



 @openforcefield

 www.openforcefield.org

An overview of the Open Force Field project, project aims and achievements

Oct. 13, 2021 | David Mobley / AstraZeneca

Particular thanks to

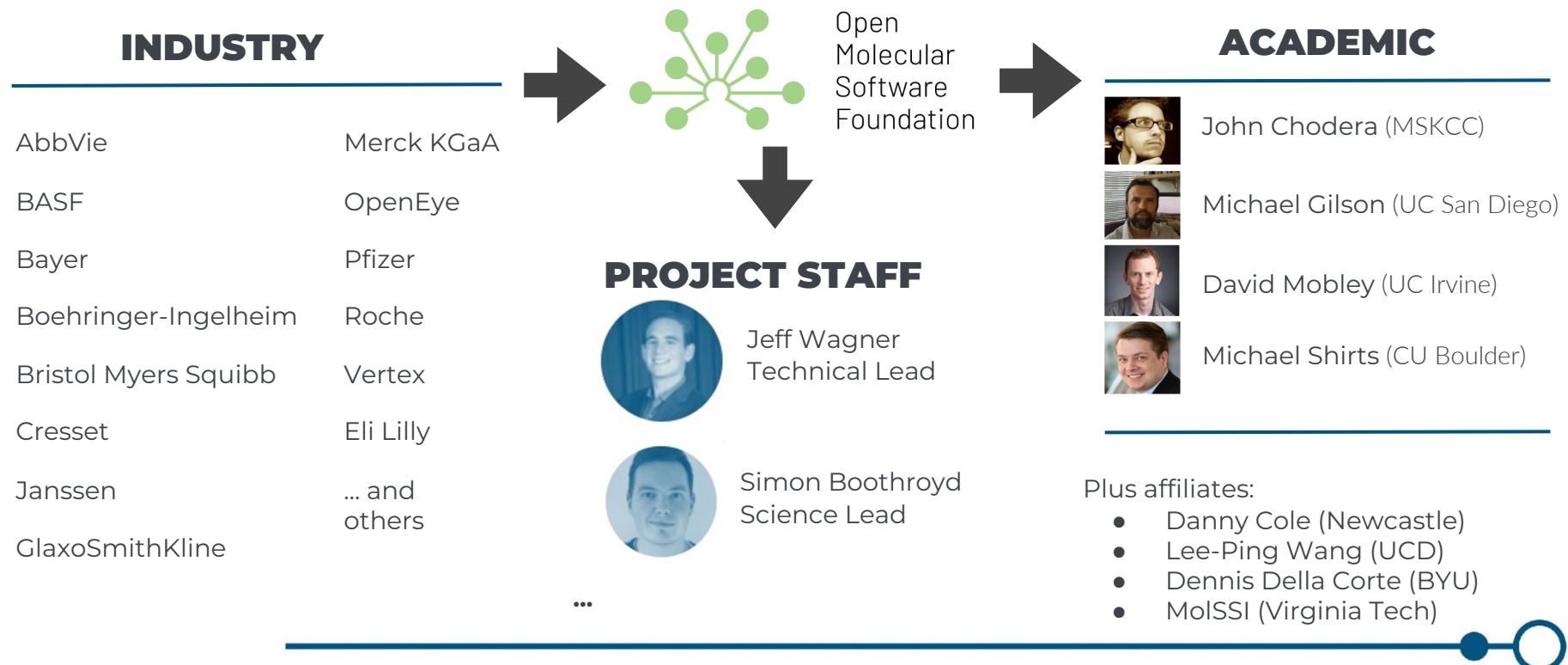


Simon Boothroyd

Jeff Wagner



The Open Force Field Consortium



How is the Open Force Field Initiative open?



Open source Python Toolkit: use the parameters in most simulation packages



Open curated QM / physical property datasets: build your own force fields



Open infrastructure: Run your own benchmarks; fit your own FFs



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

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

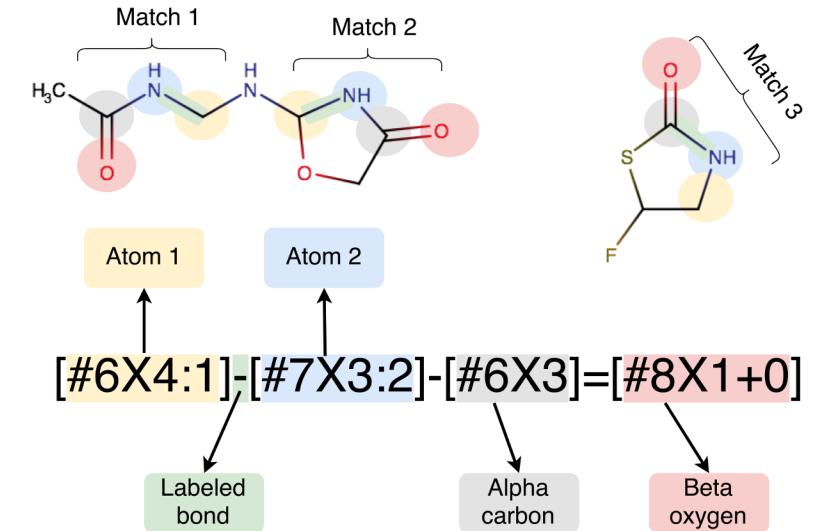
Documentation: Theory; toolkit documentation; tutorials and training materials

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



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

match bonds directly:



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

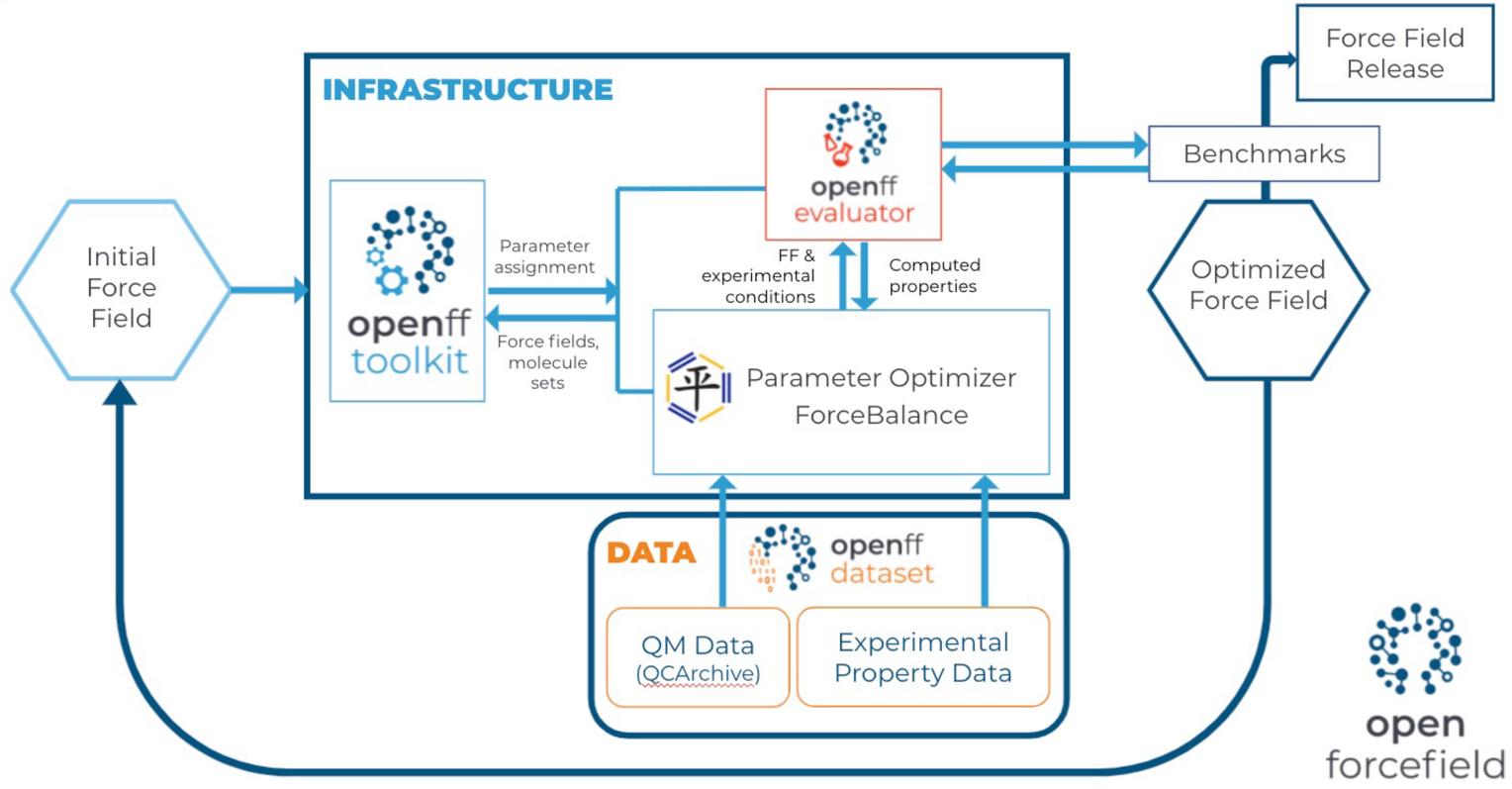
332

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



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)

- [1.0.0](#) (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

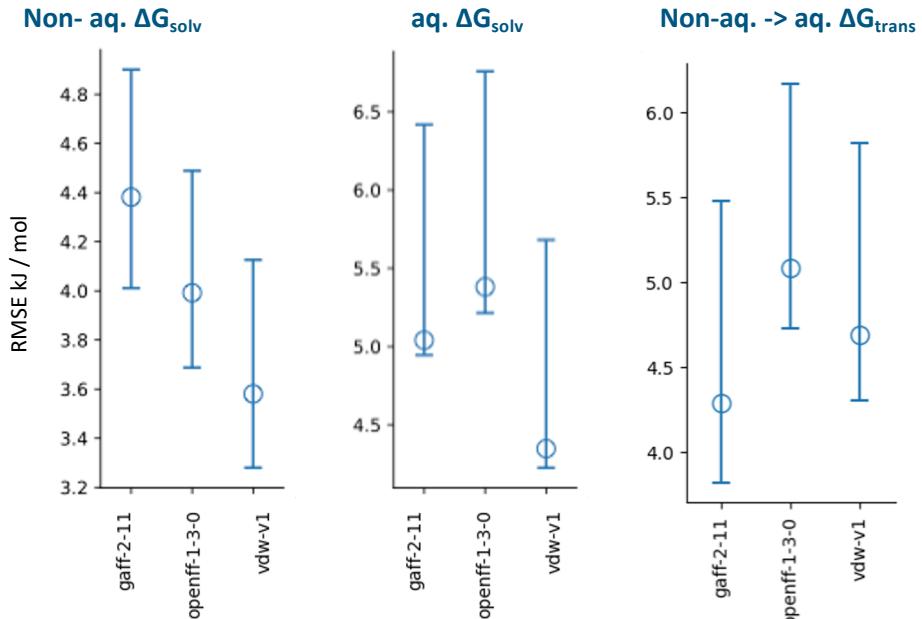
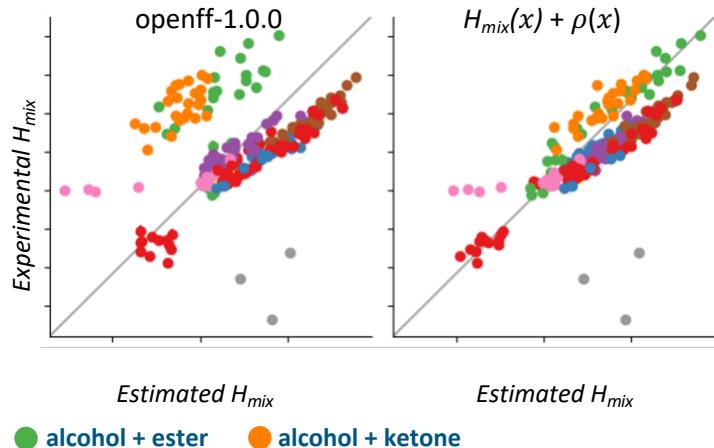


In parallel, tons of new science -- tens to hundreds of fitting experiments to test out a wide range of ideas, e.g. effect of vibrational frequency fitting

Our Sage release does better on mixture and condensed phase properties



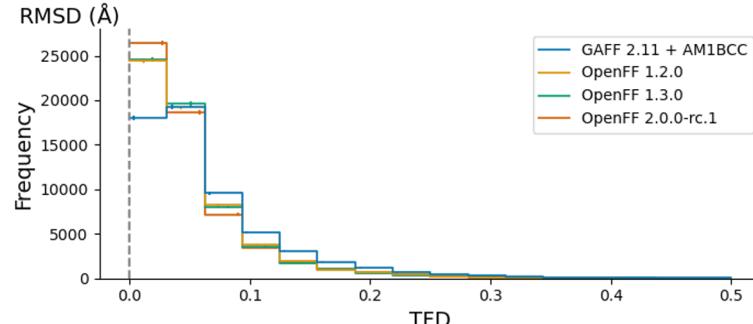
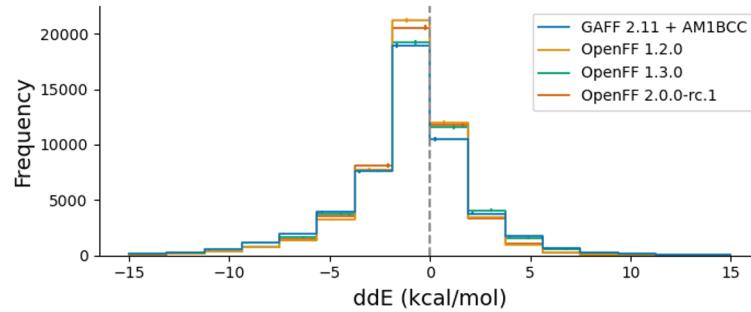
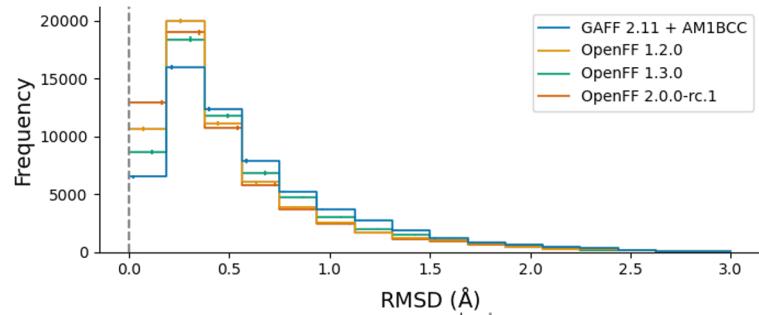
PREVIOUS STUDY SHOWS FITTING TO H_{MIX}
RESOLVES SYSTEMATIC ERRORS



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

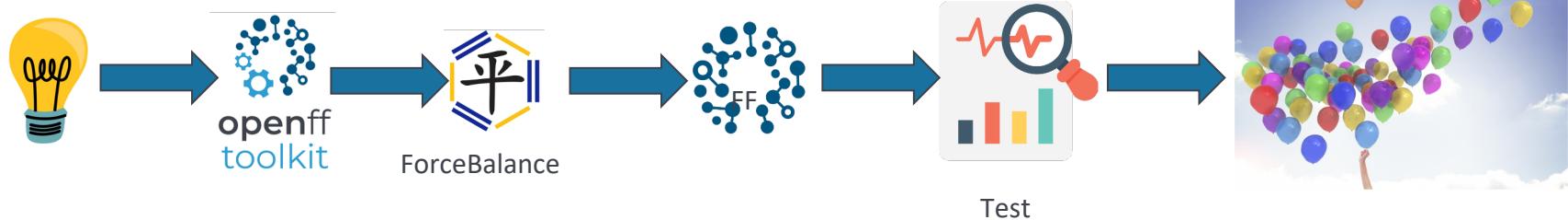


* 95% bootstrapped confidence intervals

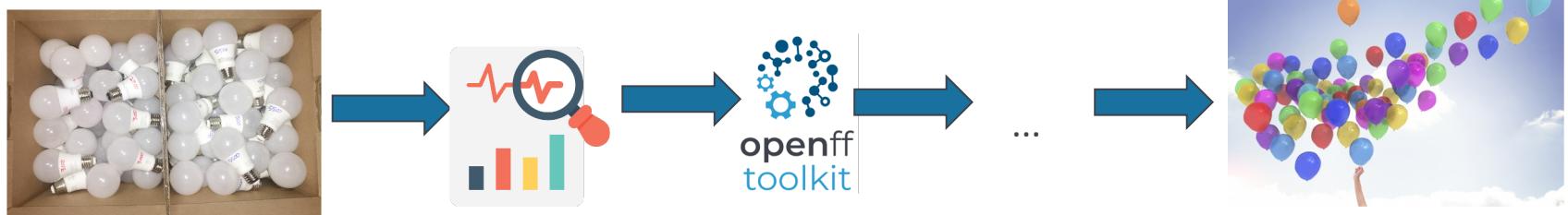
The process works, and it's changing how we think about building force fields



- 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





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!

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.

www.openforcefield.org



Search docs

Getting started on Cube and Floe development

OpenEye Orion Platform Packages Reference

OpenEye Orion Cubes and Floes Packages

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.

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 ☀

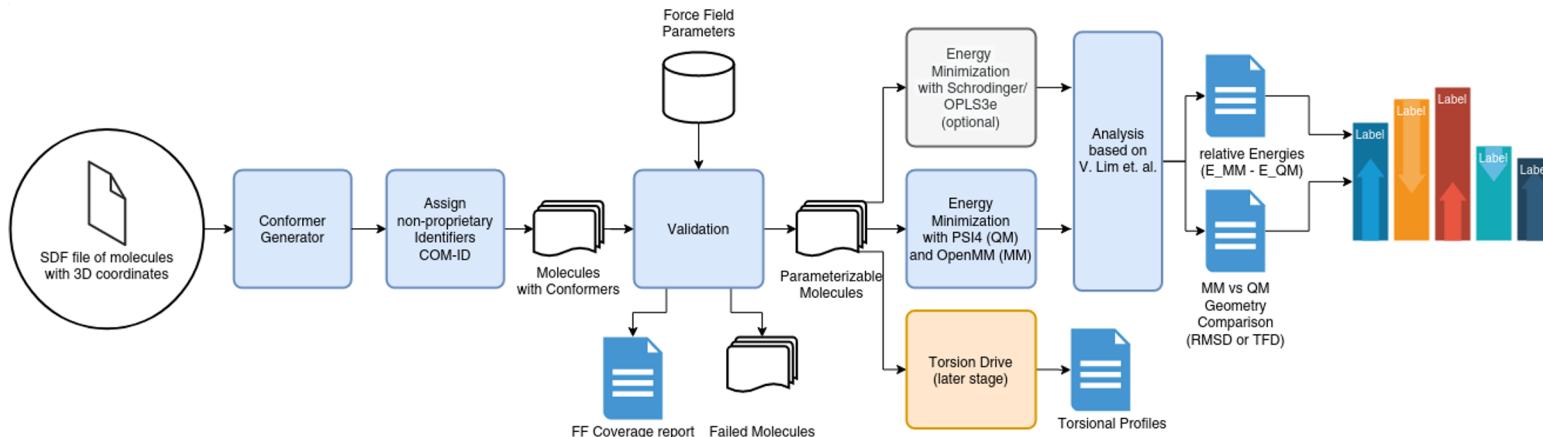
- OpenFF 1.3.0 and 1.2.1 support
- Bug Fixing





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

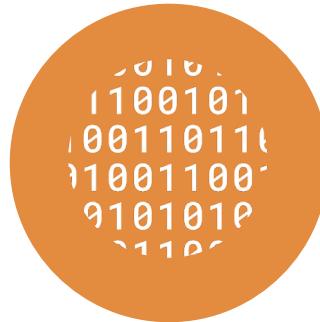


OPEN Software, OPEN Data, OPEN Science is rapidly facilitating force field science!



OPEN SOFTWARE

Automated infrastructure
enables rapid
experimentation with
minimum human
intervention



OPEN DATA

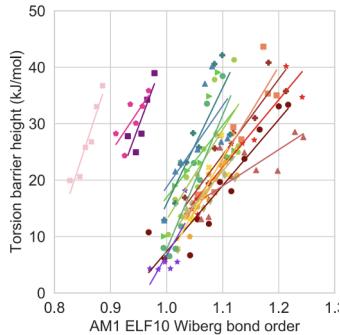
Access to large, high quality
experimental and quantum
chemical data facilities easy
curation of balanced train /
test sets



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



Chaya Stern

Jessica Maat

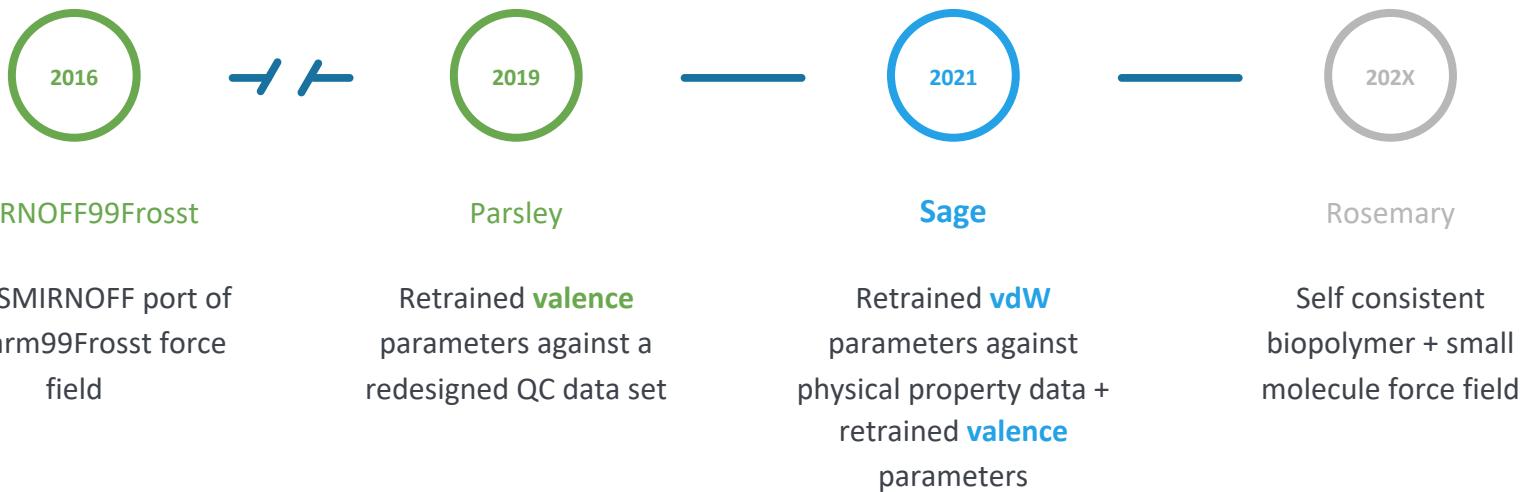
Pavan Behara



OpenFF Sage - the next INCREMENTAL generation of OpenFF force field



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

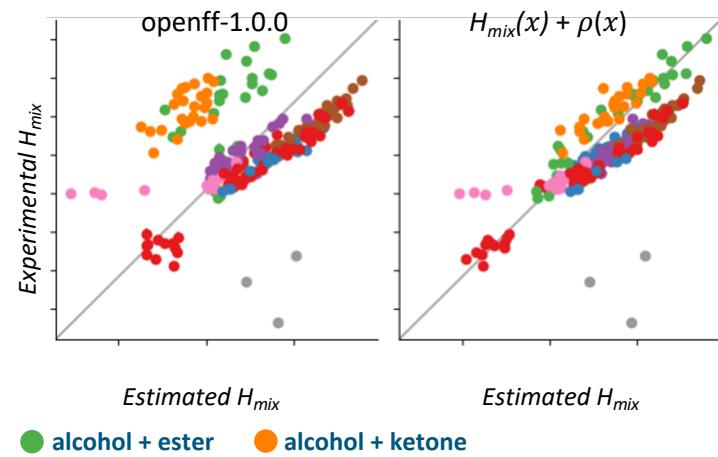


OpenFF Sage - select vdW parameters trained against experimental MIXTURE enthalpies and densities



- 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

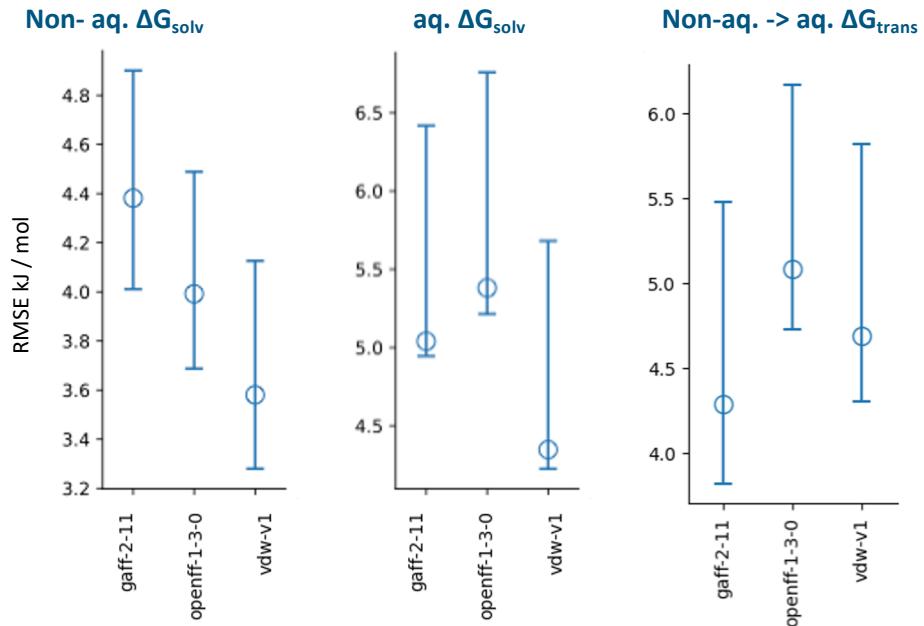


openff evaluator

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 MNSSol**
- 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 method** to derive force constants?
- Include all available data or filter to ensure balanced training set?
 - Filter optimized geometry to retain only **distinct** conformers



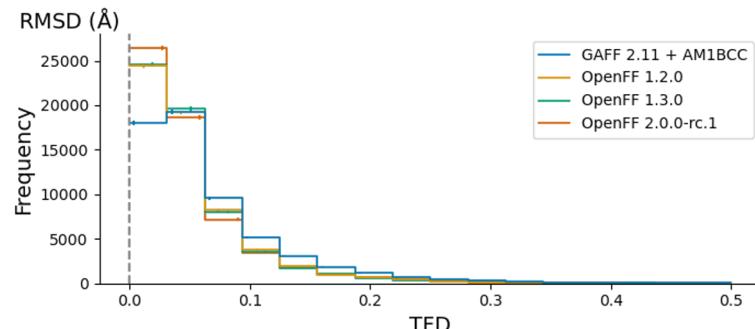
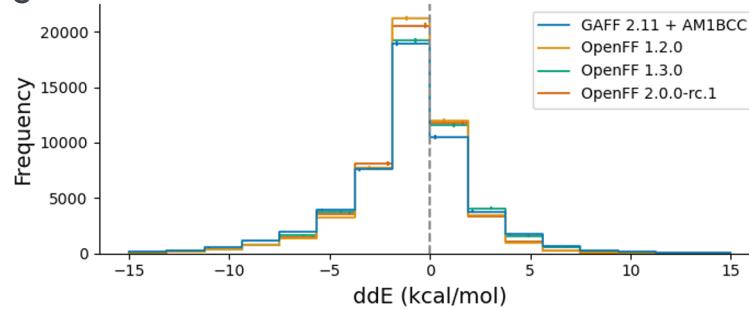
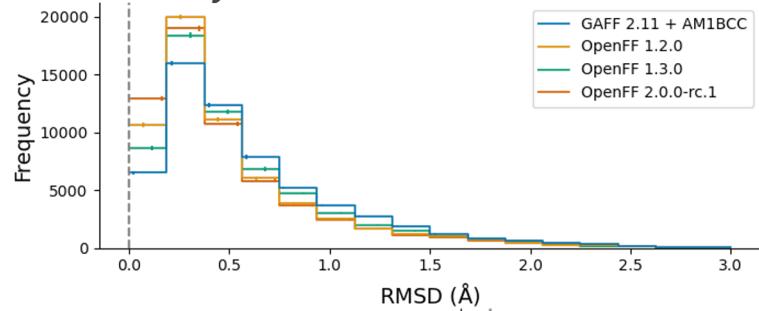
Hyesu Jang

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



- SAGE 2.0.0 showed excellent performance when benchmarked against the Public OpenFF

Industry Benchmark Season 1 v1.0

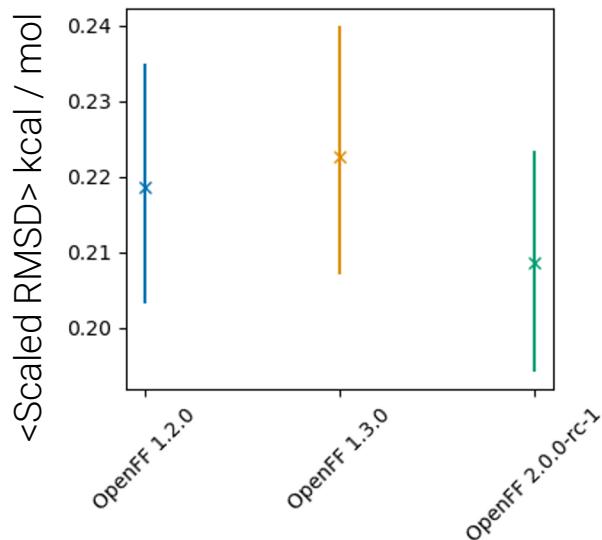
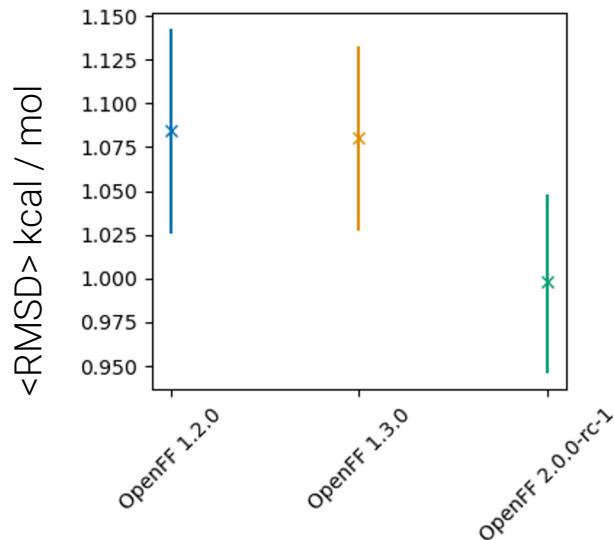


* 95% bootstrapped confidence intervals

OpenFF Sage - benchmarked torsion profiles appear to exhibit good performance



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



* 95% bootstrapped confidence intervals

The protein-ligand benchmark set consists out of 22 targets, 599 ligands and 1150 alchemical perturbations



Target	# Ligands	#Perturbations	Target	# Ligands	#Perturbations
jnk1	21	31	bace_p2	12	26
pde2	21	34	tyk2	16	24
thrombin	11	16	ros1	28	63
p38	34	56	eg5	28	65
ptp1b	23	49	cdk8	33	54
galectin	8	7	hif2a	42	92
cdk2	16	25	pfkfb3	40	66
cmet	24	74	pde10	35	36
mcl1	42	71	shp2	26	56
bace	36	58	syk	44	101
bace_hunt	32	60	tnks2	27	60
			total	599	1150

<https://github.com/openforcefield/protein-ligand-benchmark>

www.openforcefield.org

Content/preparation of Benchmark set might change in the future.

- *What is the purpose for changing (better quality, new chemistries, ...)?*
- *How do we ensure that we always compare calculations with the same input structures?*

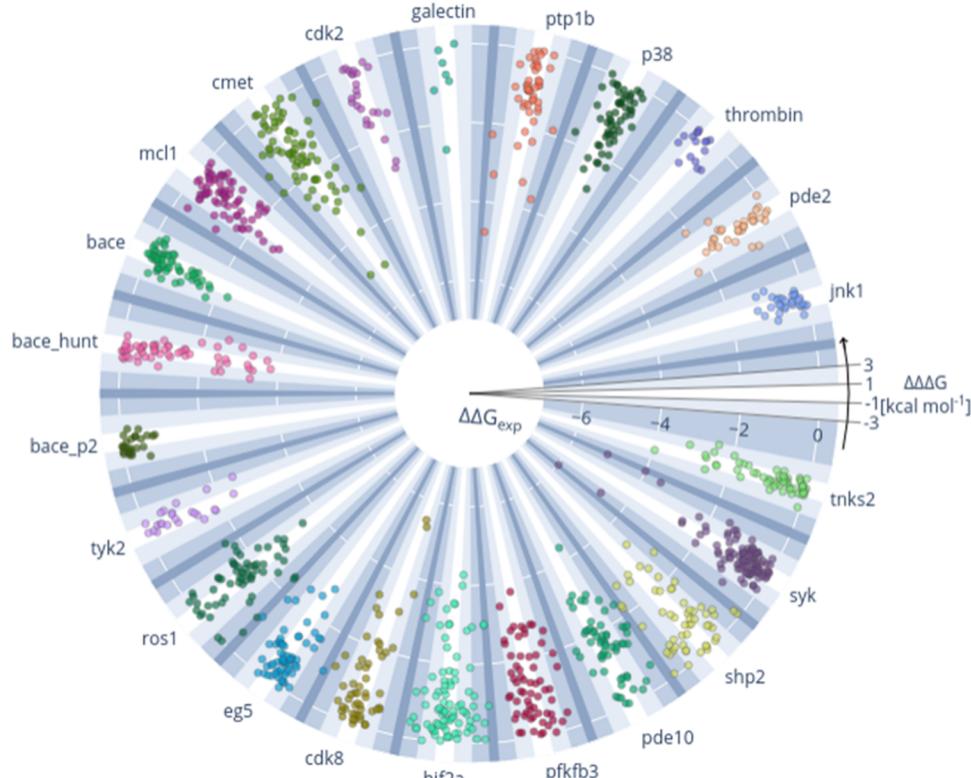
Sources

- Schrodinger JACS
[L. Wang et al., J. Am. Chem. Soc., 2015, 137, 2695 —2703.](#)
- Merck KgGA
[Christina E. M. Schindler et al., Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects, J. Chem. Inf. Model. 2020, 60, 11, 5457-5474](#)
- Janssen/Other:
[V. Gapsys et al., Large scale relative protein ligand binding affinities using non-equilibrium alchemy, Chem. Sci., 2020, 11, 1140-1152](#)
[Laura Perez Benito et al., Predicting Activity Cliffs with Free-Energy Perturbation, J. Chem. Theory Comput. 2019, 15, 3, 1884-1895](#)
- Best practices:
[Hahn DF et al., Best practices for constructing, preparing and evaluating protein-ligand binding affinity benchmarks.arXiv:210506222 \[physics.q-bio,2021 May\]](#)

Results for all perturbations using pmx and Parsley (OpenFF 1.0.0)



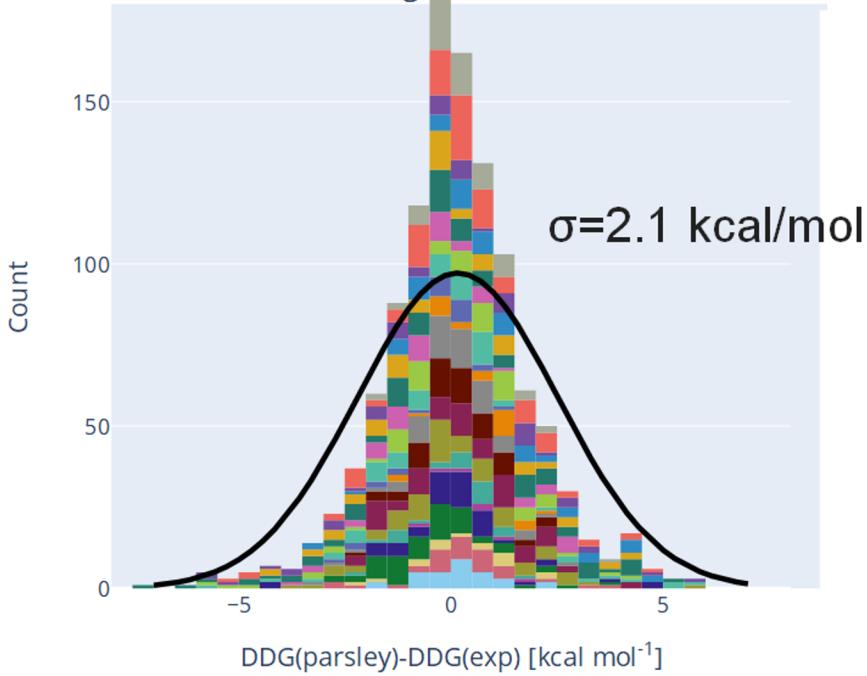
- Overview over all calculations performed
- Radial: exp. $\Delta\Delta G$ in kcal/mol
- Polar: difference between calc. and exp. $\Delta\Delta G$, $\Delta\Delta\Delta G$ in kcal/mol



Most perturbations ($\Delta\Delta G$) deviate less than 1 kcal/mol from experiment



Different colors denote different targets



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

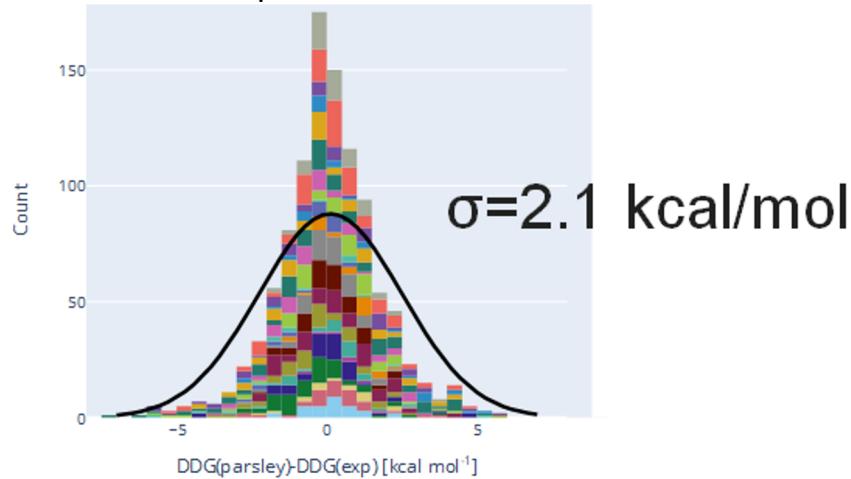
Origin of errors:

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

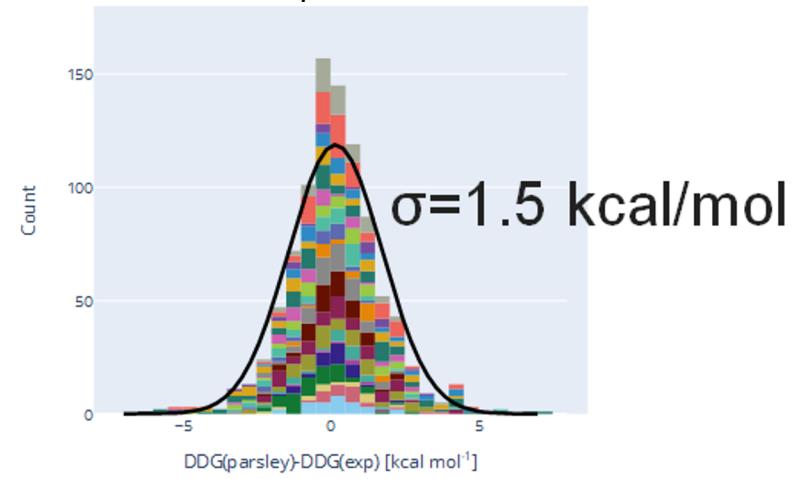
Results can be filtered based on different convergence metrics



All perturbations



Filtered 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

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

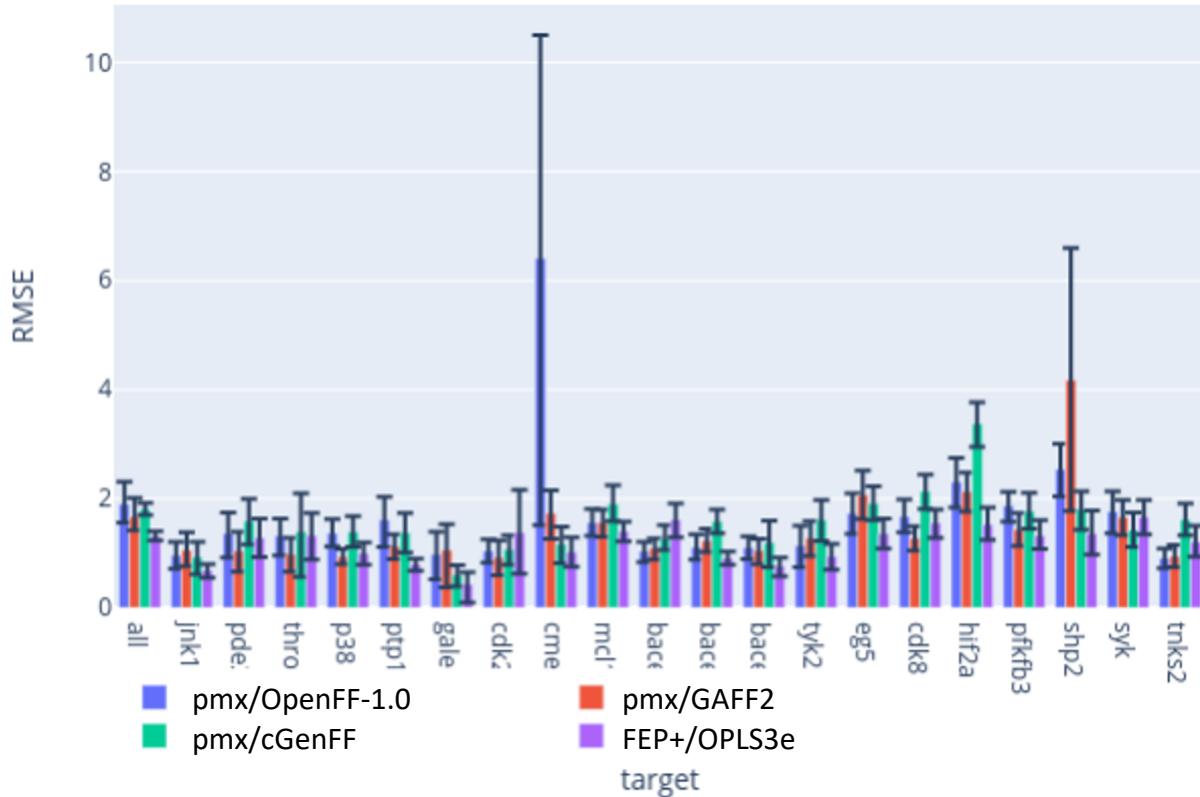
OpenFF-1.0 offers similar performance as other force fields



- RMSE based on $\Delta\Delta G$ in kcal/mol
- Error bars are 95% CI
- OPLS3e is generally slightly, but non-significantly better



Vytas Gapsys



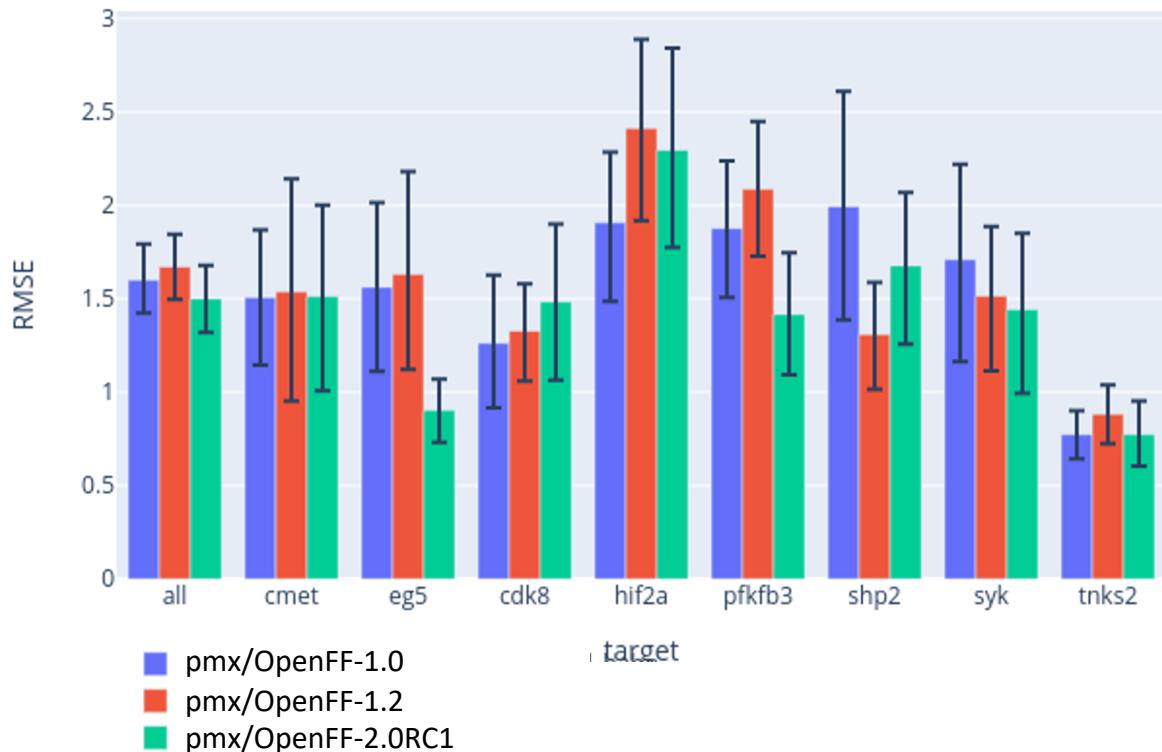
OpenFF-2.0-RC1 improved results over OpenFF-1.0



- RMSE based on $\Delta\Delta G$ in kcal/mol
- Error bars are 95% CI
- OpenFF2.0RC1 is generally slightly, but non-significantly better



Vytas Gapsys



OpenFF Sage - the next INCREMENTAL generation of OpenFF force field



- Sage 2.0.0 available now on GitHub in the [openff-sage](#) repository



SMIRNOFF99Frosst



Parsley



Sage



Rosemary

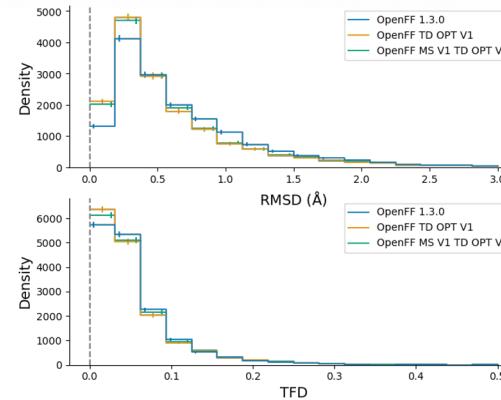
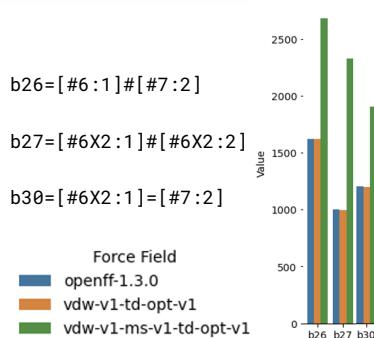
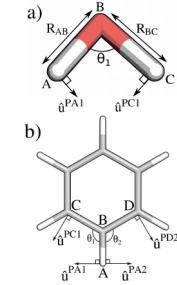
Initial SMIRNOFF port of
the parm99Frosst force
field

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



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

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



33

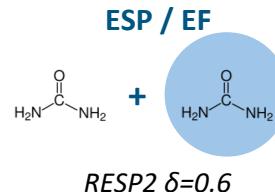
Double-bonded oxygen in a lactone or lactam



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



openff
evaluator



Mixture enthalpies +
densities

AM1BCC Ported to
SMIRNOFF



Integrate Into Fitting
Infrastructure



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

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

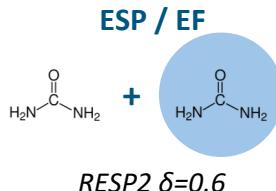
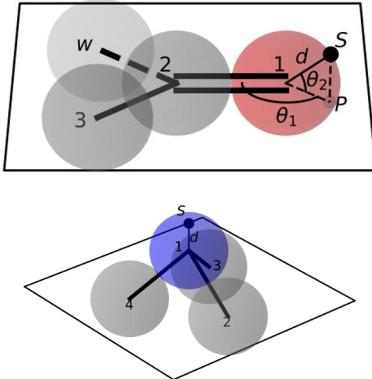
Test fits being performed
against a combination of
QC ESP / EF data and
mixture exp. data

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



Trevor Gokey

Hypothesis

The inclusion of off-site charges should improve the accuracy of a force fields electrostatic interactions

Software

Virtual site support added to the OpenFF toolkit. Support for training to QC ESP + EF data in progress

Training

Virtual sites will be trained against ESP / EF QC data, based on input from the Cole group

Testing

Trained parameters will be benchmarked against experimental and physical property data

Interoperability

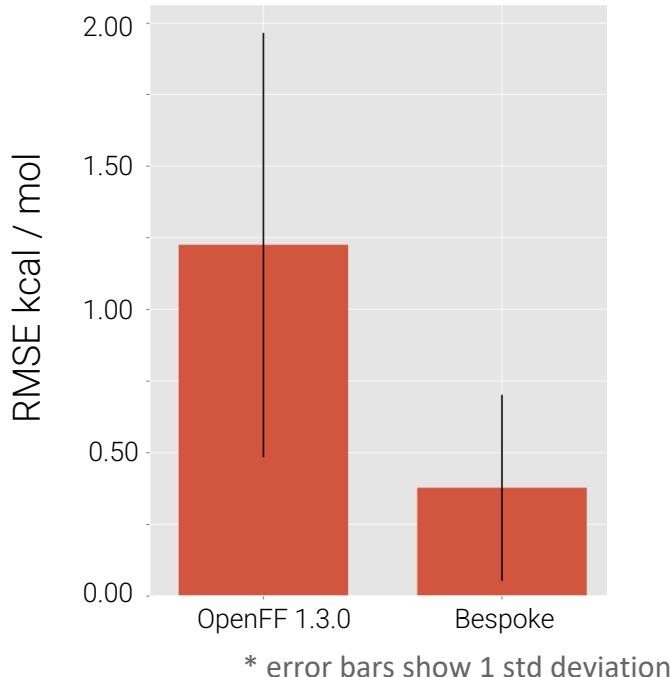
In order to include into mainline force field need major simulation packages to support proposed v-sites

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



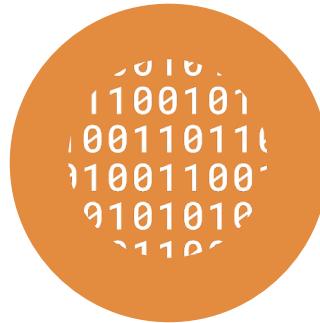
Josh Horton

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 facilitates easy curation of balanced train / test sets



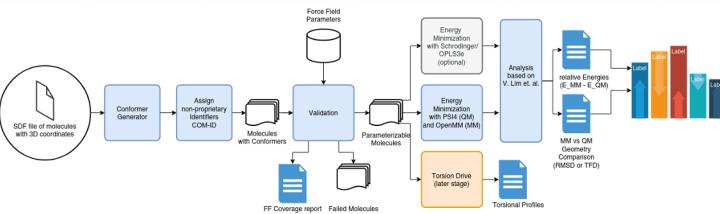
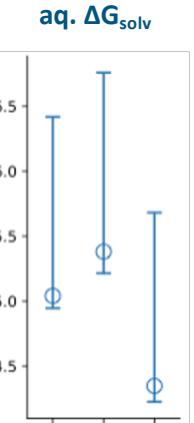
OPEN SCIENCE

Exploring new force field science:
*hypothesis - build
software - train - test -
iterate*
is now almost routine

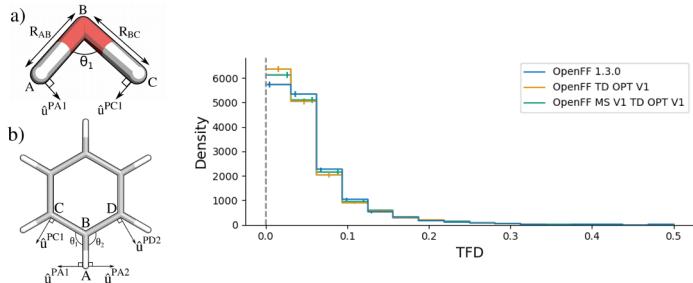
Conclusions



Sage looks even better than we expected!



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



New technology yields better starting points and better FFs

www.openforcefield.org

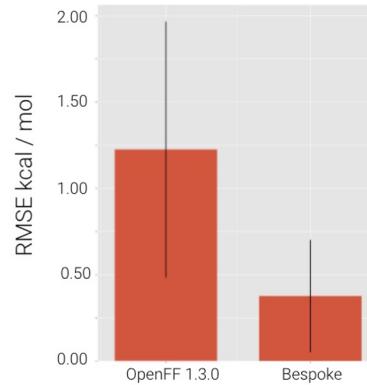
PELE Force Field Yielder



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.

We see community uptake, with and without our help



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

OpenEye for support + Chris Bayly's sabbatical which helped launch this