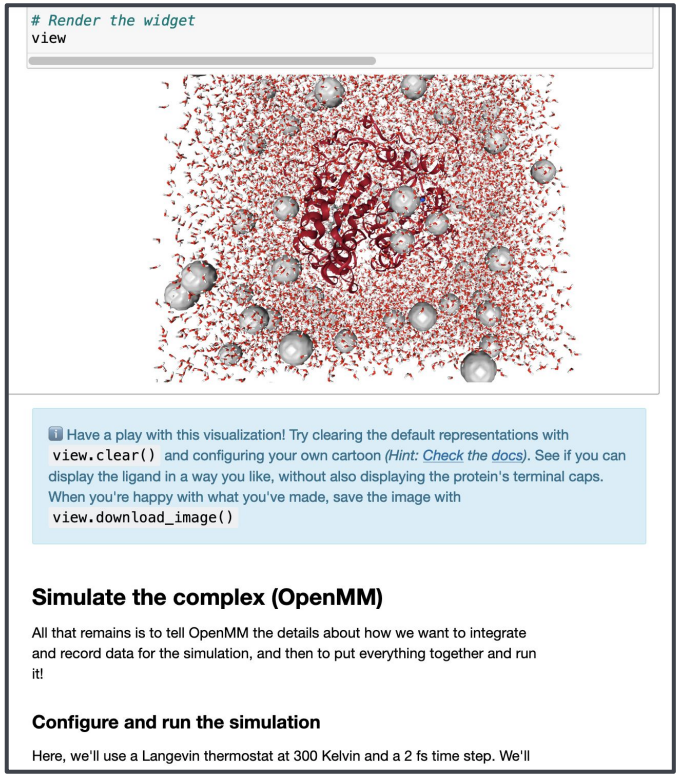
- Documentation and example refresh
- `openff-toolkit-base` package
- Continuous bugfixes+improvements
- 1 major API-breaking release
- Biopolymer update coming soon!

**Old** (0.8.3 and earlier)

```
from openforcefield.topology import Molecule
from openforcefield.typing.engines.smirnoff import ForceField
```

**New** (0.9.0 and newer)

```
from openff.toolkit.topology import Molecule
from openff.toolkit.typing.engines.smirnoff import ForceField
```




https://github.com/openforcefield/openff-toolkit/tree/master/examples/toolkit_showcase
https://github.com/openforcefield/openff-toolkit/tree/master/examples/forcefield_modification




## Q1 2021: OpenFF and OpenMM migrate to conda-forge

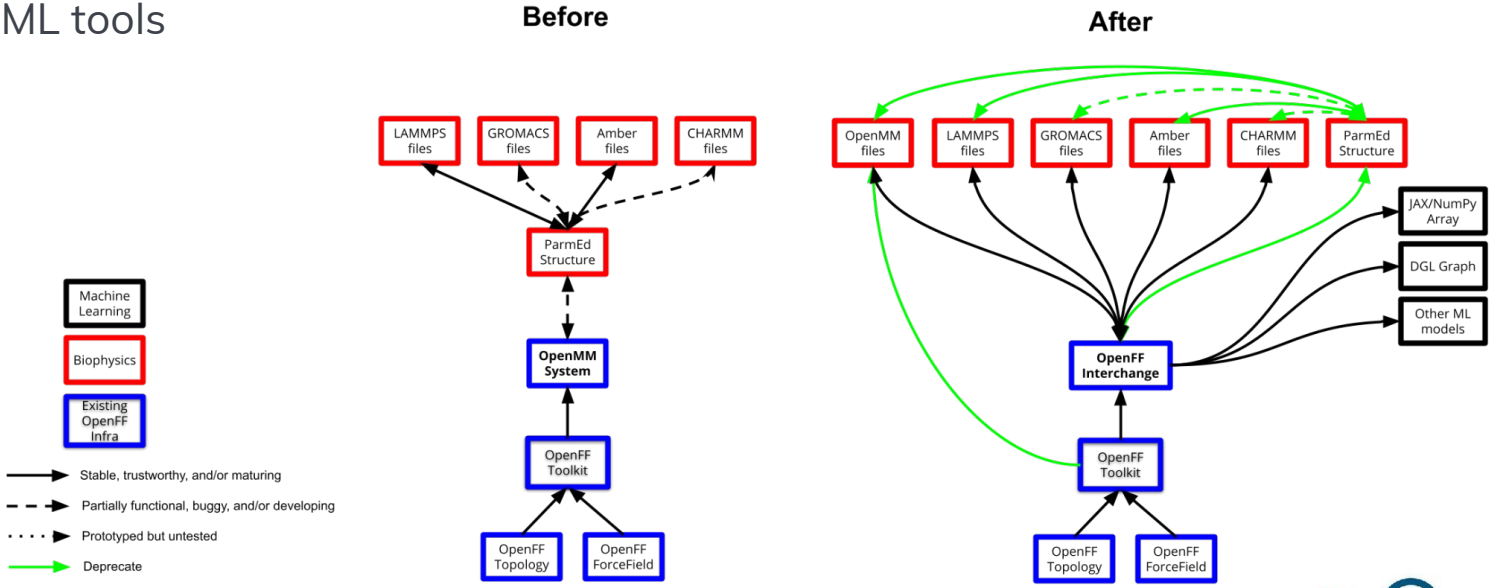
- Broader interoperability
- Better low-level library interfaces
- Better dependency resolution
- No more maintaining Omnia!

conda-forge / packages / **openff-toolkit** 0.9.2

License: MIT
Home: https://openforcefield.org/
</> Development: https://github.com/openforcefield/openff-toolkit
Documentation: https://open-forcefield-toolkit.readthedocs.io/
13001 total downloads
Last upload: 1 month and 9 days ago

## New Interchange object

- A “ParmEd replacement”
- Post-assignment parameter modification
- Interface with ML tools



## Use of OpenFF tools outside the Initiative

- Companies including Cresset and OpenEye
- Collaborators including MoSDeF, Rowley, and Cole labs
- Unaffiliated groups around the world
- Strangers on GitHub!

cresset
discover • create • innovate

This release expands the choice of force fields available for this type of calculations with the addition of the Open force field. As the Open FF Consortium provides frequent updates and improvements to the Open FF, we opted for a flexible implementation within Flare, enabling you to easily upgrade to the latest available version simply by dropping the related files into the appropriate Flare installation folder.

**PELE Force Field Yielder**

PELE Force Field Yielder

The `peleffy` (PELE Force Field Yielder) is a Python package that builds PELE-compatible force field templates. The current supported force fields are:

- Any force field from the Open Force Field toolkit.
- OPLS2005.
- A combination of them.

OpenEye SCIENTIFIC

Search docs

Getting started on Cube and Floe development

OpenEye Orion Platform Packages Reference

**OpenEye Orion Cubes and Floes Packages**

Chebuu/3VTE-model
exp.02/01-Assembly.ipynb

```
"from simtk.openmm.app import PDBFile, NoCutoff, HBonds\n",
"\n",
"from openforcefield.topology import Molecule\n",
"from openmmforcefields.generators import SystemGenerator\n",
"from openforcefield.typing.engines.smirnoff import ForceField\n",
```

**Release Notes**

**v3.0.0 November 2020**

**General Notice**

- OpenFF 1.3.0 and 1.2.1 support
- Bug Fixing

Molecular Simulation MoSDeF Design Framework

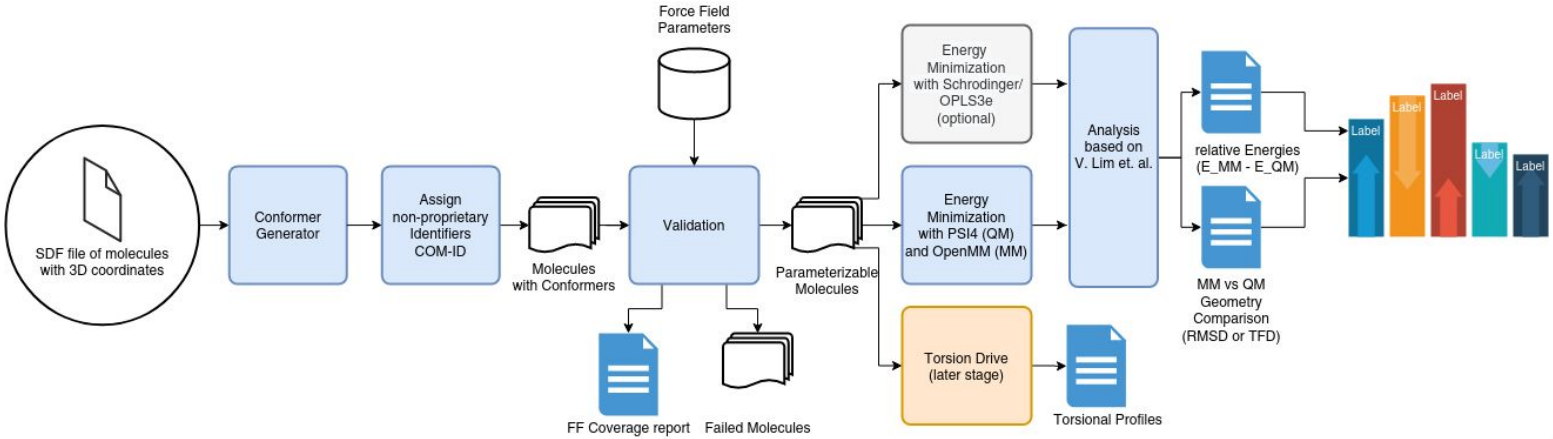
Making it easy to USE force fields

**Making it easy to COMPARE force fields**

Making it easy to CREATE force fields

## OpenFF Benchmark

- Automates running the Lim and Hahn benchmark
- CLI-first approach, using OpenFF Python backend
- Heavily pinned environments and conda installers to provide consistent results








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)

## Simulation-based properties

- Automated infrastructure for setting up and running simple physical property calculations
- Standard protein-ligand benchmark set




*The physical properties which are natively supported by the framework.*

| | Direct Simulation | | MBAR Reweighting | |
|---|---|---|---|---|
| | **Supported** | **Gradients** | **Supported** | **Gradients** |
| Density | ✓ | ✓ | ✓ | ✓ |
| Dielectric Constant | ✓ | ✓* | ✓ | ✓* |
| $\Delta H_{vaporization}$ | ✓ | ✓ | ✓ | ✓ |
| $\Delta H_{mixing}$ | ✓ | ✓ | ✓* | ✓ |
| $\Delta V_{excess}$ | ✓ | ✓ | ✓ | ✓ |
| $\Delta G_{solvation}$ | ✓ | ✓* | ✗ | ✗ |
| $\Delta G_{host\text{-}guest}$ **(beta)** | ✓* | ✗ | ✗ | ✗ |

### Protein-Ligand Benchmarks for Free Energy Calculations

**Contents:**

- Installing the Protein Ligand Benchmark Set
  - Installation from Source
- Getting Started
  - Get the whole set of targets in the dataset
  - The `Target` class
  - The `LigandSet` and `Ligand` class
  - The `EdgeSet` and `Edge` class
- Data
  - Data file tree and file description

https://github.com/openforcefield/protein-ligand-benchmark
https://github.com/openforcefield/openff-evaluator

Making it easy to USE force fields

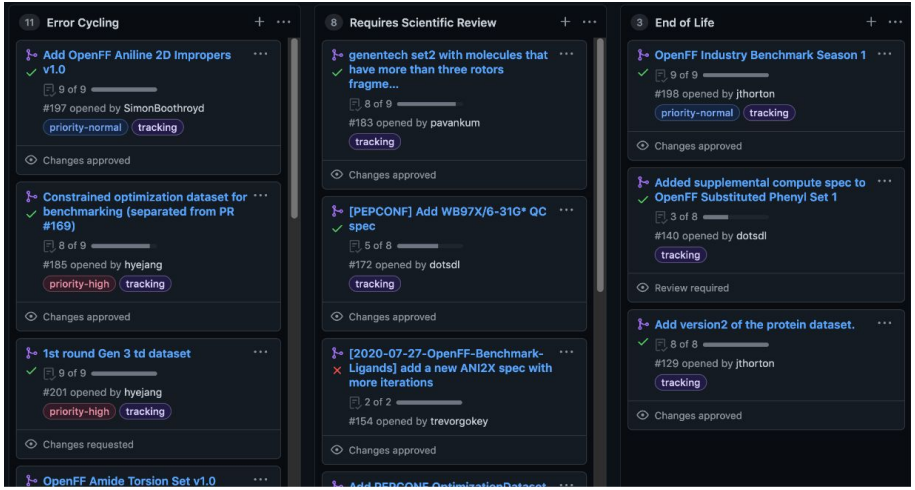
Making it easy to COMPARE force fields

**Making it easy to CREATE force fields**




## QCSubmit and QCA-Dataset-Submission

- Standards
- Automatic validation
- Priority assignment and error handling




**`TorsionDriveRecord`** current status

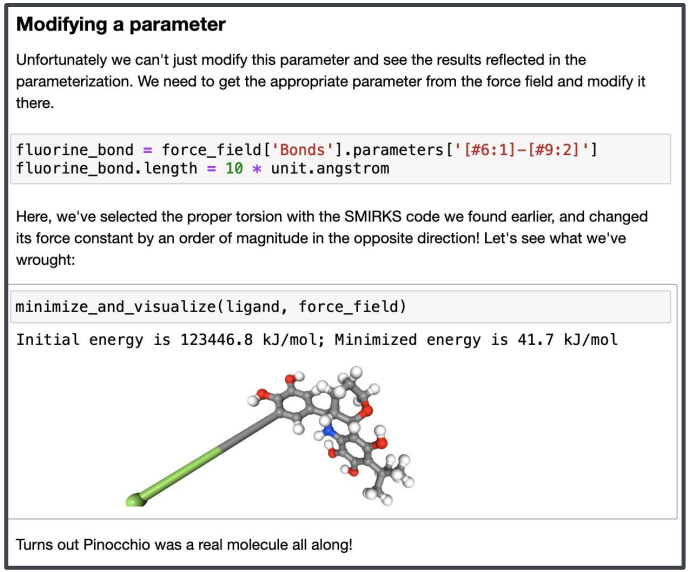
| specification | COMPLETE | ERROR | INCOMPLETE | RUNNING |
|---|---|---|---|---|
| default | 872 | 15 | 0 | 1 |

**`OptimizationRecord`** current status

| specification | COMPLETE | ERROR | INCOMPLETE |
|---|---|---|---|
| default | 55376 | 25 | 3 |

## OpenFF Toolkit

- Added support for:
  - Virtual sites
  - More partial charge methods
  - SMARTS-based charge increments
  - Wiberg bond order-interpolated torsions and bonds
  - Molecule subclasses
- Absorbed CMILES
- Can use AmberTools backend for Fragmenter




## OpenFF Forcefields

- Formerly “openforcefields”
- Release automation:
  - Canary tests for known issues prevent known instabilities, eg C-C#C
  - Parsing/formatting tests
  - Packaging

# OPEN Software, OPEN Data, OPEN Science is rapidly facilitating force field science!

## OPEN SOFTWARE

Automated infrastructure enables rapid experimentation with minimum human intervention


## OPEN DATA

Access to large, high quality experimental and quantum chemical data facilities easy curation of balanced train / test sets

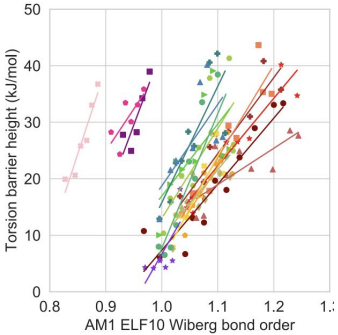
## OPEN SCIENCE

Exploring new force field science:

hypothesis - build software - train - test - iterate

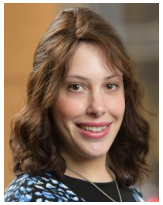
is now almost routine

# WBO Interpolation - promising force field science not quite ready for the prime time




openff toolkit

forcebalance


Chaya Stern


Jessica Maat


Pavan Behara

**Hypothesis** → **Software** → **Training** → **Testing** → **Data Driven Decision**

Wiberg Bond Order and torsion barrier height strong correlated

Add support for interpolating torsion barrier height using WBO to OpenFF toolkit

New WBO interpolated torsion parameters added and FF refits performed

Performance of the refit parameters assessed against QC data

FF performance after including new parameters - inclusion in Sage deferred

# OpenFF Sage - the next INCREMENTAL generation of OpenFF force field

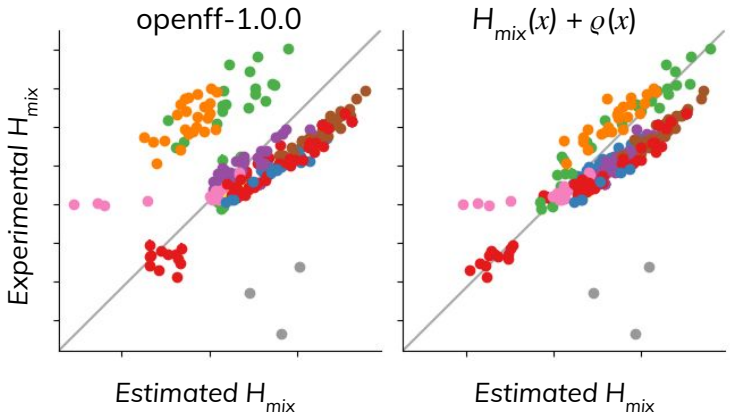
- OpenFF **Sage** commences the next generation of OpenFF force fields


| SMIRNOFF99Frosst | Parsley | **Sage** | Rosemary |
|---|---|---|---|
| Initial SMIRNOFF port of the parm99Frosst force field | Retrained **valence** parameters against a redesigned QC data set | Retrained **vdW** parameters against physical property data + retrained **valence** parameters | Self consistent biopolymer + small molecule force field |

# OpenFF Sage - select vdW parameters trained against experimental MIXTURE enthalpies and densities

- Mixture properties offer large benefits over pure properties alone
  - Easily incorporate interactions between solvent, ligands, amino acids, sugars etc
- Training set includes ~**1000** mixture enthalpy and density data points (NIST ThermoML)
  - Directly includes **aqueous** (TIP3P) mixtures
  - Small organic molecules (C, H, N, O, Cl, Br), ambient conditions, 3 concentrations
- Made possible by the **OpenFF Evaluator**

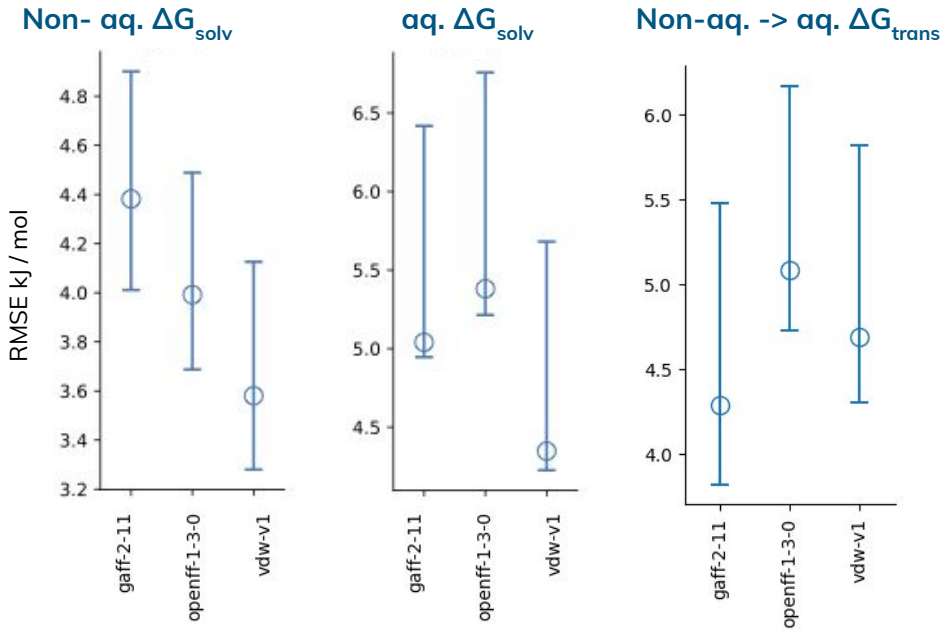
## PREVIOUS STUDY SHOWS FITTING TO $H_{MIX}$ RESOLVES SYSTEMATIC ERRORS




openff evaluator

# OpenFF Sage - positive improvements observed for solvation / transfer free energies

- Benchmarked refit vdW parameters against solvation free energies + transfer free energies
  - Subset of FreeSolv and MNSol
- Some caution required given fixed charged force field
- While outlook looks positive, biggest gains likely once electrostatics consistently refit



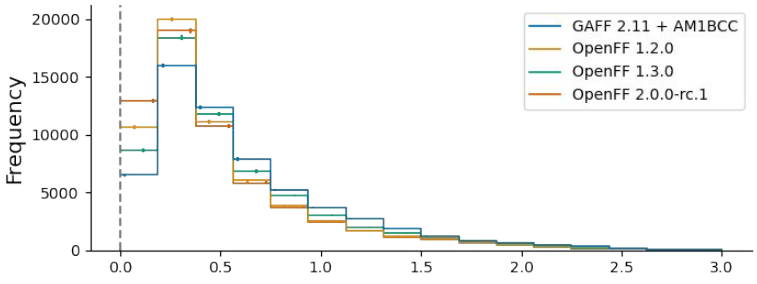
# OpenFF Sage - exploring which data best informs valence parameters

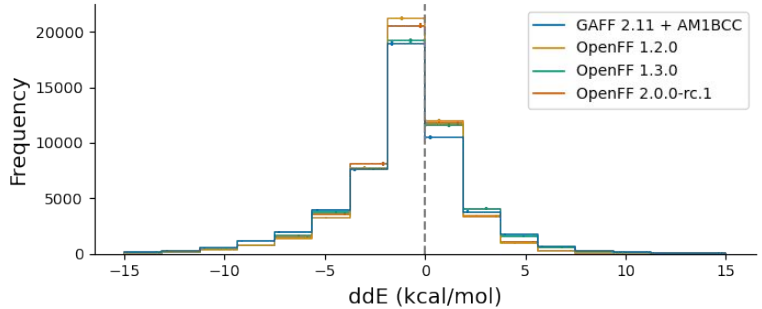
- Explored constructing torsion drive sets by combinatorially combining small, chemically diverse fragments (**Gen 3**)
  - reduce steric hindrance / electrostatic interactions while retaining a diversity of chemistry around central bond
  - ultimately **not yet included** due to poor test set performance - still **WIP**
- Revisited which data should valence terms be trained against
  - Train against *vibrational frequency* or not?
  - Use **modified seminario** to derive force constants?
- Include all available data or filter to ensure balanced training set?
  - Filter optimized geometry to retain only **distinct** conformers

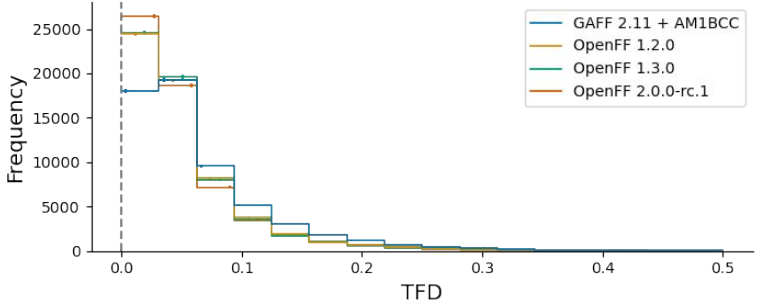
Hyesu Jang

# OpenFF Sage - without introducing new QC data we've again improved opt-geo RMSD and ddE

- SAGE **2.0.0-rc.1** showed excellent performance when benchmarked against the Public OpenFF Industry Benchmark Season 1 v1.0



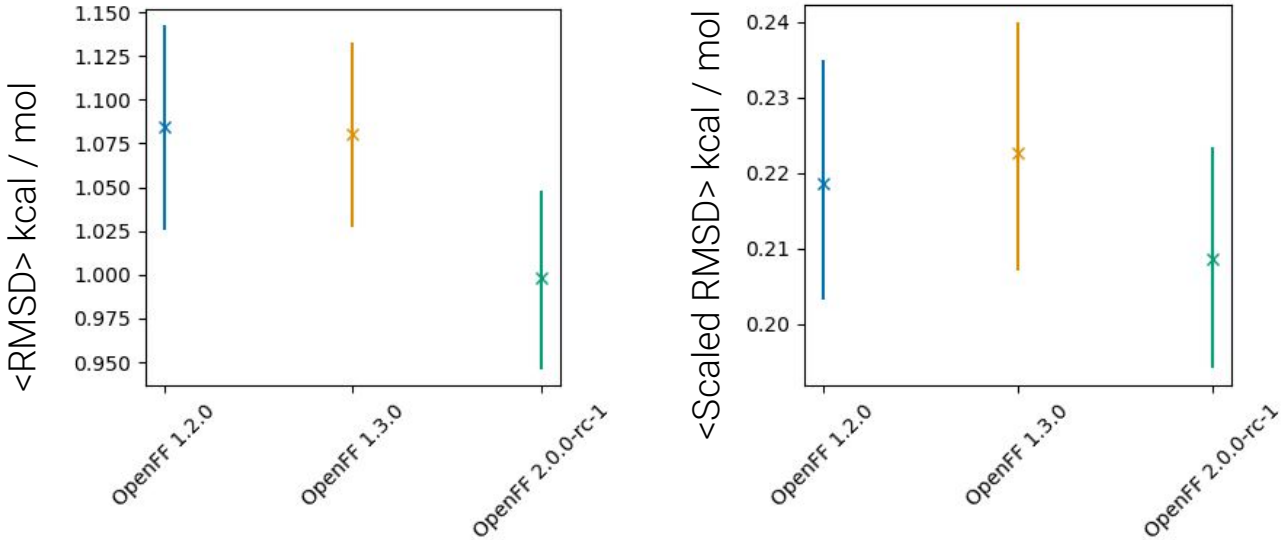




* 95% bootstrapped confidence intervals

# OpenFF Sage - benchmarked torsion profiles appear to exhibit good performance

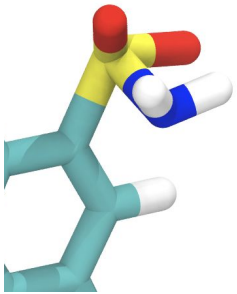
- SAGE **2.0.0-rc.1** likewise predicts well the torsion profiles of a fragmented set of JACS ligands (`OpenFF-benchmark-ligand-fragments-v1.0`)




* 95% bootstrapped confidence intervals

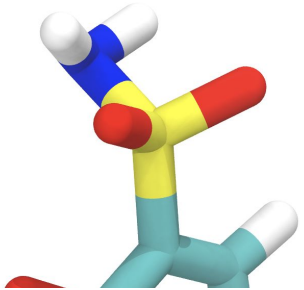
# OpenFF Sage - sulfonamide valence angle now correctly tetrahedral after valence refits

BEFORE



| | OpenFF 1.2.0 | OpenFF 1.3.0 |
|---|---|---|
| a30 | θ=104.2<br>k=201.1 | θ=99.1<br>k=220.7 |
| a31 | θ=96.3<br>k=152.3 | θ=89.8<br>k=188.4 |

AFTER



**Problem Identified** → **Diagnostics** → **Triage** → **Longer Term Solution**

Pharma partners report incorrect sulfonamide valence angle (~75 degrees) when simulating with OpenFF 1.3.0

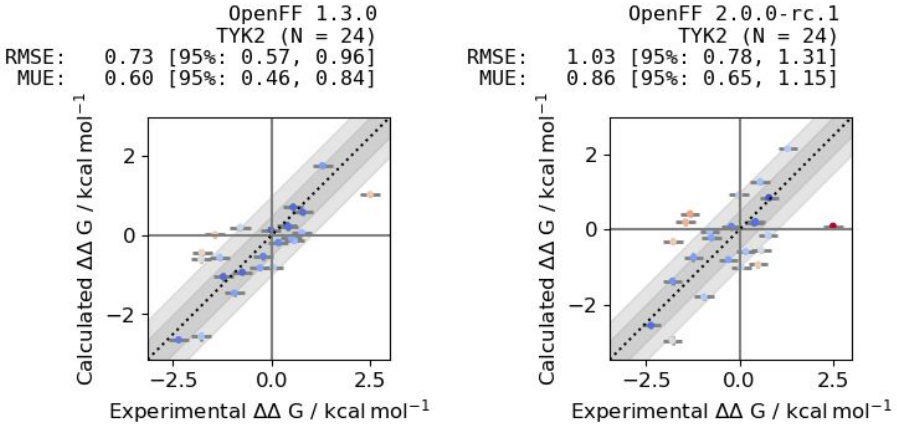
OpenFF identified sulfonamide equilibrium valence angle parameters decreased 'unphysically' between releases

Problematic parameters rolled back to previous values and OpenFF 1.3.1-alpha.1 made available for testing

Excluding vibrational frequency training data from 2.0.0-rc.1 appears to have resolved issue after refit

# OpenFF Sage - aim to routinely incorporate P-L binding free energies into force field assessment

- The relative binding free energies of the TYK2 JACS set were computed using each new force field produced during the test fits using PERSES as ‘sanity’ check
- In future hope to distribute across Folding@Home so can routinely benchmark against large P-L bind free energy data sets

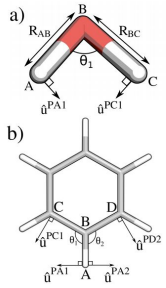


# OpenFF Sage - the next INCREMENTAL generation of OpenFF force field

- **Sage 2.0.0-rc.1** available now on GitHub in the openff-sage repository


| SMIRNOFF99Frosst | Parsley | **Sage** | Rosemary |
|---|---|---|---|
| Initial SMIRNOFF port of the parm99Frosst force field | Retrained **valence** parameters against a redesigned QC data set | Retrained **vdW** parameters against physical property data + retrained **valence** parameters | Self consistent biopolymer + small molecule force field |

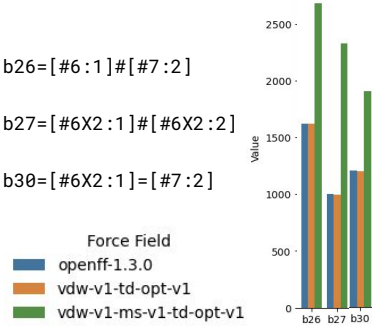
# Modified Seminario* - heavily automated fitting pipeline used to explore new fitting target in days

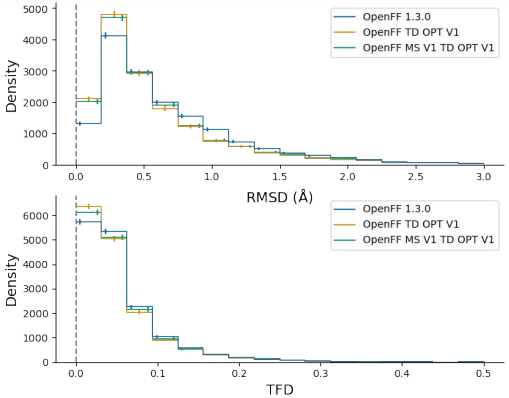



```
b26=[#6:1]#[#7:2]

b27=[#6X2:1]#[#6X2:2]

b30=[#6X2:1]=[#7:2]
```




Josh Horton

## Hypothesis

Using the modified seminario method to derive bond and angle force constants directly from QC data yields more 'physical' values

## Force Constants Computed

The Cole group retrieved all hessian data generated by OpenFF and from this computed average bond and angle force constants

## Remaining Parameters Refit

Within ~1 day OpenFF refit the rest of the valence parameters while restraining the force constants

## New Parameters Benchmarked

Within a further ~1 day the new force field had been benchmarked against the QC data

* Alice E. A. Allen, Michael C. Payne, and Daniel J. Cole 10.1021/acs.jctc.7b00785

# Refit Charge Models - AM1BCC charge model currently being re-trained against QC and exp. data

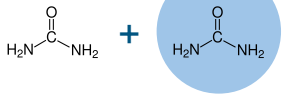
33 Double-bonded oxygen in a lactone or lactam

[#8X1$(*=[#6r]@[#7r,#8r]):1]

+

openff evaluator

ESP / EF



*RESP2* $\delta$=0.6

**Mixture enthalpies + densities**

**AM1BCC Ported to SMIRNOFF**

A majority of the original AM1BCC parameters have been ported to SMIRNOFF

→

**Integrate Into Fitting Infrastructure**

ForceBalance and the OpenFF Evaluator extended to support co-optimising against QC and exp. data

→

**Training**

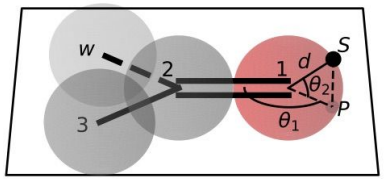
Test fits being performed against a combination of QC ESP / EF data and mixture exp. data

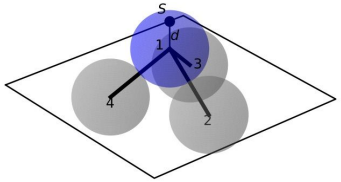
→

**Testing**

XtalPi benchmarking test fits against experimental solvation / transfer free energy data

# Virtual Sites - fitting infrastructure becoming available in weeks, followed by new science in months

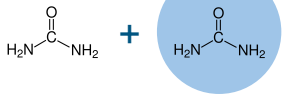



+

openff toolkit

ESP / EF




*RESP2 δ=0.6*

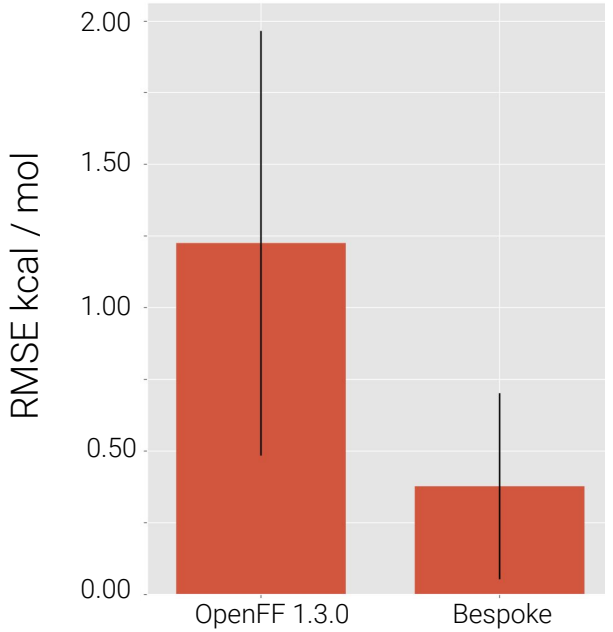
Trevor Gokey

| Hypothesis | Software | Training | Testing | Interoperability |
|---|---|---|---|---|
| The inclusion of off-site charges should improve the accuracy of a force fields electrostatic interactions | Virtual site support added to the OpenFF toolkit. Support for training to QC ESP + EF data in progress | Virtual sites will be trained against ESP / EF QC data, based on input from the Cole group | Trained parameters will be benchmarked against experimental and physical property data | In order to include into mainline force field need major simulation packages to support proposed v-sites |

# Bespoke Fitting - expect initial release of the bespoke fitting package in the next 1-2 months

- Can retrain torsion parameters to bespoke torsion scans generated for ‘fragments’ of original molecule
- Working to expand to other valence terms
  - force constants from **modified seminario method**
- Investigating the use of **ANI2x** models to rapidly generate torsion scan training data

TORSION PROFILE BEFORE AND AFTER FITTING THE JACS LIGAND FRAGMENTS

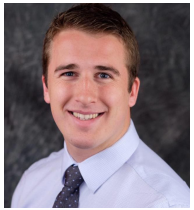



* error bars show 1 std deviation

Josh Horton

# AND MUCH MORE

Trevor Gokey



Owen Madin

Hyesu Jang

## AUTOMATED CHEMICAL PERCEPTION

When and where do we need new parameters

## SURROGATE MODELS FOR RAPID vdW FITTING

Learning a surrogate model for rapid physical properties estimates

## INTERNAL COORDINATE HESSIAN FITTING

A more robust approach for fitting against hessians

# AND MUCH MORE

Jeff Setiadi

Jessica Maat

Pavan Behara

## TRAIN TO HOST-GUEST BINDING AFFINITY

Can we use H-G binding affinities as surrogates for P-L?

## WBO INTERPOLATED IMPROPER TORSIONS

Reduced noise from steric and electrostatic interactions

## QM THEORY BENCHMARK

Which level of theory should we train / test against?

# OPEN Software, OPEN Data, OPEN Science is rapidly facilitating force field science!

## OPEN SOFTWARE

Automated infrastructure enables rapid experimentation with minimum human intervention

## OPEN DATA

Access to large, high quality experimental and quantum chemical data facilities easy curation of balanced train / test sets

## OPEN SCIENCE

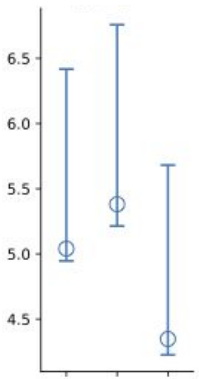
Exploring new force field science:

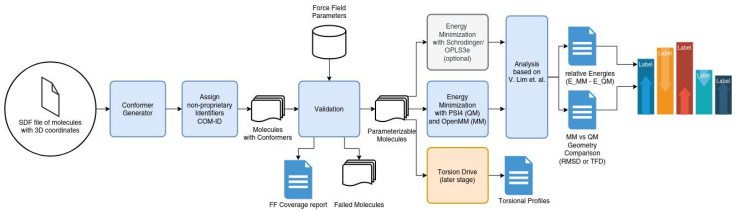
hypothesis - build software - train - test - iterate

is now almost routine

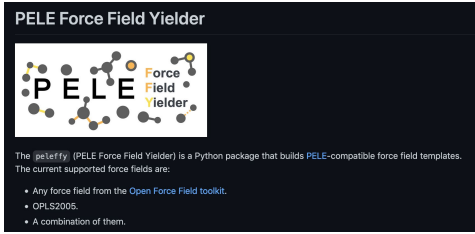
# Conclusions

Sage candidate looks promising on diverse data

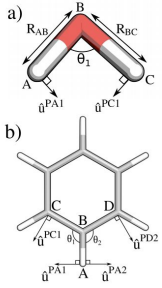


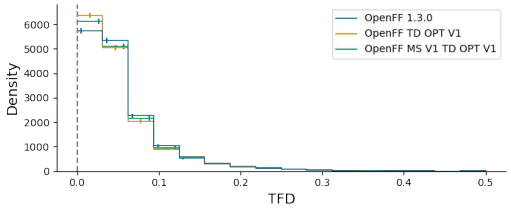




Automated benchmarking has been a major focus and will point the way forward

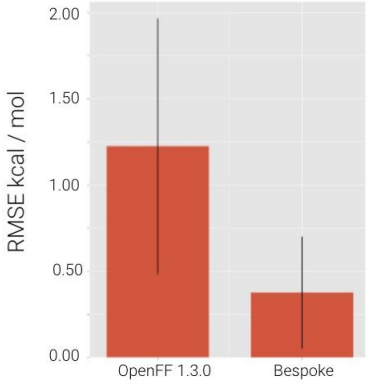



We see community uptake, with and without our help




New technology yields better starting points and better FFs




Bespoke torsion fitting will be ready soon and yields accuracy gains

# Acknowledgements

https://openforcefield.org/about/organization/

Too many other folks over the years to list, including the AMBER FF community, the GAFF/GAFF2 developers, etc.

NIH and NSF for funding work that helped pave the way to this effort

Consortium and NIH for current funding, plus MolSSI and others for fellowship funding

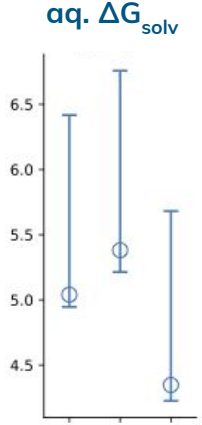
# Potential Follow-up Workshops

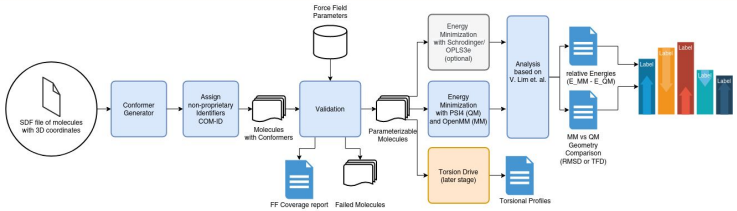
- **Bespoke fitting**
  - This presentation session would focus on the generation of bespoke parameters for a molecule using the soon-to-be-released BespokeFit package, and analysis of the performance of the resulting force fields.
- **FF optimization and debugging/Chemical perception**
  - This presentation session would recap our experiences with force field optimization and debugging in greater detail, and discuss current and future efforts to automate generating new parameters.
- **Benchmarking**
  - This presentation session would review current benchmarking infrastructure and preliminary results with respect to (vote on each of the following:)
    - conformer energetics
    - physical properties
    - protein-ligand free energies
- **QC interface notebook**
  - This demo session would cover how to use QCSubmit and other tools to navigate, submit, and retrieve data from OpenFF's datasets on QCArchive. This would be accompanied by a notebook and instructions on how to install the QCSubmit package locally.

Vote here! https://forms.gle/mEyK2Mzq8ea8kRfT9

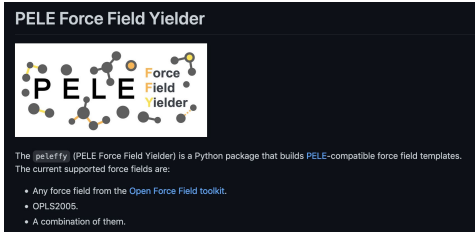
# Questions?

Sage candidate looks promising on diverse data

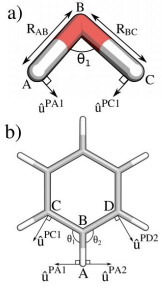


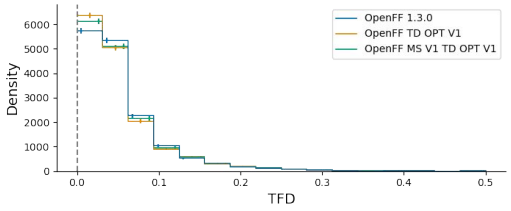




Automated benchmarking has been a major focus and will point the way forward

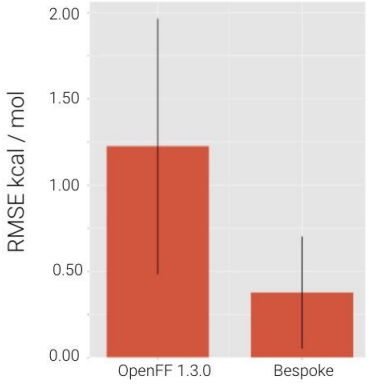



We see community uptake, with and without our help

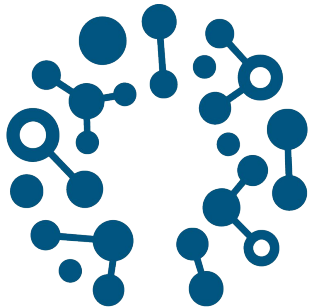



New technology yields better starting points and better FFs




Bespoke torsion fitting will be ready soon and yields accuracy gains



open
forcefield

**Website**
https://openforcefield.org/

**GitHub**
https://github.com/openforcefield/

**Zenodo**
https://zenodo.org/communities/openforcefield/

**Twitter**
https://twitter.com/openforcefield

**YouTube**
https://www.youtube.com/channel/UCh0qJSUm_sYr7nuTzhW806g/videos

**LinkedIn**
https://www.linkedin.com/company/openforcefield

**Email**
info@openforcefield.org