open
forcefield

@openforcefield

www.openforcefield.org

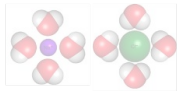
The Open Force Field Initiative: An Open Source, Open Science Approach to Better Biomolecular Force Fields

16 June 2022 | John Chodera / Andy Vinter Memorial Meeting

Molecular mechanics force fields have traditionally been products of heroic human effort

proteins

post-translational modifications



water

ions

small molecules

nucleic acids

lipids

carbohydrates

Amber20 recommendations

J. A. Maier; C. Martinez; K. Kasavajhala; L. Wickstrom; K. E. Hauser; C. Simmerling. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.*, 2015, *11*, [illegible]

[illegible] T. Fox; J. W. Caldwell; P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **1995**, *117*, 5179–5197.

N. Homeyer; A. H. C. Horn; H. Lanig; H. Sticht. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *J. Mol. Model.*, **2006**, *12*, 281–289.

H. W. Horn; W. C. Swope; J. W. Pitera; J. D. Madura; T. J. Dick; G. L. Hura; T. Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.*, **2004**, *120*, 9665–9678.

I. S. Joung; T. E. Cheatham, III. Molecular dynamics simulations of the dynamic and energetic properties [illegible] *Chem. B*, **2009**, *113*, 13279–[illegible]

P. Li; B. P. Roberts; D. K. Chakravorty; K. M. Merz, Jr. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.*, **2013**, *9*, [illegible]

[illegible] Development and testing of a general Amber force field. *J. Comput. Chem.*, **2004**, *25*, 1157–1174.

R. Galindo-Murillo; J. C. Robertson; M. Zgarbovic; J. Sponer; M. Otyepka; P. Jureska; T. E. Cheatham. [illegible] *J. Chem. Theory Comput.*, **2016**, [illegible]

A. Perez; I. Marchan; D. Svozil; J. Sponer; T. E. Cheatham; C. A. Laughton; M. Orozco. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of alpha/gamma Conformers. *Biophys. [illegible]*

[illegible] Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.*, **2011**, *7*, 2886–2902.

Å. Skjevik; B. D. Madej; R. C. Walker; K. Teigen. Lipid11: A modular framework for lipid simulations using amber. *J. Phys. Chem. B*, **2012**, *116*, 11124–11136.

C. J. Dickson; B. D. Madej; Å. A. Skjevik; R. M. Betz; K. Teigen; I. R. Gould; R. C. Walker. Lipid14: The [illegible]

K. N. Kirschner; A. B. Yongye; S. M. Tschampel; J. González-Outeiriño; C. R. Daniels; B. L. Foley; R. J. Woods. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.*, **2008**, *29*, 622–655.

Quickly adds up to >100 human-years

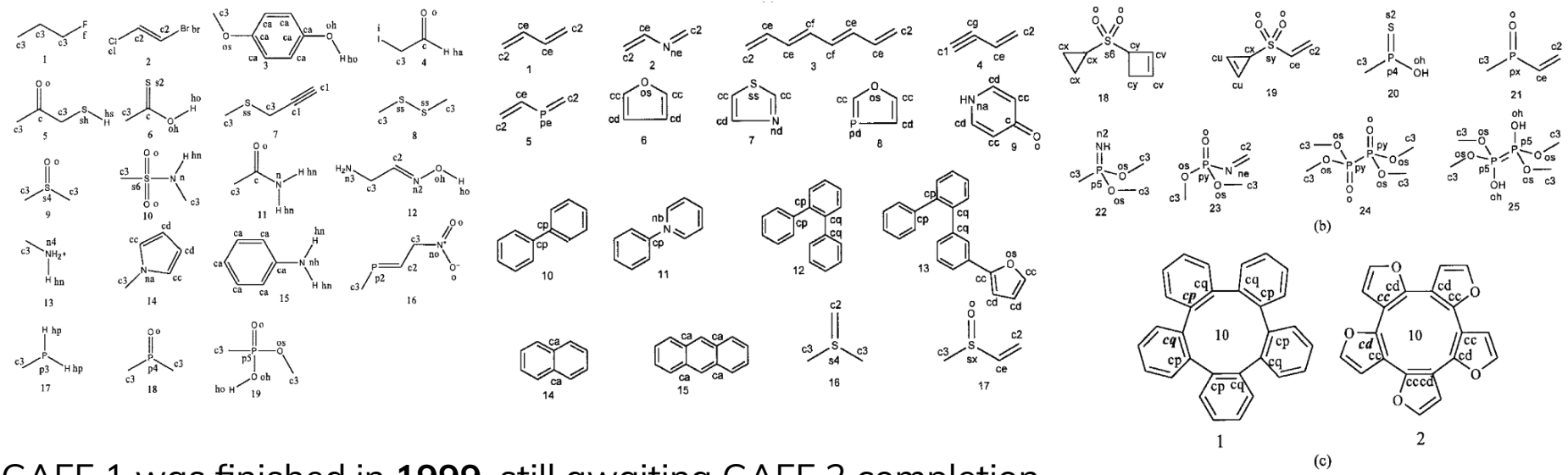
Intended to be compatible, but not co-parameterized

Significant effort is required to extend to new areas (e.g. covalent inhibitors, bio-inspired polymers, etc.)

Nobody is going to want to refit this based on some new data

How can we bring this problem into the modern era?

As drug discovery explores new parts of chemical space, how can force fields keep up?



GAFF 1 was finished in **1999**, still awaiting GAFF 2 completion
Extension to new chemical space is **nontrivial**
Parameter fitting code was **never released**
Atom types have introduced numerous **errors**

How can we bring force field science into the modern era?

Open source Python Toolkit: modern infrastructure for building/using force fields

Open curated QM / physical property datasets: public data to build on

Open infrastructure: Extend our tools; run your own benchmarks

Open science: Everything done in the open; everyone can get involved

The Open Force Field Consortium

INDUSTRY

AbbVie

BASF

Bayer

Boehringer-Ingelheim

Bristol Myers Squibb

Cresset

Janssen

GlaxoSmithKline

Merck KGaA

OpenEye

Pfizer

Roche

Vertex

Eli Lilly

... and others

Open Molecular Software Foundation

PROJECT STAFF

Jeff Wagner
Technical Lead

Lily Wang
Science Lead

...

ACADEMIC

John Chodera (MSKCC)

Michael Gilson (UC San Diego)

David Mobley (UC Irvine)

Michael Shirts (CU Boulder)

Plus affiliates:

- **Danny Cole** (Newcastle)
- **Lee-Ping Wang** (UCD)
- **Dennis Della Corte** (BYU)
- **MolSSI** (Virginia Tech)

What is the Open Force Field Initiative producing?

Toolkits: Modern toolkits for rapid development, application, and evaluation of force fields

Parameters: Parameterized datasets for different model resolutions

Datasets: Curated collections of physical property measurements

Community: Bringing together top developers & users and working together to solve our problems in the open so everyone benefits

Best Practices: Measurement and calculation of physical properties

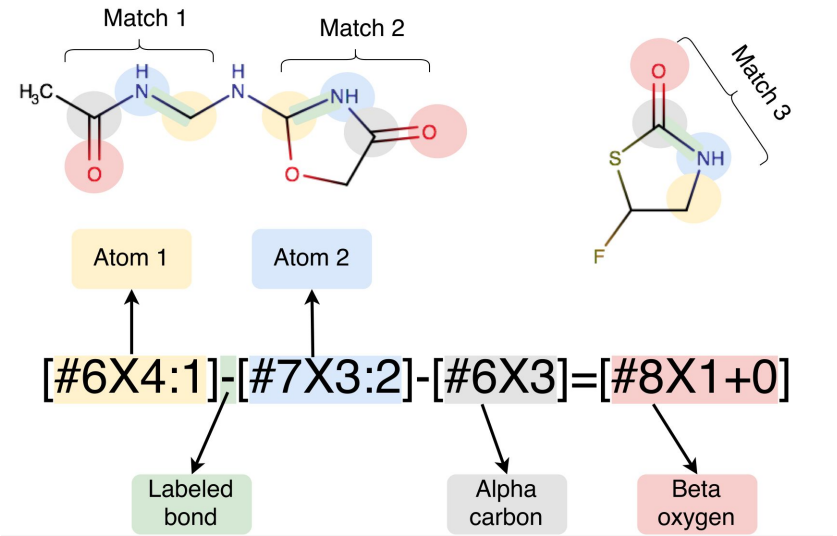
Standards: Representation of molecular systems; forcefield descriptions

Documentation: Theory; toolkit documentation; tutorials and training materials

Publications: Communicating the ideas behind our work to the scientific community

The SMIRKS Native Open Force Field spec (SMIRNOFF) avoids atom typing and simplifies parameter assignment

match valence terms **directly** and **consistently** for small molecules, proteins, and other biomolecules



Use of industry-standard SMARTS/SMIRKS chemical perception greatly simplifies tooling for parameter assignment while solving issues with extensibility and flexibility

SMIRNOFF allowed significant compression of smirnoff99Frosst, our AMBER-lineage starting point

Description	Force Field	Lines of parameters
Basic Amber FF:	parm99	720
Merck Frosst small mol:	parm@Frosst	2893
Total:		3613

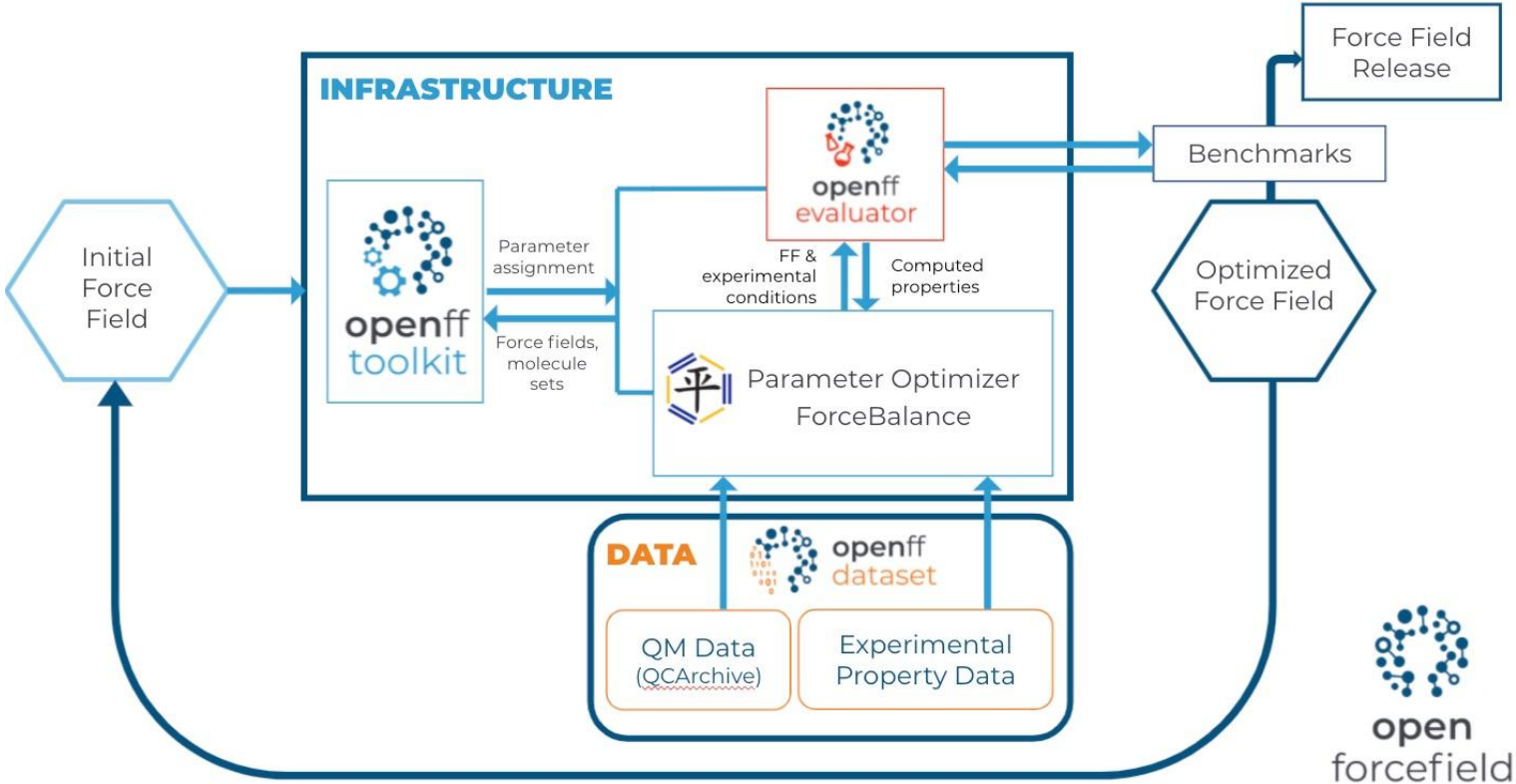
smirnoff99Frosst	332

Chemical Space Coverage

Database	smirnoff 99Frosst	parm @Frosst
DrugBank	99.7%	60%
ZINC	99.8%	52%
eMolecules	99.5%	--

- Less than 1/10 the size of the original force field
- Removes redundancy
- Almost completely covers pharmaceutical chemical space

OpenFF has built open source infrastructure for automating the construction of general biomolecular force fields



Our first-generation small molecule force field was Amber-compatible but significantly improved on GAFF

Parsley (openff-1.x) small molecule force field

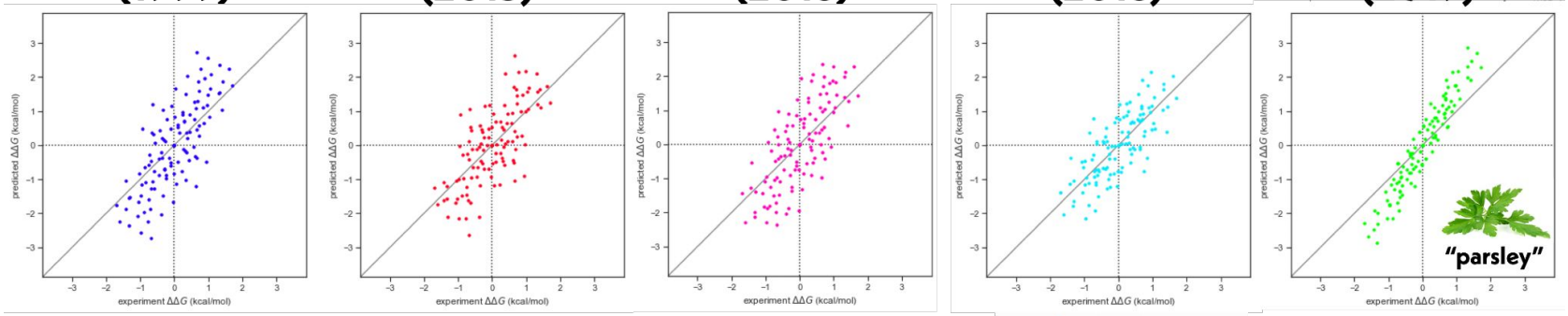
- 1.0.0 (10/2019): The first optimized force field
- 1.1.0 (03/2020): More valence parameter refits and some fixes
- 1.2.0 (06/2020): Expanded and redesigned QM dataset dramatically improved accuracy
- 1.2.1 (09/2020): Bugfix for propynes/HMR
- 1.3.0 (10/2020): Addresses some amide issues
- 1.3.1 (06/2021): Bugfix for sulfonamide geometries

In parallel, we are able to do new science: We've run hundreds of fitting experiments to test out a wide range of ideas, e.g. effect of vibrational frequency fitting

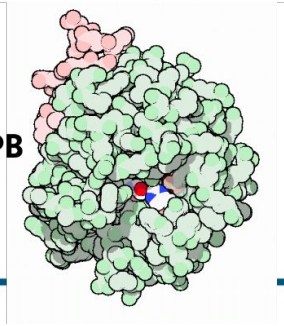
We've made significant progress in the accuracy of free energy calculations

Open Force Field Initiative

GAFF 1 (1999) | OPLS2.1 (2015) | GAFF 2 (2016) | smirnoff99Frosst (2018) | openff 1.0 (2019)



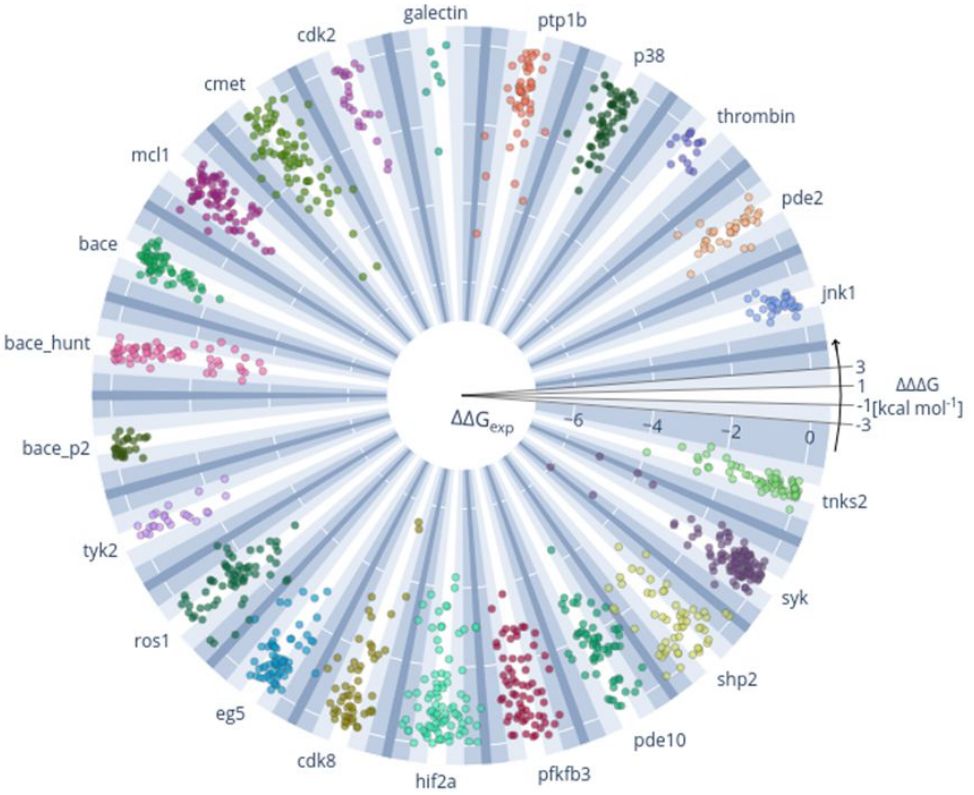
thrombin
PDB101: 1PPB



perses
http://github.com/choderalab/perses

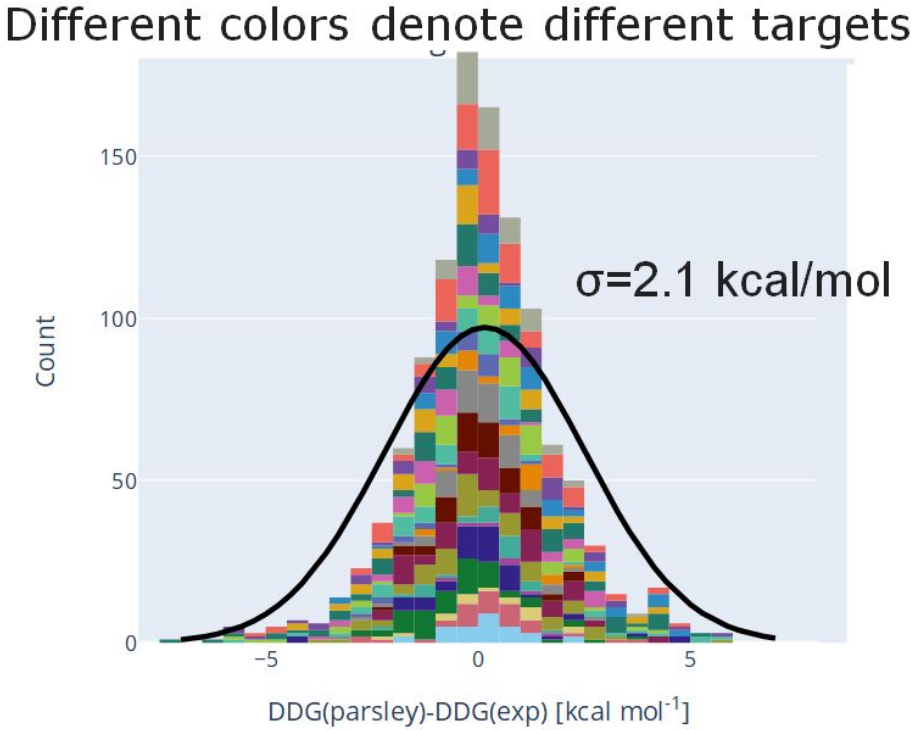
A comprehensive benchmark of free energy calculations performance shows openff-1.0.0 is quite good

- Overview over all calculations performed
- Radial: exp. ΔΔG in kcal/mol
- Polar: difference between calc. and exp. ΔΔG, ΔΔΔG in kcal/mol



Protein ligand benchmark
https://github.com/openforcefield/protein-ligand-benchmark

Most perturbations (ΔΔG) deviate less than 1 kcal/mol from experiment (openff-1.0.0)



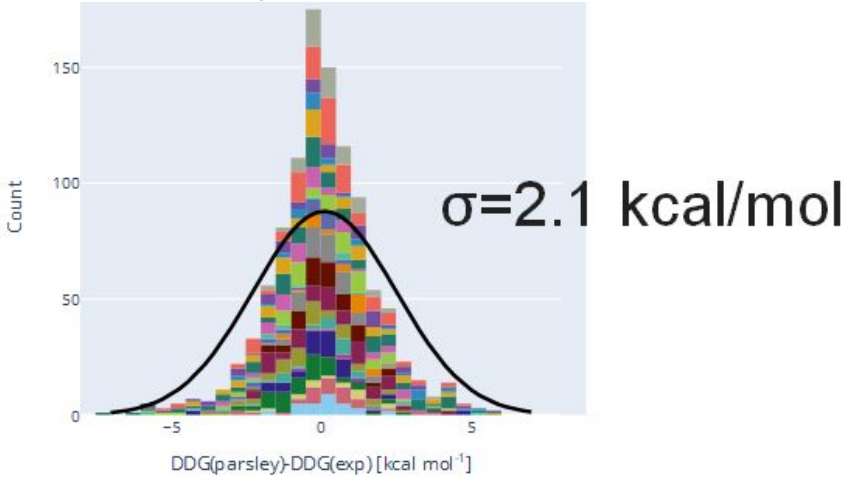
Abs. Error [kcal/mol]	# Perturbations	% of total
< 0.5	322	29
<1.0	592	52
<2.0	911	79
<3.0	1052	92
total	1149	100

Origin of errors:

- Set-up (poses, charges)
- Sampling (simulation time)
- Model accuracy (Force Field)
- Experimental data

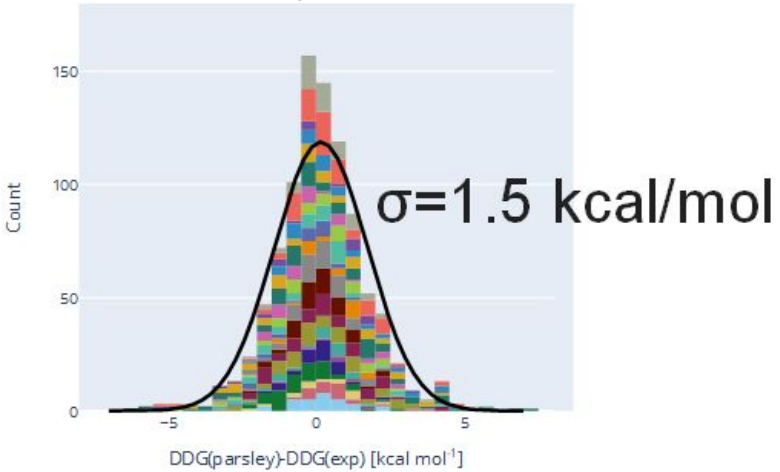
If we filter for just fully converged free energy calculations, accuracy is quite good (openff-1.0.0)

All perturbations



Abs. Error [kcal/mol]	# Perturbations	% of total
< 0.5	322	29
<1.0	592	52
<2.0	911	79
<3.0	1052	92
total	1149	100

Filtered perturbations



Abs. Error [kcal/mol]	# Perturbations	% of total
< 0.5	383	32
<1.0	508	57
<2.0	748	85
<3.0	835	94
total	885	100

Our second-generation small molecule force field Sage makes numerous improvements over Parsley

Sage (openff-2.x) makes significant improvements:

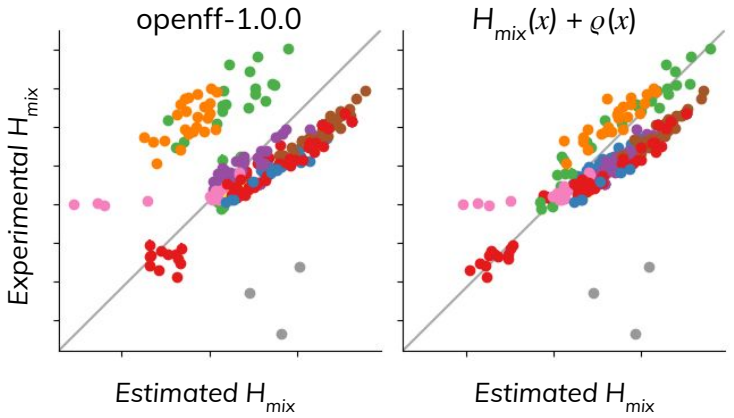
- 2.0.0 (08/2021): Further improvements to the **valence parameters** and **select van der Waals** parameters retrained against experimental mixture enthalpies and densities
- Working on minor releases that include:
  - **Significantly expanded quantum chemical datasets** (torsion drives on combinatorial fragments, vibrational frequencies)
  - **Refited key impropers**
  - **Virtual sites**
  - **Co-optimization of charge models and vdW**

https://github.com/openforcefield/openff-forcefields/releases/tag/2.0.0-rc.1

Sage includes key 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** measurements (NIST ThermoML)
  - Directly includes **aqueous** 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

Rosemary will self-consistently parameterize small molecules and biopolymers (proteins and friends)

SMIRNOFF99Frosst

Initial SMIRNOFF port of the parm99Frosst force field

Parsley

Retrained **valence** parameters against a redesigned QC data set

Sage

Retrained **vdW** parameters against physical property data + retrained **valence** parameters

Rosemary

Self consistent **biopolymer** + small molecule force field

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.

OpenMM
High performance, customizable molecular simulation.

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

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

Release Notes

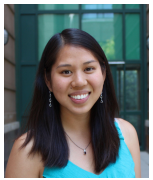
v3.0.0 November 2020

General Notice

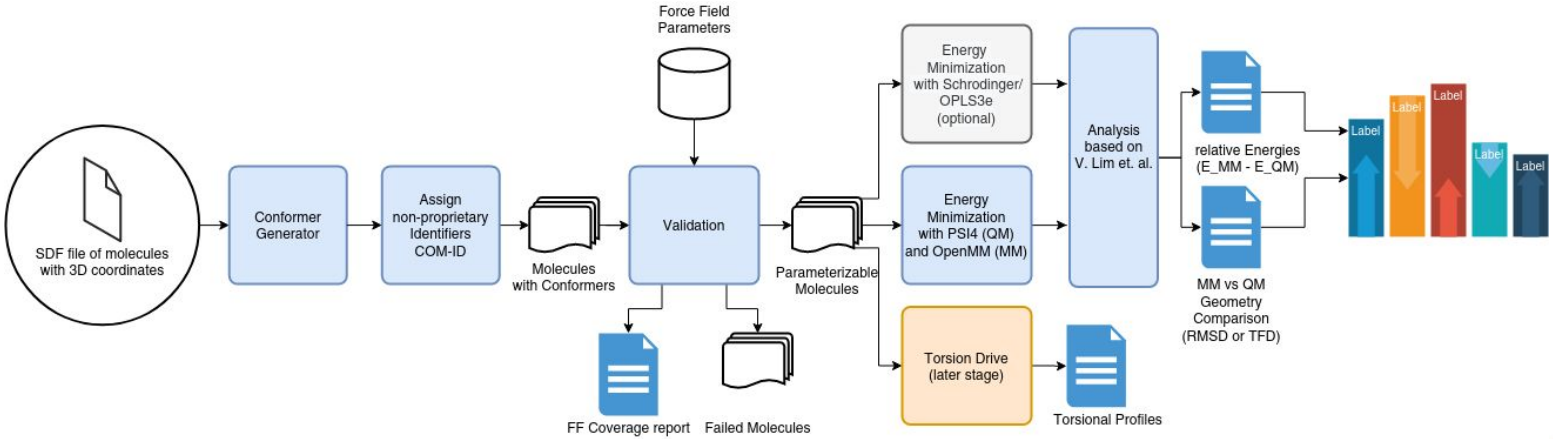
- OpenFF 1.3.0 and 1.2.1 support
- Bug Fixing

Molecular Simulation MoSDeF Design Framework

OpenFF Benchmark



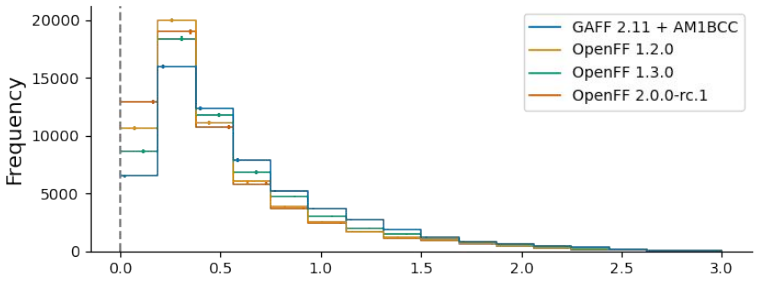

- Automates running QM conformational energy benchmark
- CLI-first approach, using OpenFF Python backend
- Pinned conda environments and conda installers to provide consistent results

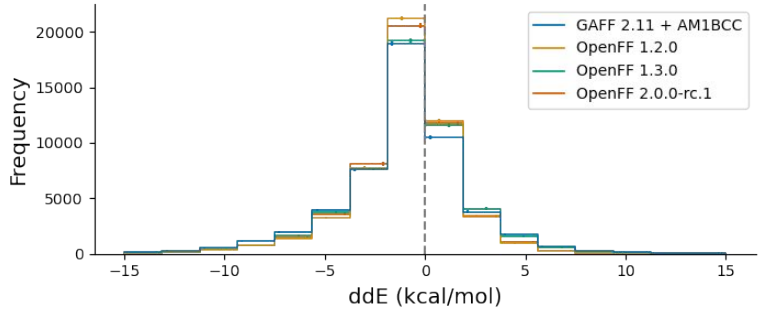


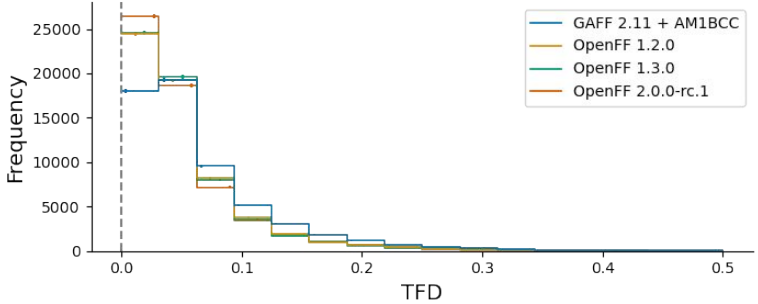
Lim VT, Hahn DF, Tresadern G *et al.* Benchmark assessment of molecular geometries and energies from small molecule force fields [version 1; peer review: 2 approved]. *F1000Research* 2020, **9**(Chem Inf Sci):1390 (https://doi.org/10.12688/f1000research.27141.1)

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

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

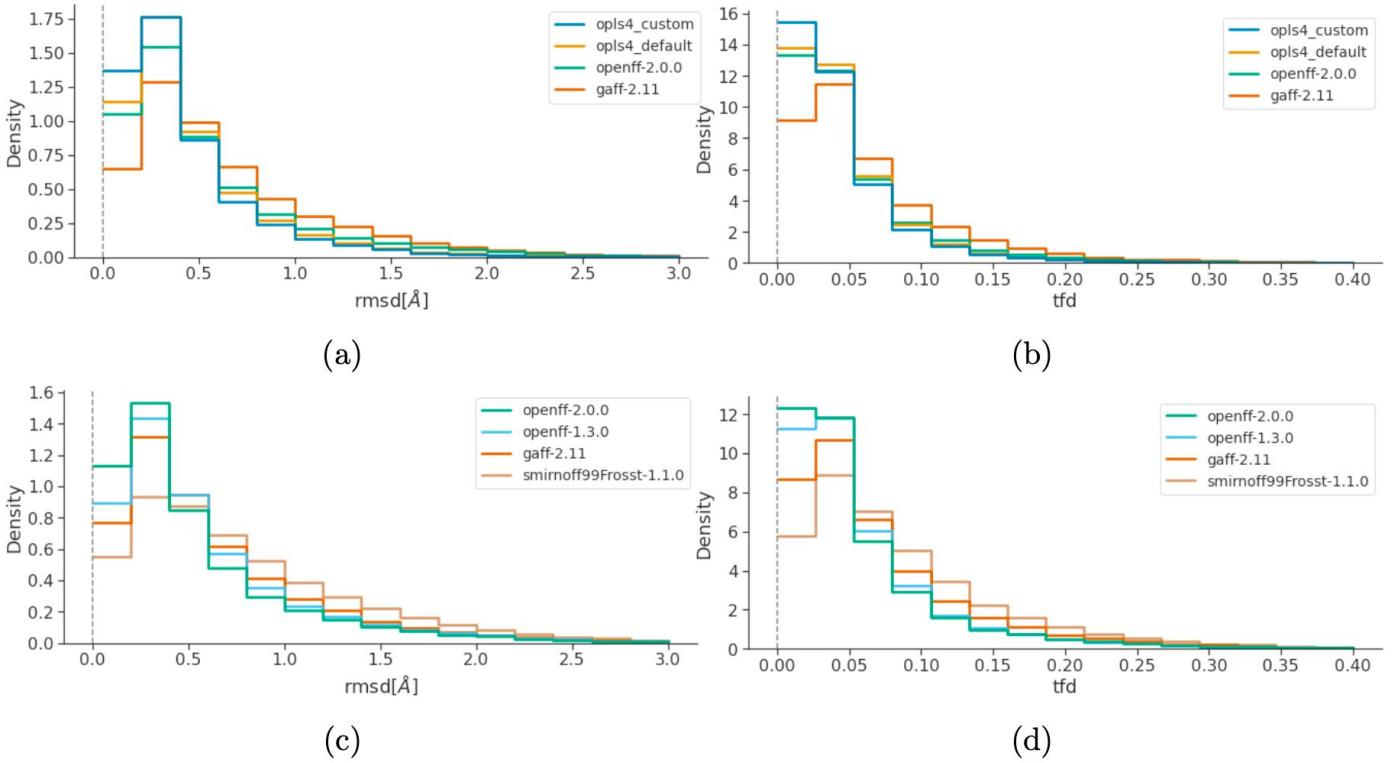






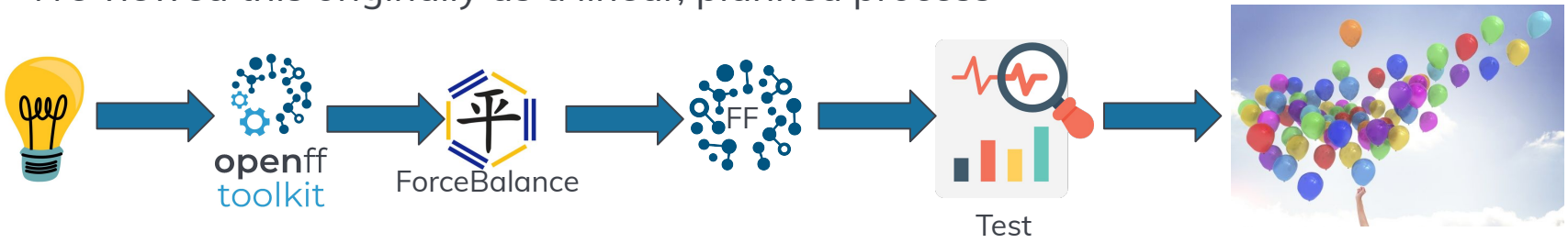
* 95% bootstrapped confidence intervals

Benchmarking on proprietary industry datasets shows similar performance, significant improvement

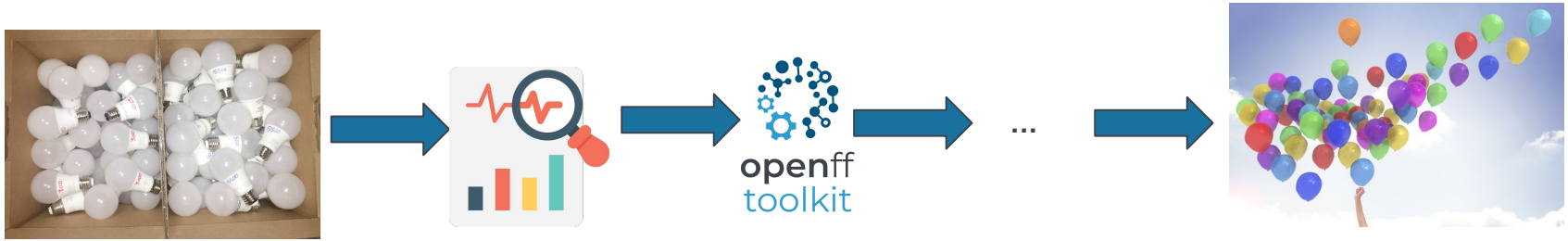


The process works, and it's changing how we think about building force fields

- 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



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

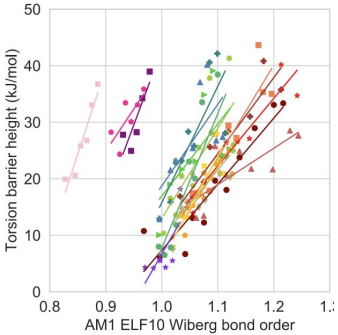
OPEN SCIENCE

Exploring new force field science:

hypothesize - build develop - train - test - iterate

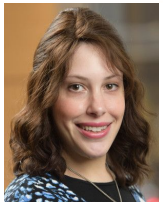
is now almost routine

WBO Interpolation - promising force field science that will drastically simplify valence parameters



openff toolkit

forcebalance


Chaya Stern


Jessica Maat

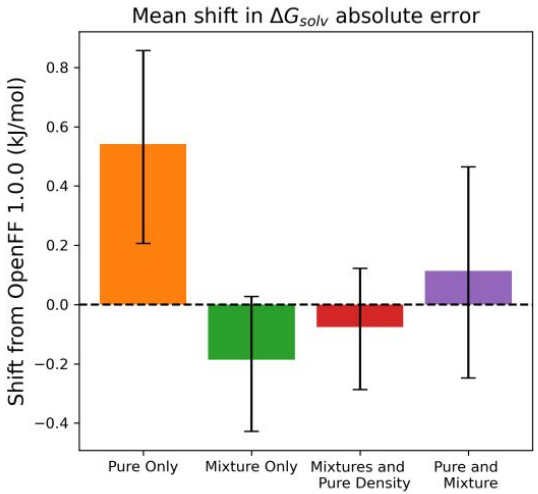

Pavan Behara

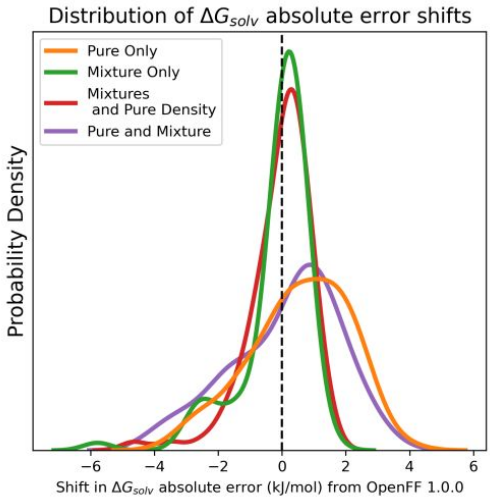
Hypothesis → **Software** → **Training** → **Testing** → **Data Driven Decision**

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 - improvements observed for solvation / transfer free energies

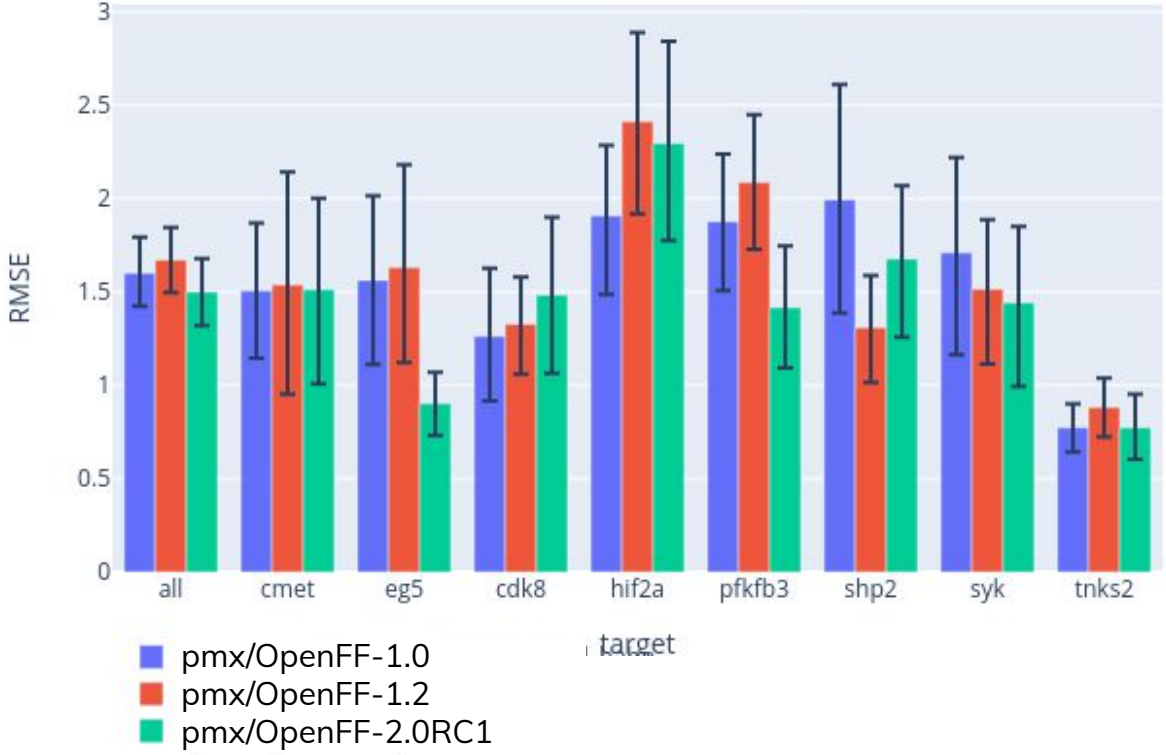
- Benchmarked refit vdW parameters against solvation free energies + transfer free energies
  - Subset of FreeSolv and MNSol
- Training to mixture data significantly improves performance relative to training to pure data only, or pure + mixture





Sage improves free energy benchmarks over Parsley

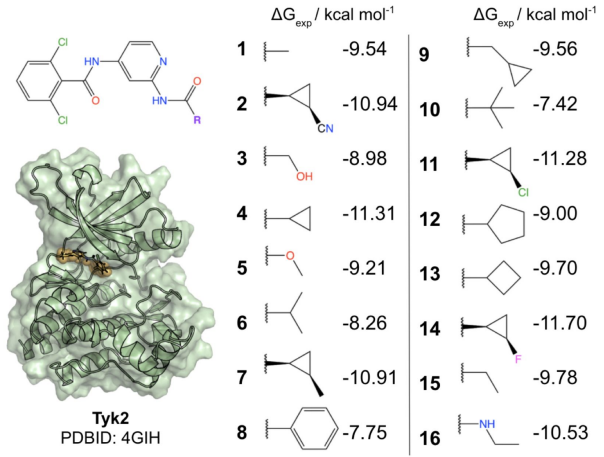
- RMSE based on ΔΔG in kcal/mol
- Error bars are 95% CI

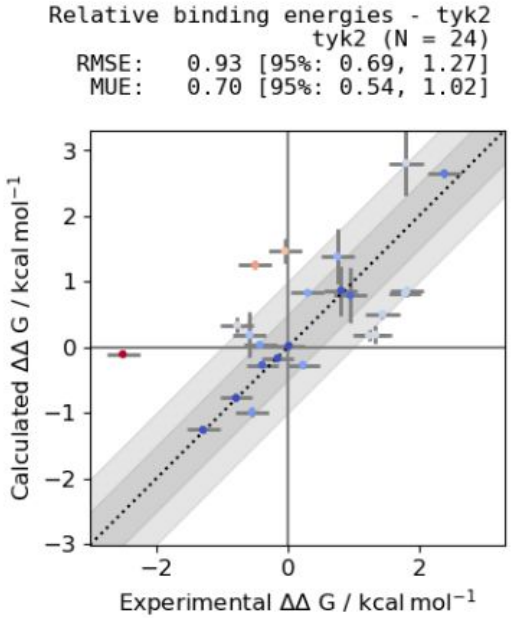


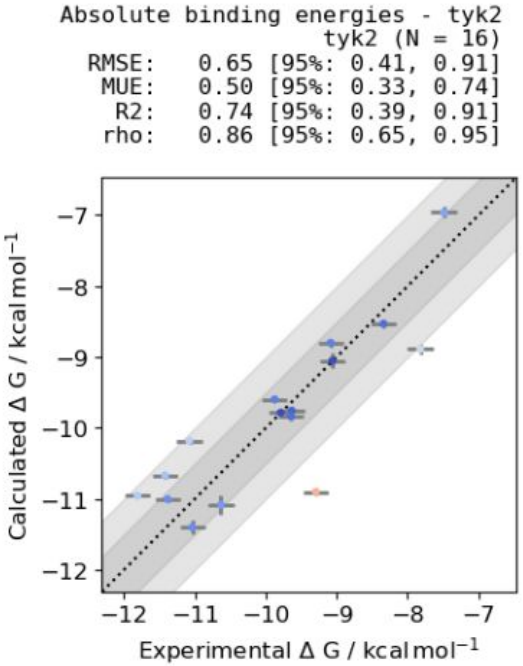
Vytas Gapsys

Sage provides excellent performance on some systems

- Error bars are 95% CI

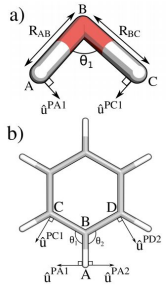






perses 0.10.0
http://github.com/choderalab/perses

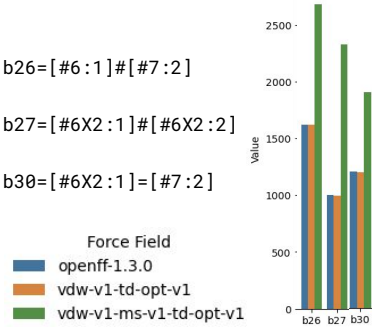
Modified Seminario* - heavily automated fitting pipeline used to explore new fitting target in days

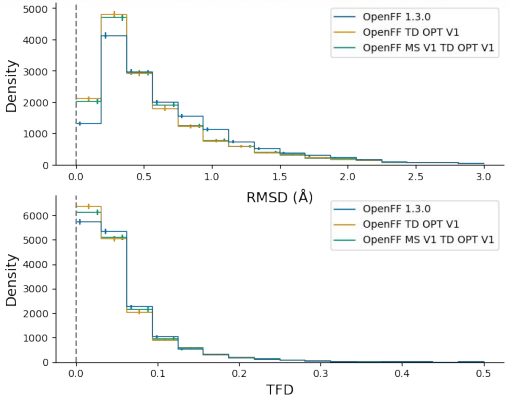



```
b26=[#6:1]#[#7:2]

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 - AM1-BCC charge model currently being re-trained against QC and exp. data

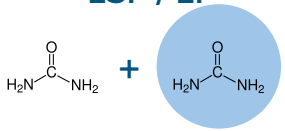
33 Double-bonded oxygen in a lactone or lactam

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

+

openff evaluator

ESP / EF



RESP2 δ=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

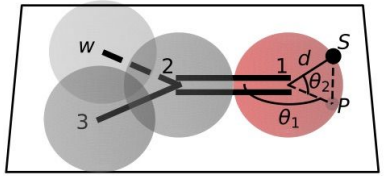
Test fits being performed against a combination of QC ESP / EF data and mixture exp. data

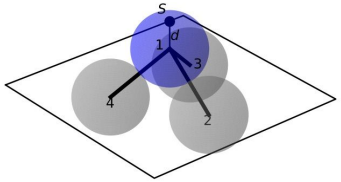
→

Testing

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

Virtual Sites - We're now testing virtual sites in force fields and they will probably make Rosemary

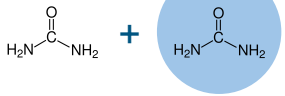




+

openff
toolkit

ESP / EF



RESP2 δ=0.6

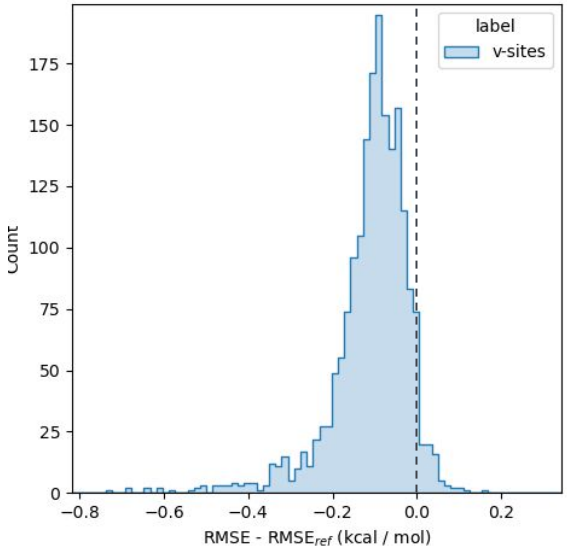
Trevor Gokey

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

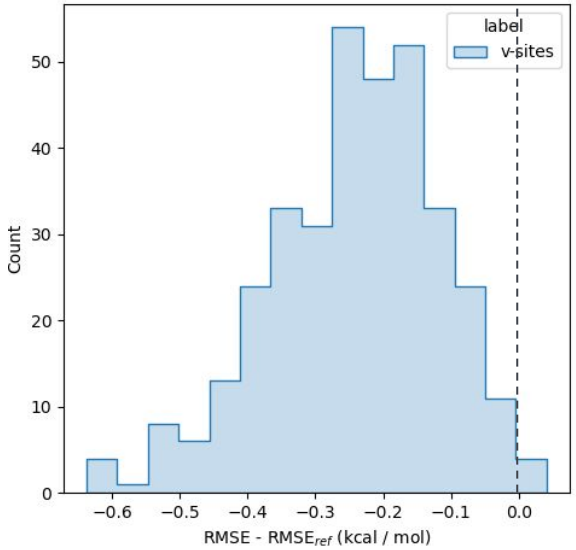
Addition of virtual sites led to better performance on electrostatic potentials (ESPs)

Shift in ESP RMSE after addition of virtual sites

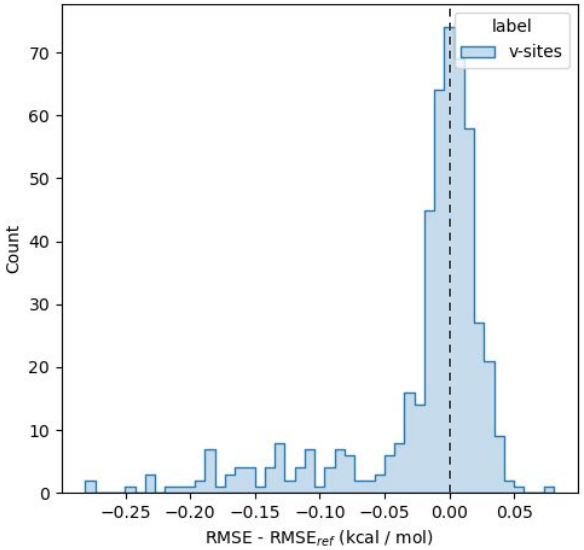
Chlorine



Bromine

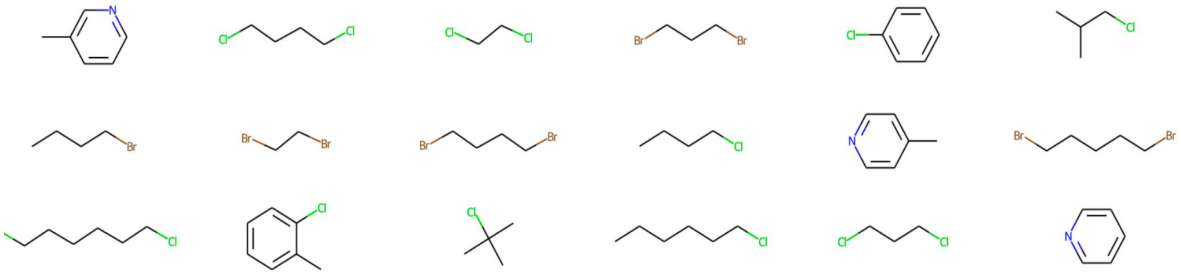


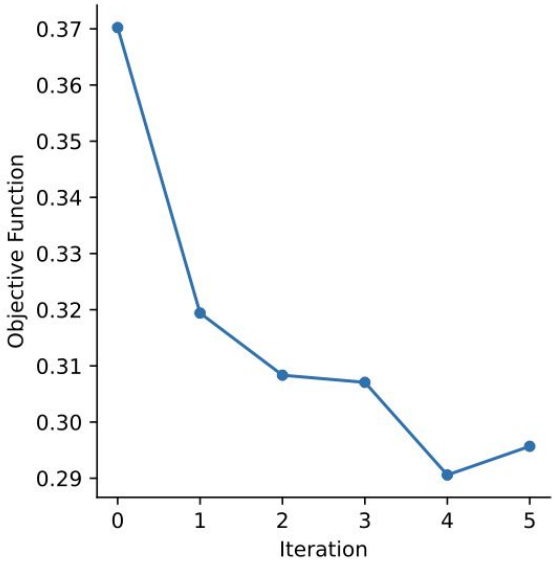
Pyridine



vdW parameters were refit after integrating new charge model

- 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

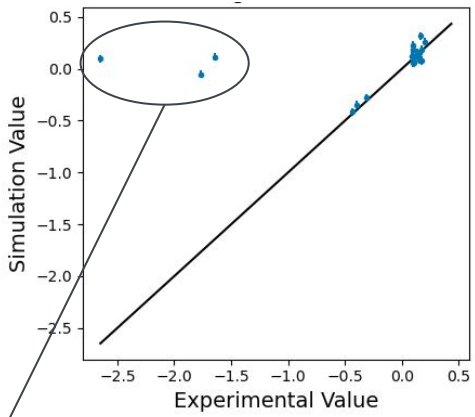




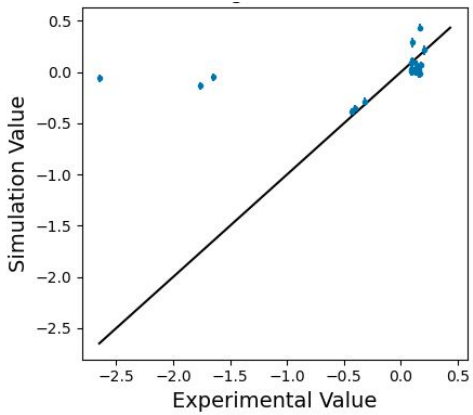
Sage vdW refits did not improve enthalpy of mixing of pyrrole and pyridine

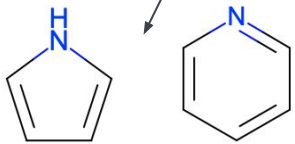
$\Delta H_{mix}(x)$ of Heteroaromatics

OpenFF 1.3.0



OpenFF 2.0.0

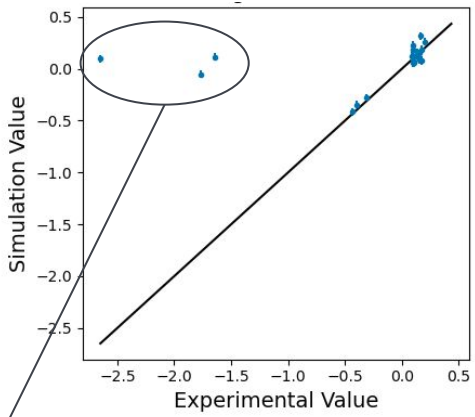




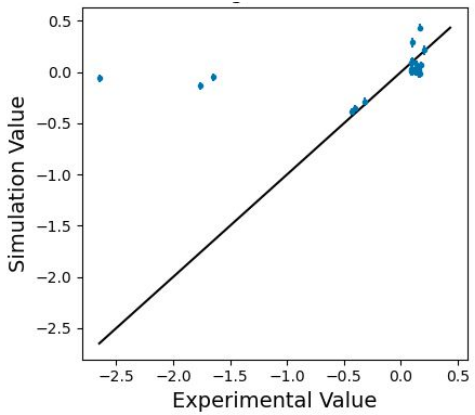
The addition of virtual sites markedly improved modeling of interactions

$\Delta H_{mix}(x)$ of Heteroaromatics

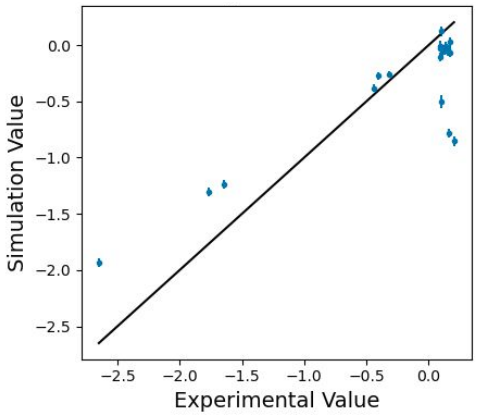
OpenFF 1.3.0

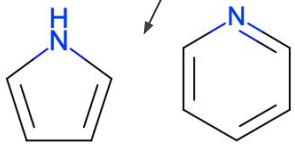


OpenFF 2.0.0



FF + new BCCs / v-sites

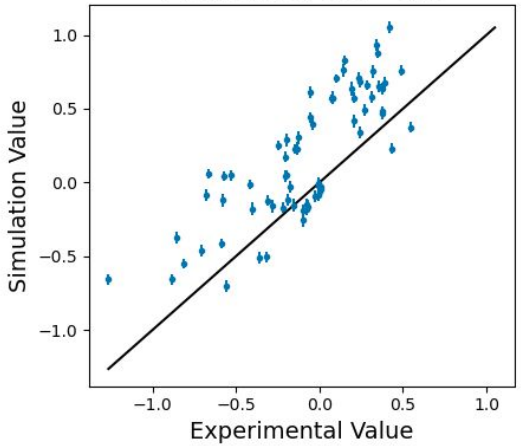




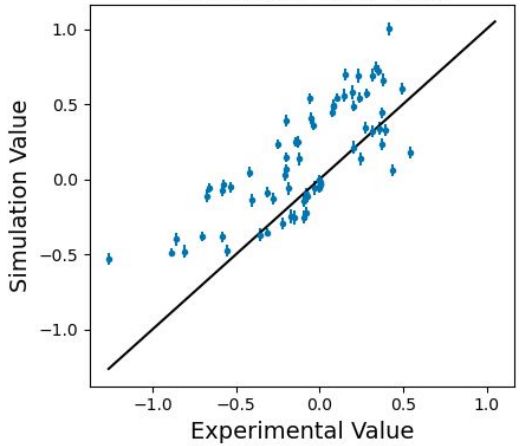
Similarly, the addition of v-sites also resulted in significantly improved performance with chlorine

$\Delta H_{mix}(x)$ of mixtures containing Chlorine

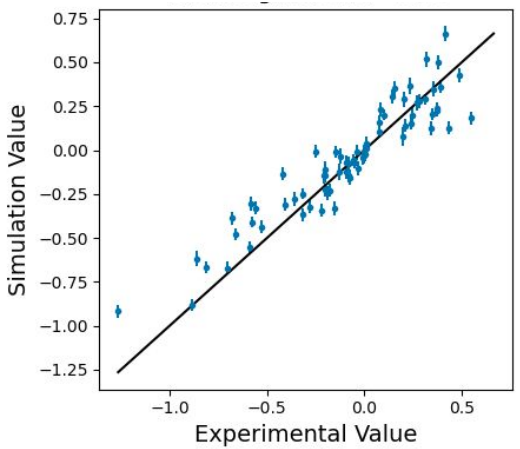
OpenFF 1.3.0



OpenFF 2.0.0

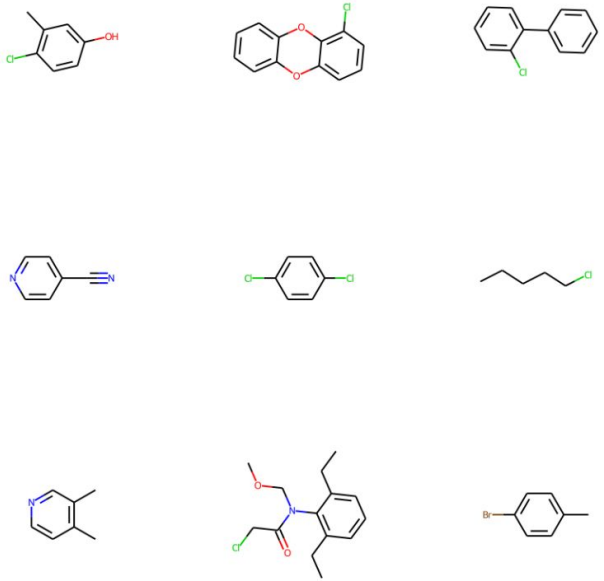


FF + new BCCs / v-sites

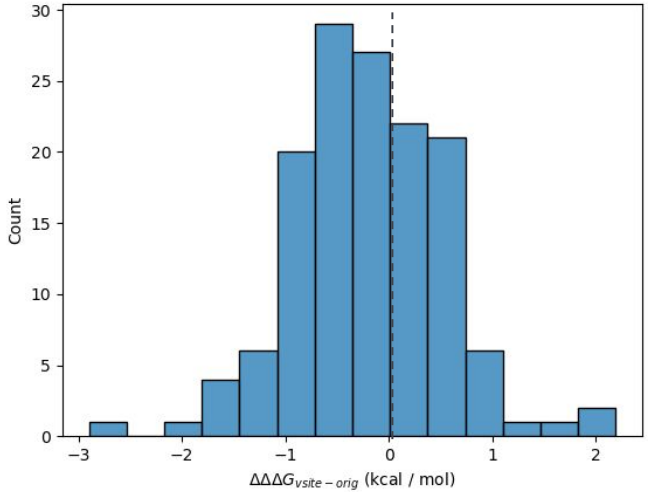


Hydration free energies largely improved after addition of virtual sites

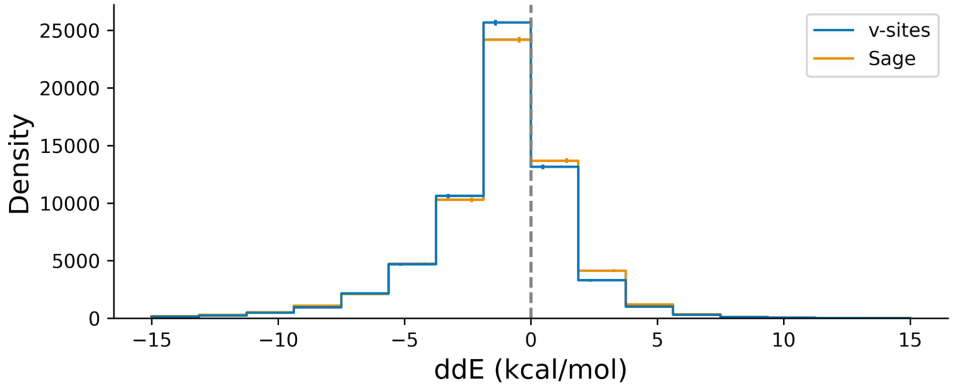
- Test set comprised of subset of FreeSolv that would be assigned virtual sites

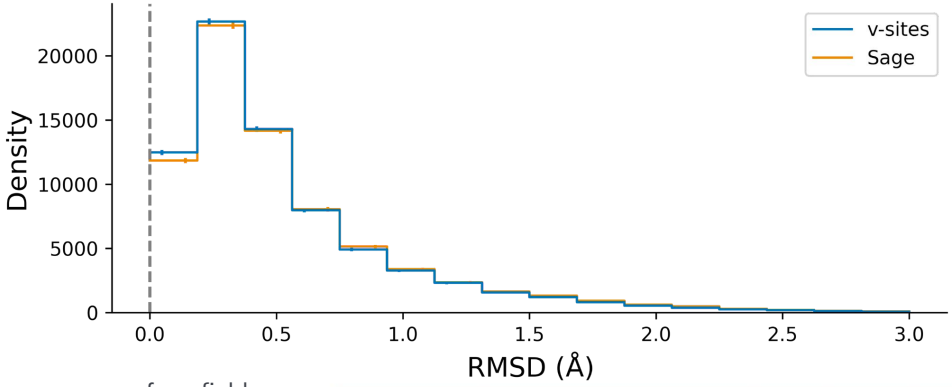


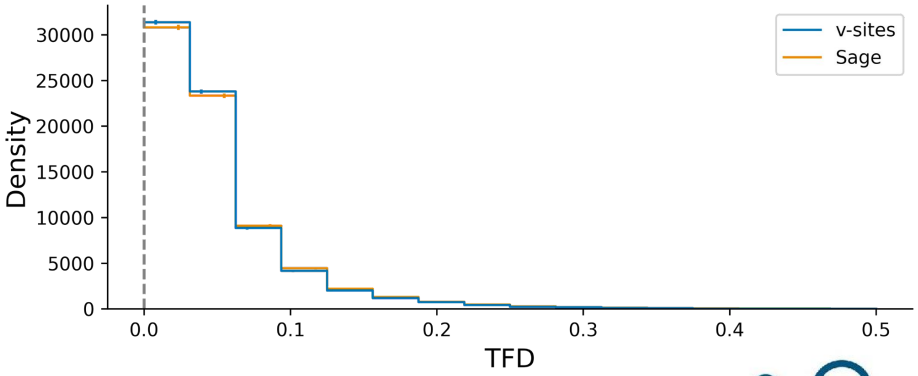
Shift in **absolute error** after addition of virtual sites



After refitting remaining valence terms, benchmarks show moderate improvement*

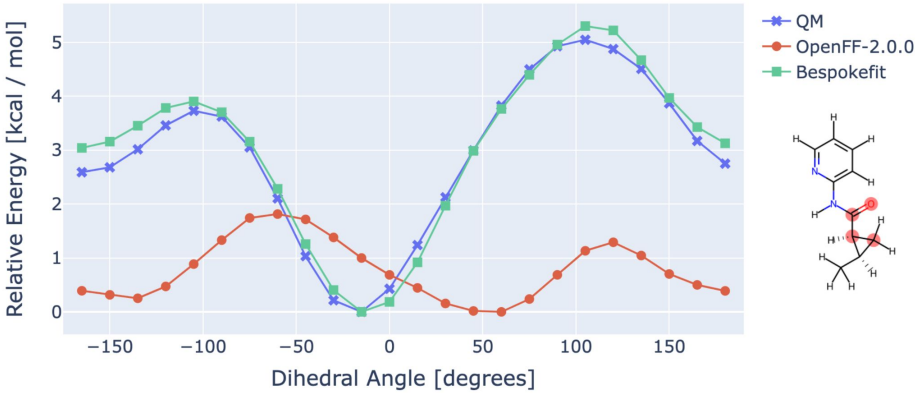




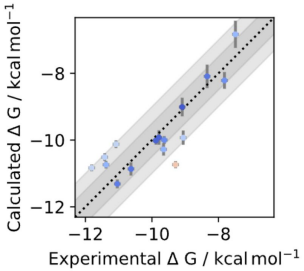


OpenFF BespokeFit enables custom torsions to be fit to QM for a molecule or chemical series

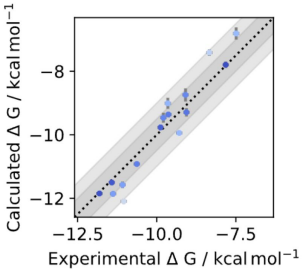
- Can retrain torsion parameters to bespoke torsion scans generated for ‘fragments’ of original molecule
- Very fast with GFN-XTB or similar; also works with QM method of choice



OpenFF-1.3.0
tyk2 (N = 16)
RMSE: 0.68 [95%: 0.49, 0.91]
MUE: 0.56 [95%: 0.41, 0.80]
R2: 0.72 [95%: 0.35, 0.88]
rho: 0.85 [95%: 0.62, 0.94]



Bespoke default-1.3.0
tyk2 (N = 16)
RMSE: 0.51 [95%: 0.35, 0.69]
MUE: 0.42 [95%: 0.28, 0.59]
R2: 0.93 [95%: 0.84, 0.97]
rho: 0.97 [95%: 0.92, 0.99]



openmm-bespokefit: https://github.com/openforcefield/openff-bespokefit

Our infrastructure creates opportunities for new science - maybe you want to contribute?

- Polarizable force fields: An area of interest, but we've not worked there yet
  - Create two force fields which differ only in polarizability vs fixed-charge
  - Train them to the same data
  - Benchmark on the same data
- Likewise, but for different types of multipole expansion…
- (Currently in progress: Experiments replacing 12-6 LJ with buffered exponentials, etc.)

open forcefield

@openforcefield

www.openforcefield.org

Newcastle University | 150 YEARS | SCIENCE, AGRICULTURE AND ENGINEERING

Functional Form Exploration

Smirnoff-Plugins update

Joshua Horton, Daniel Cole & Simon Boothroyd

smirnoff-plugins enables exploration of functional forms

Newcastle University | 150 YEARS SCIENCE, AGRICULTURE AND ENGINEERING

smirnoff-plugins[1]

A plugin framework to rapidly prototype and test extensions to the force field functional form.

Feature	Progress
Automatic scaled 1-4 support	✓
Virtual site support	✓
Long-range correction support	✓*
Free energy support	✓
On conda-forge	✓
Openff stack integration	✓
Trivial to add new forms	✓

Build new force fields using the same OpenFF API

```
buckingham_handler = force_field.get_parameter_handler("DampedBuckingham68")
buckingham_handler.gamma = 35.8967 * unit.nanometer ** -1
buckingham_handler.add_parameter(
    {
        "smirks": "[#1:1]-[#8X2H2+0]-[#1]",
        "a": 0.0 * unit.kilojoule_per_mole,
        "b": 0.0 / unit.nanometer,
        "c6": 0.0 * unit.kilojoule_per_mole * unit.nanometer ** 6,
        "c8": 0.0 * unit.kilojoule_per_mole * unit.nanometer ** 8,
    }
)
buckingham_handler.add_parameter(
    {
        "smirks": "[#1]-[#8X2H2+0:1]-[#1]",
        "a": 1600000.0 * unit.kilojoule_per_mole,
        "b": 42.00 / unit.nanometer,
        "c6": 0.003 * unit.kilojoule_per_mole * unit.nanometer ** 6,
        "c8": 0.00003 * unit.kilojoule_per_mole * unit.nanometer ** 8,
    }
)
```


1 https://github.com/openforcefield/smirnoff-plugins
* https://github.com/openmm/openmm/issues/3277

Functional Forms

Lennard-Jones:

$$U_{ij}^{LJ} = \epsilon_{ij}\left[\left(\frac{r_{m,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{m,ij}}{r_{ij}}\right)^{6}\right] \qquad r_{m,ij} = 2^{1/6}\sigma_{ij}$$

Double Exponential potential:

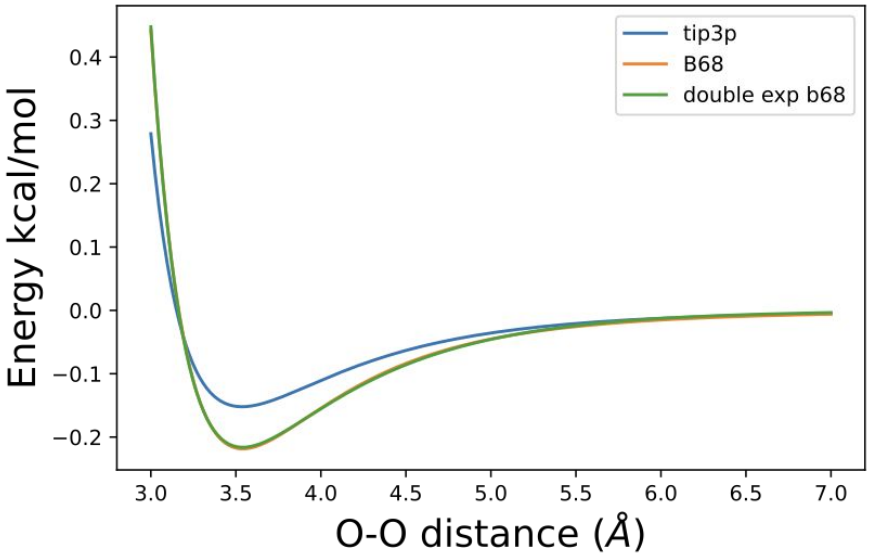
$$U_{ij}^{DEXP} = \epsilon_{ij}\left[\frac{\beta e^{\alpha}}{\alpha-\beta}exp\left(-\alpha\frac{r_{ij}}{r_{m,ij}}\right) - \frac{\alpha e^{\beta}}{\alpha-\beta}exp\left(-\beta\frac{r_{ij}}{r_{m,ij}}\right)\right]$$

- Natural soft core
- flexibility (α defines steepness of repulsion and β the decay of the attraction) can be fit to mimic other potentials.
- Cheap to compute, factors can be pre-computed

Buckingham damped 6-8 potential:

$$U(r) = E_{bonded} + A\exp(-br) - f_{damp,6}\frac{C_6}{r^6} - f_{damp,8}\frac{C_8}{r^8}$$

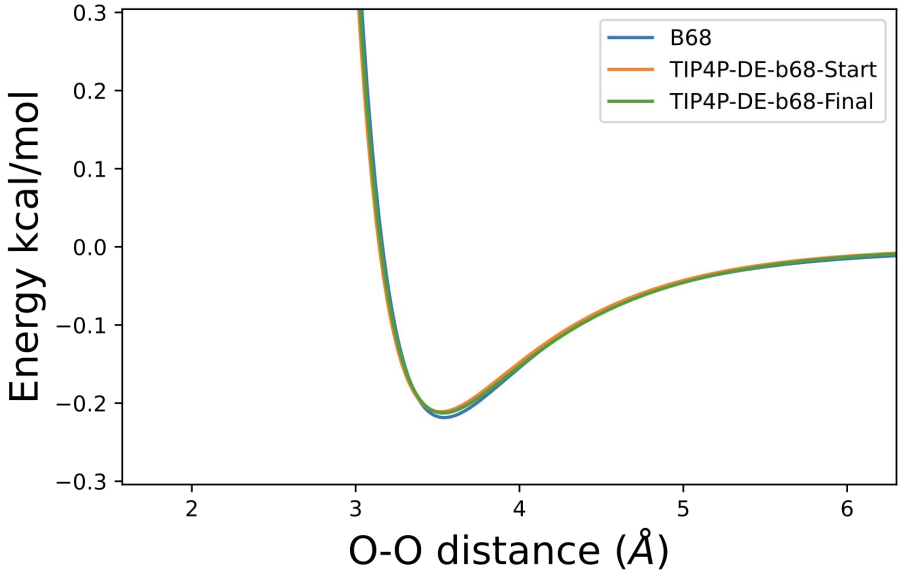
- Physically motivated, inclusion of C8 which is normally absorbed by C6
- Exponential repulsion
- Needs damping function as U(0) = ∞



B68 https://doi.org/10.1063/5.0014469
DEXP https://doi.org/10.1021/acs.jctc.0c01267

Results for solution phase properties on diverse set

Property	DEXP-ALL (Error)	SAGE (Error)
Pure Density	0.022 g/mL (0.019, 0.026)	0.031 g/mL (0.026,0.037)
Binary Density	0.011 g/mL (0.01, 0.012)	0.014 g/mL (0.013, 0.015)
Enthalpy of mixing	0.400 kJ/mol (0.356, 0.446)	0.541 kJ/mol (0.5, 0.59)



Summary

Can we fit a new functional form?

Yes we can do large scale force field fits using OpenFF infrastructure easily with minimal extra steps compared to standard fits.

Next steps

- Refits with reduced scope for a proof of concept publication
- More solvation free energy benchmark calculations
- Valence fits to build out any missing infrastructure.

UK Research and Innovation

We are exploring new machine learning technologies for advancing force field science

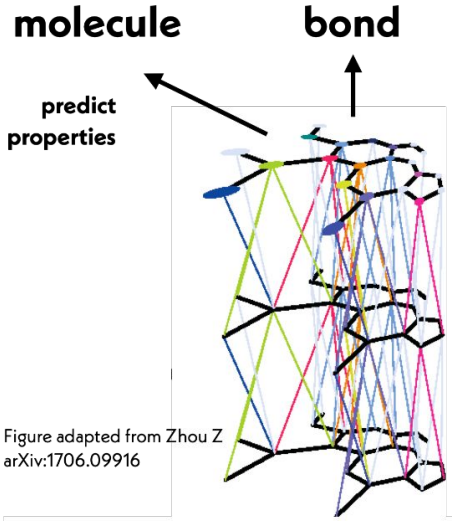


Figure adapted from Zhou Z arXiv:1706.09916

Learns **electronegativity** (e_i) and **hardness** (s_i) subject to fixed charge sum constraint:

$$\{\hat{q}_i\} = \underset{q_i}{\operatorname{argmin}} \sum_i \hat{e}_i q_i + \frac{1}{2}\hat{s}_i q_i^2$$

$$\sum_i \hat{q}_i = \sum_i q_i = Q$$

$$\mathbf{e}_k^{(t+1)} = \phi^e(\mathbf{e}_k^{(t)}, \sum_{i \in \mathcal{N}_k^e} \mathbf{v}_i, \mathbf{u}^{(t)}), \quad \text{(edge update)}$$

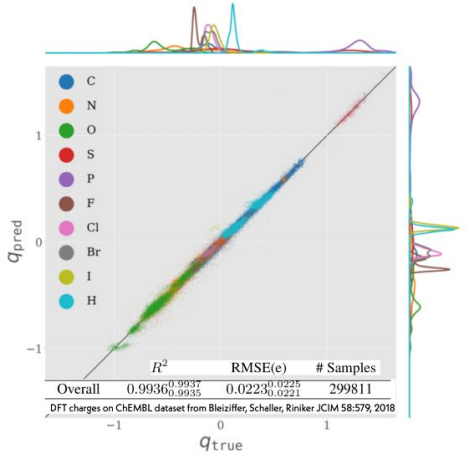
$$\bar{\mathbf{e}}_i^{(t+1)} = \rho^{e\to v}(E_i^{(t+1)}), \quad \text{(edge to node aggregate)}$$

$$\mathbf{v}_i^{(t+1)} = \phi^v(\bar{\mathbf{e}}_i^{(t+1)}, \mathbf{v}_i^{(t)}, \mathbf{u}^{(t)}), \quad \text{(node update)}$$

$$\bar{\mathbf{e}}^{(t+1)} = \rho^{e\to u}(E^{(t+1)}), \quad \text{(edge to global aggregate)}$$

$$\bar{\mathbf{v}}^{(t+1)} = \rho^{v\to u}(V^{(t)}), \quad \text{(node to global aggregate)}$$

$$\mathbf{u}^{(t+1)} = \phi^u(\bar{\mathbf{e}}^{(t+1)}, \bar{\mathbf{v}}^{(t+1)}, \mathbf{u}^{(t)}), \quad \text{(global update)}$$



control experiment:
direct prediction of charges: RMSE **0.2800 e**

Gimlet

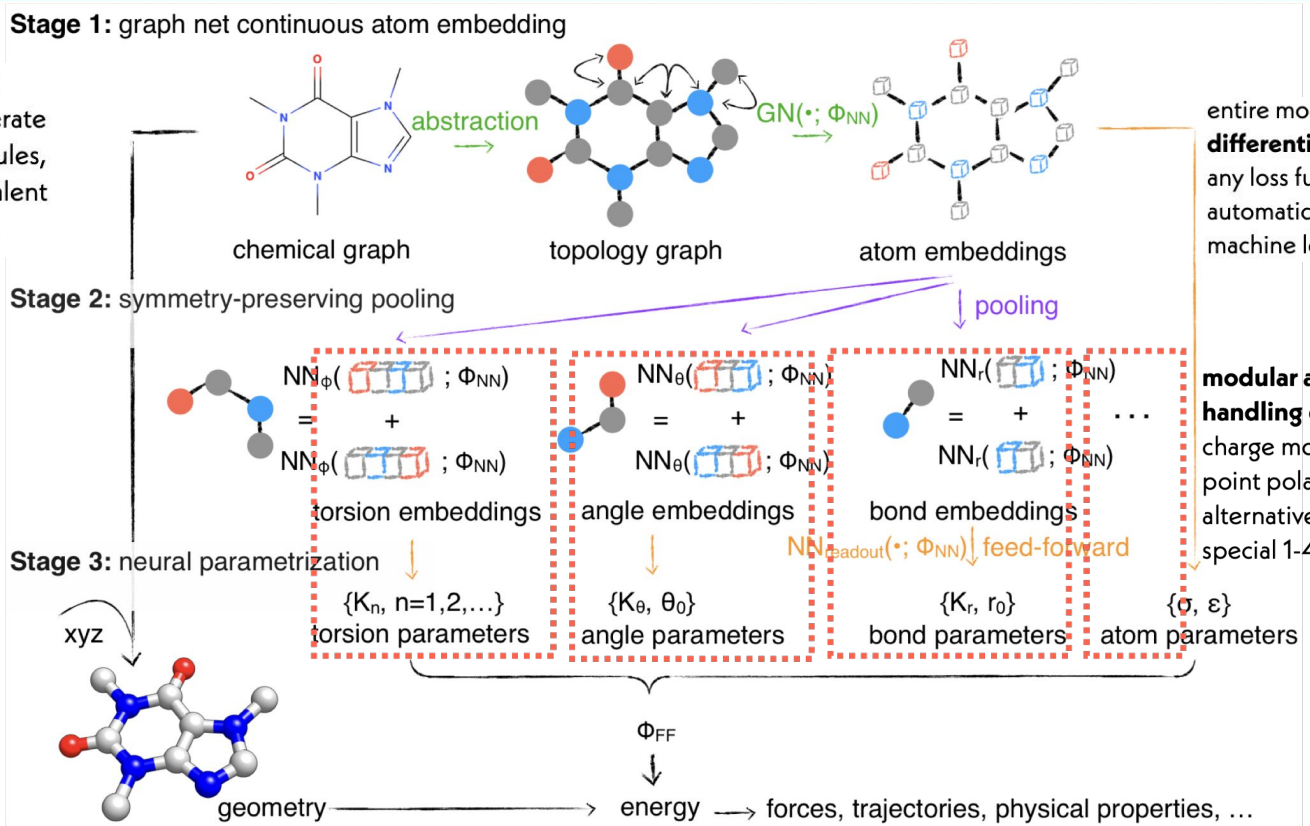
Graph Inference on MoLEcular Topology

preprint: https://arxiv.org/abs/1909.07903
code: http://github.com/choderalab/gimlet

YUANQING WANG

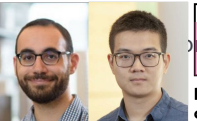
We are exploring new machine learning technologies for advancing force field science

use of only **chemical graph** means that model can generate parameters for small molecules, proteins, nucleic acids, covalent ligands, carbohydrates, etc.



entire model is **end-to-end differentiable** so can be fit to any loss function by standard automatic differentiation machine learning frameworks

modular and extensible handling of potential terms: charge model parameters, point polarizabilities, alternative vdW forms, special 1-4 parameters, etc.

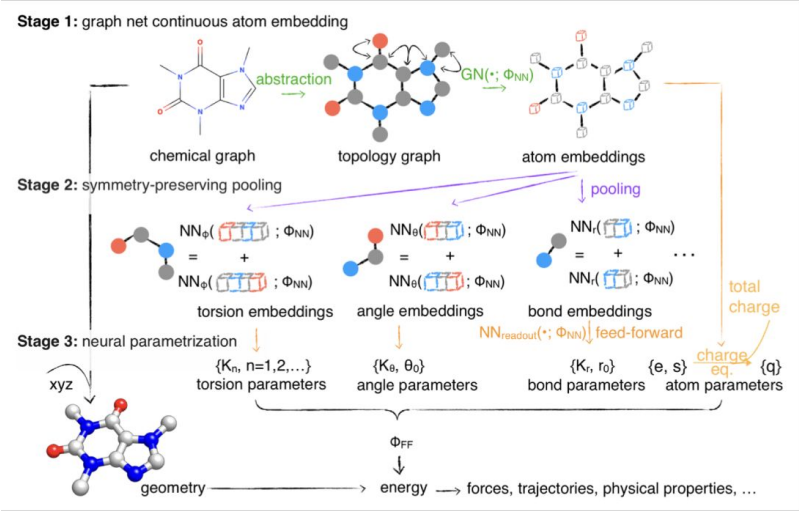


orcefield.org

preprint: https://arxiv.org/abs/2010.01196
code: https://github.com/choderalab/espaloma

Machine learning frameworks could greatly simplify our infrastructure for building new force fields

espaloma architecture



(implemented in pytorch)

http://github.com/choderalab/espaloma

building a new force field

```python
import torch, dgl, espaloma as esp

# retrieve OpenFF Gen2 Optimization Dataset
dataset = esp.data.dataset.GraphDataset.load("gen2").view(batch_size=128)

# define Espaloma stage I: graph -> atom latent representation
representation = esp.nn.Sequential(
    layer=esp.nn.layers.dgl_legacy.gn("SAGEConv"), # use SAGEConv implementation in DGL
    config=[128, "relu", 128, "relu", 128, "relu"], # 3 layers, 128 units, ReLU activation
)

# define Espaloma stage II and III:
# atom latent representation -> bond, angle, and torsion representation and parameters
readout = esp.nn.readout.janossy.JanossyPooling(
    in_features=128, config=[128, "relu", 128, "relu", 128, "relu"],
    out_features={                  # define modular MM parameters Espaloma will assign
        1: {"e": 1, "s": 1}, # atom hardness and electronegativity
        2: {"coefficients": 2}, # bond linear combination
        3: {"coefficients": 3}, # angle linear combination
        4: {"k": 6}, # torsion barrier heights (can be positive or negative)
    },
)

# compose all three Espaloma stages into an end-to-end model
espaloma_model = torch.nn.Sequential(
                 representation, readout,
                 esp.mm.geometry.GeometryInGraph(), esp.mm.energy.EnergyInGraph(),
                 esp.nn.readout.charge_equilibrium.ChargeEquilibrium(),
)

# define training metric
metrics = [
    esp.metrics.GraphMetric(
            base_metric=torch.nn.MSELoss(), # use mean-squared error loss
            between=['u', "u_ref"],         # between predicted and QM energies
            level="g", # compare on graph level
    )
    esp.metrics.GraphMetric(
            base_metric=torch.nn.MSELoss(), # use mean-squared error loss
            between=['q', "q_hat"],         # between predicted and reference charges
            level="n1", # compare on node level
    )
]

# fit Espaloma model to training data
results = esp.Train(
    ds_tr=dataset, net=espaloma_model, metrics=metrics,
    device=torch.device('cuda:0'), n_epochs=5000,
    optimizer=lambda net: torch.optim.Adam(net.parameters(), 1e-3), # use Adam optimizer
).run()

torch.save(espaloma_model, "espaloma_model.pt") # save model
```

Listing 1. Defining and training a modular Espaloma model.

JOSH FASS

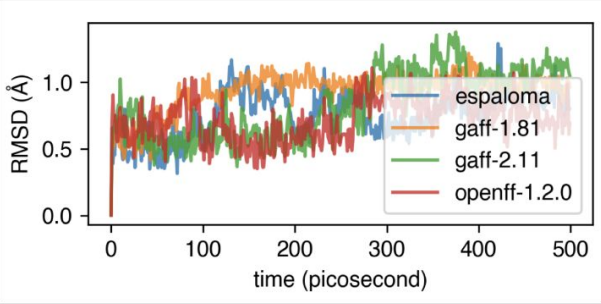
YUANQING WANG

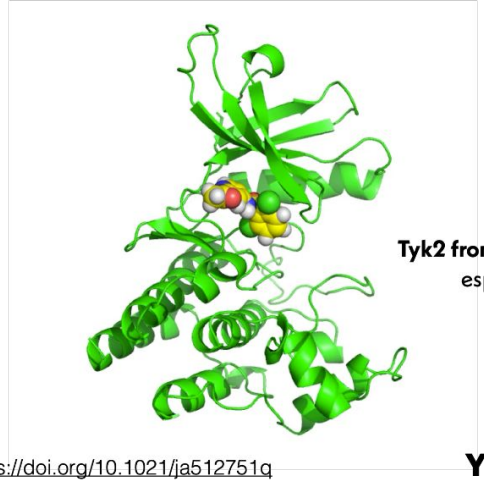
orcefield.org

preprint: https://arxiv.org/abs/2010.01196
code: https://github.com/choderalab/espaloma

Initial experiments suggest a bright future

	(a) dataset	# mols	# trajs	# snapshots	Espaloma RMSE Train	Espaloma RMSE Test	Legacy FF RMSE (kcal/mol) (Test molecules) OpenFF 1.2.0	GAFF-1.81	GAFF-2.11	Amber ff14SB
	PhAlkEthOH (simple CHO)	7408	12592	244036	$0.8656^{0.9131}_{0.8225}$	$1.1398^{1.2332}_{1.0715}$	$1.6071^{1.6915}_{1.5197}$	$1.7267^{1.7935}_{1.6543}$	$1.7406^{1.8148}_{1.6679}$	
	OpenFF Gen2 Optimization (druglike)	792	3977	23748	$0.7413^{0.7920}_{0.6914}$	$0.7600^{0.8805}_{0.6644}$	$2.1768^{2.3388}_{2.0380}$	$2.4274^{2.5207}_{2.3300}$	$2.5386^{2.6640}_{2.4370}$	
	VEHICLe (heterocyclic)	24867	24867	234326	$0.4476^{0.4690}_{0.4273}$	$0.4233^{0.4414}_{0.4053}$	$8.0247^{8.2456}_{7.8271}$	$8.0077^{8.2313}_{7.7647}$	$9.4014^{9.6434}_{9.2135}$	
	PepConf (peptides)	736	7560	22154	$1.2714^{1.3616}_{1.1899}$	$1.8727^{1.9749}_{1.7309}$	$3.6143^{3.7288}_{3.4870}$	$4.4446^{4.5738}_{4.3386}$	$4.3356^{4.4641}_{4.1965}$	$3.1502^{3.1859,*}_{3.1117}$
joint	OpenFF Gen2 Optimization	1528	11537	45902	$0.8264^{0.9007}_{0.7682}$	$1.8764^{1.9947}_{1.7827}$	$2.1768^{2.3388}_{2.0380}$	$2.4274^{2.5207}_{2.3300}$	$2.5386^{2.6640}_{2.4370}$	
joint	PepConf	1528	11537	45902	$1.2038^{1.3056}_{1.1178}$	$1.7307^{1.8439}_{1.6053}$	$3.6143^{3.7288}_{3.4870}$	$4.4446^{4.5738}_{4.3386}$	$4.3356^{4.4641}_{4.1965}$	$3.1502^{3.1859,*}_{3.1117}$





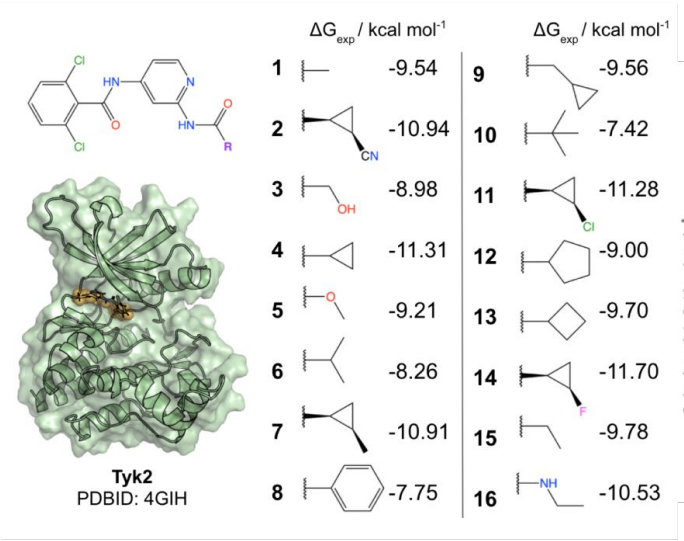
Tyk2 from OpenFF benchmark set
espaloma **joint** model
+ TIP3P water

preprint: https://arxiv.org/abs/2010.01196
code: http://github.com/choderalab/espaloma

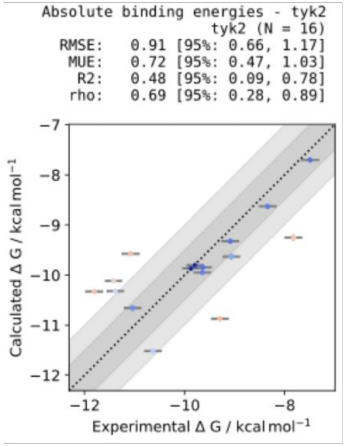
Tyk2 benchmark doi: https://doi.org/10.1021/ja512751q

YUANQING WANG

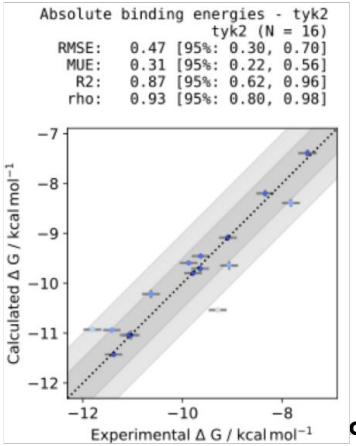
Initial experiments suggest a bright future



OpenFF 1.2.0 small molecule
Amber ff14SB protein
TIP3P water



espaloma "joint" 0.2.2 small molecule
Amber ff14SB protein
TIP3P water

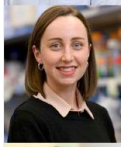


MIKE HENRY

IVÁN PULIDO

IVY ZHANG

DOMINIC RUFA

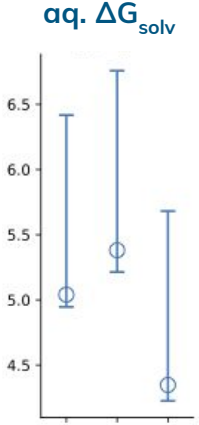
HANNAH BRUCE CDONALD

YUANQING WANG

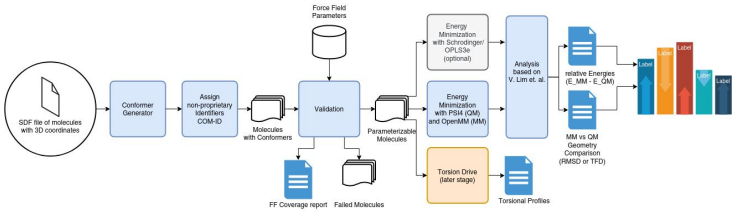

preprint: https://arxiv.org/abs/2010.01196
code: http://github.com/choderalab/espaloma
free energy calculations with http://github.com/choderalab/perses

Conclusions

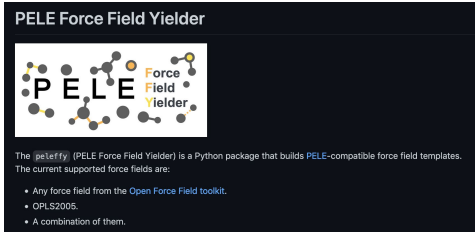
Sage greatly improved performance

Rosemary is coming soon

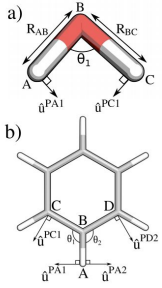


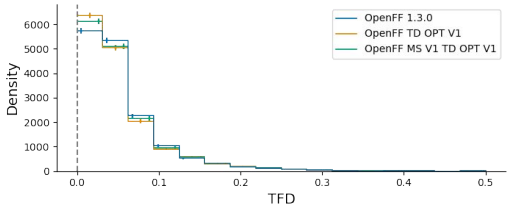


Automated benchmarking has been a major focus and will point the way forward

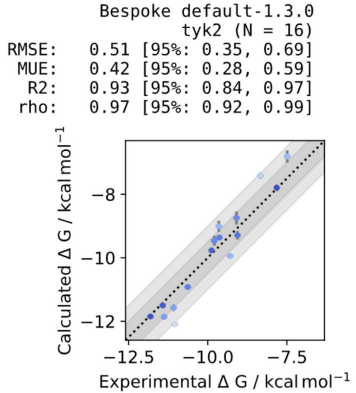


We see **community uptake**, with and without our help





New technology yields better starting points and better FFs



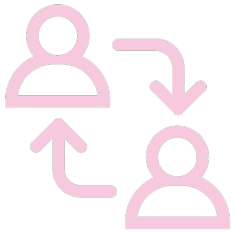
Bespoke torsion fitting is now available

The Open Molecular Software Foundation is a new way to broadly support the biomolecular modeling community

https://omsf.io

OMSF is a **non-profit** organization developing **open source** software and supporting research activities to improve **computational** methods and models for **molecular sciences**.

OMSF is currently seeking 501c3 tax-exempt status (under review).



OMSF facilitates **collaboration between academic and industry stakeholders, software teams, and funders** to design, build and test open source infrastructure for molecular sciences.

The OMSF Mission is broad

https://omsf.io/about/mission/

Accelerate scientific, technical, and commercial progress through rapid dissemination of knowledge and research tools in computational molecular sciences, following open science principles.

Improve modeling accuracy, quality and the overall user experience through more rigorous and sustainable research software development practices and continuous method validation.

Improve sustainability and reduce software development costs across industry and academia by pooling resources, eliminating duplication of efforts, sharing know-how and allowing reuse of software components through open licenses.

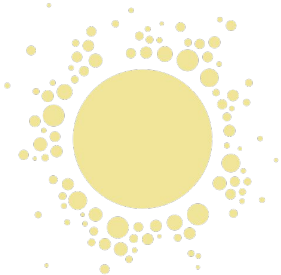
Remove or reduce barriers to collaboration between industry and academia by providing a clear legal and organizational framework, while making all collaboration materials and results available to the general public.

Create a hub of experts in molecular sciences committed to improving research software development and application practices by connecting all interested stakeholders — researchers, developers and funders — and providing organizational support.

How it works

https://omsf.io/about/mission/

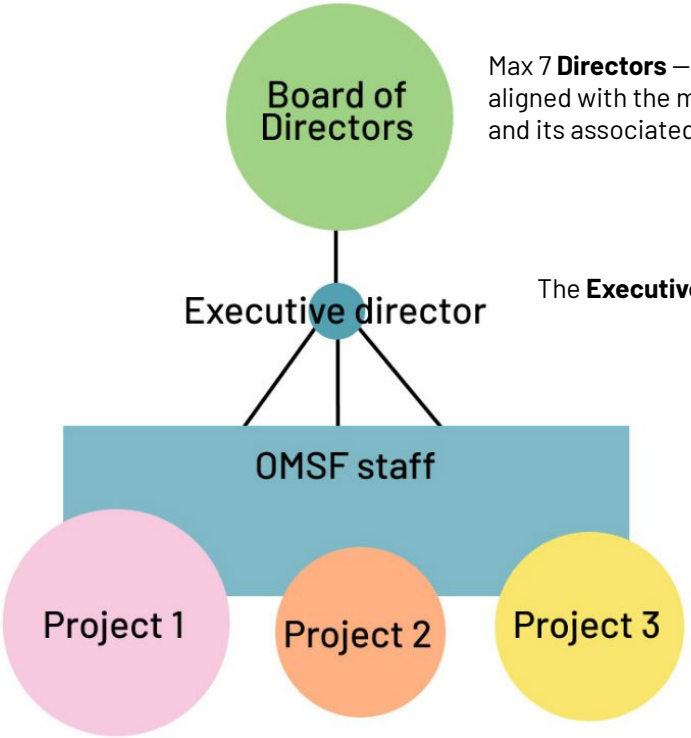
Fiscal sponsorship. OMSF provides an adminstrative home and operational base for open source software projects through fiscal sponsorship. Fiscal sponsors confer their legal and tax-exempt status to sponsored projects (groups). OMSF was founded to fill the need for better administrative, legal and project management support for distributed, cross-collaborative teams focused on (open source) research software development.



Community oriented programs. In addition to hosting specific software projects, OMSF plans to support other programs directed at advancing molecular sciences and research software development through education and training, creating paths to improved software interoperability and sustainability, better user support, and other relevant community activities and resources.

OMSF hosts open software projects

https://omsf.io/about/mission/



Max 7 **Directors** — 4 representing Hosted Projects + 3 "outside" Directors. BoD ensures that all activities are aligned with the mission and engages in strategic planning and decisions making in the best interest of OMSF and its associated stakeholders.

The **Executive Director** is the main link between the Board of Directors, staff and Hosted Projects.

OMSF's **staff** includes software engineers, project managers, etc. They support Hosted Projects, run the operations and implement strategy envisioned by the Board of Directors.

Each **Hosted Project** has their own governance model, and their **Governing Board** is responsible for running the Hosted Project and communicating with OMSF. OMSF may hire team members directly to contribute to each Hosted Project or provide fellowships and grants to academics in support of work relevant for the Hosted Project. Each Hosted Project may maintain a larger network of external collaborators and contributors.

Acknowledgements

Open Force Field Initiative researchers and software scientists
https://openforcefield.org/about/organization/

NIH R01 GM132386 and the Open Force Field Consortium,
plus MolSSI and others for fellowship funding

So many collaborators and contributors over the years

Getting involved: We need financial support, and benefit from advice and collaboration

- Three funding tiers (yearly): $100K, $50K, $20K; each with different benefits
- We need industry help
  - Prioritizing chemistry
  - Benchmarking
  - Identifying problems (& fixes)
- Our meetings are open
  - Get involved as much as you want
- **OpenFF creates an environment** where partners share with one another and us pre-competitively **to solve our common problems**