Workshop Schedule

- Presentations (45 minutes)
  - Protein-ligand benchmarking
  - Small molecule conformer benchmarking
- Q&A Time
- Break (5 min)
- Interactive session (45 minutes)
- Q&A Time

open
forcefield

@openforcefield

www.openforcefield.org

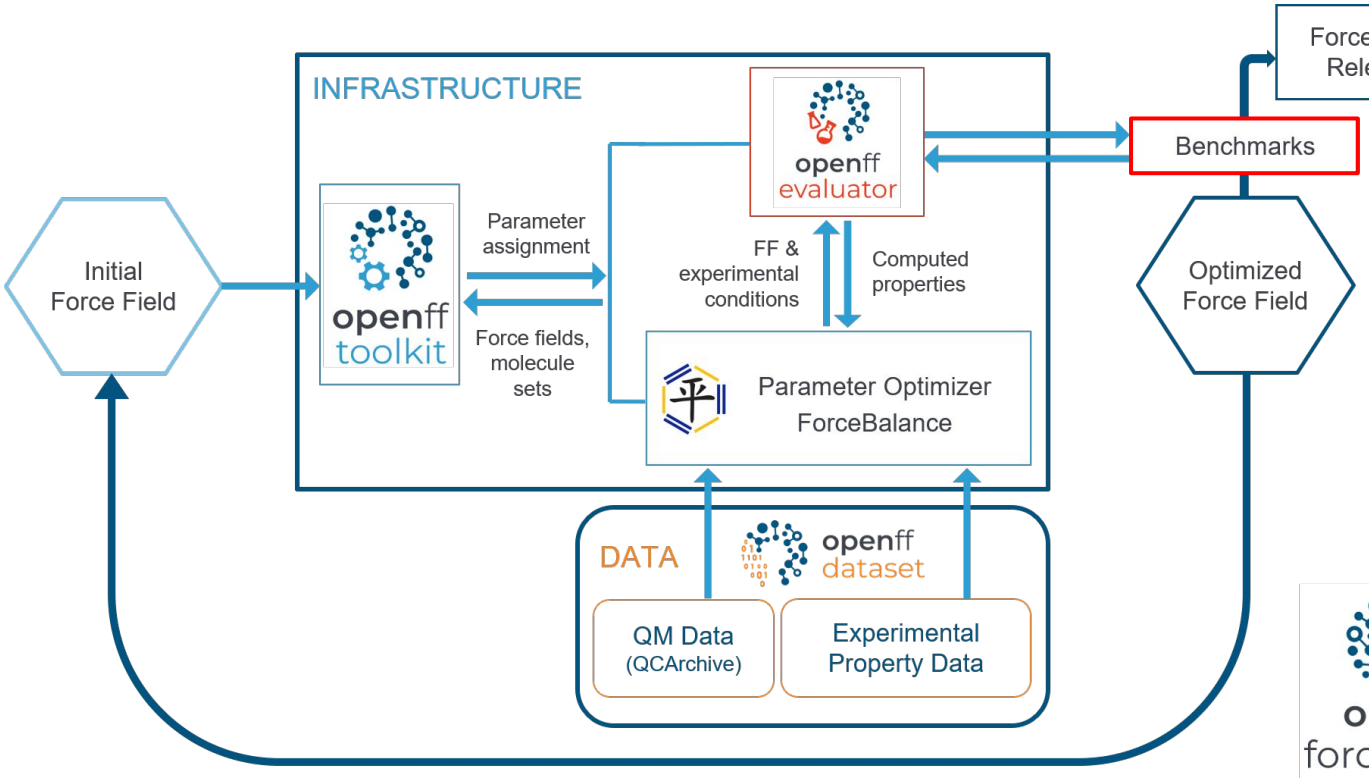
5th open force field follow-up workshop

Sept. 1, 2021 | Benchmarking + interactive session

Workshop Schedule

- Presentations (45 minutes)
  - Protein-ligand benchmarking
  - Small molecule conformer benchmarking
- Q&A Time
- Break (5 min)
- Interactive session (45 minutes)
- Q&A Time

How the Open Forcefield framework is drawn up



- **Geometry optimization benchmarks**
- **Protein-ligand binding FE benchmarks**

open forcefield

Large scale benchmarking of force fields in protein-ligand free energy calculations

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

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

➔ *What is the purpose for changing (better quality, new chemistries, …)?*

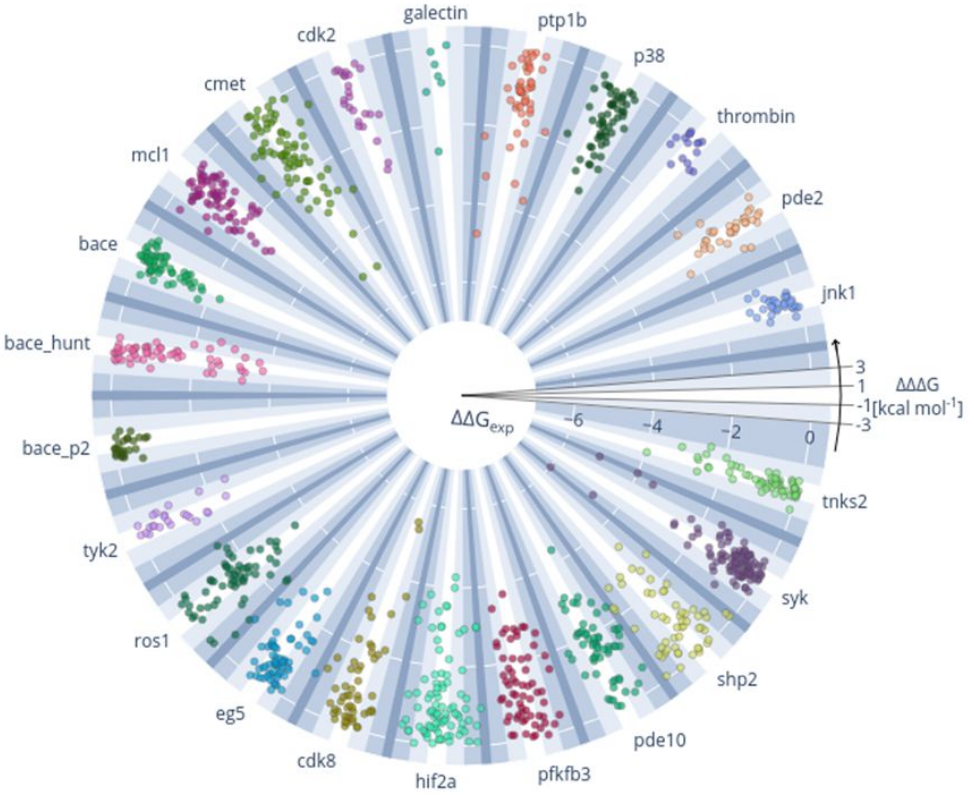
➔ *How do we ensure that we always compare calculations with the same input structures?*

Sources

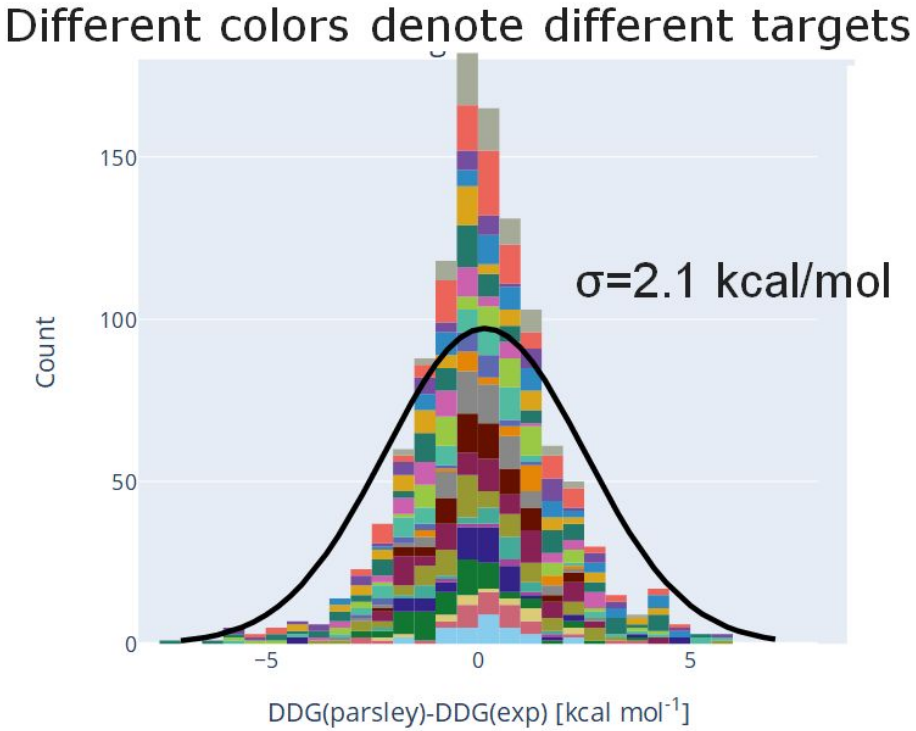
- Schrodinger JACS
  L. Wang et al. , *J. Am. Chem. Soc.*, 2015, **137** , 2695 —2703.
- Merck KgGA
  Christina E. M. Schindler et al., Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects, J. Chem. Inf. Model. 2020, 60, 11, 5457–5474
- Janssen/Other:
  V. Gapsys 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. ΔΔG in kcal/mol
- Polar: difference between calc. and exp. ΔΔG, ΔΔΔG in kcal/mol



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



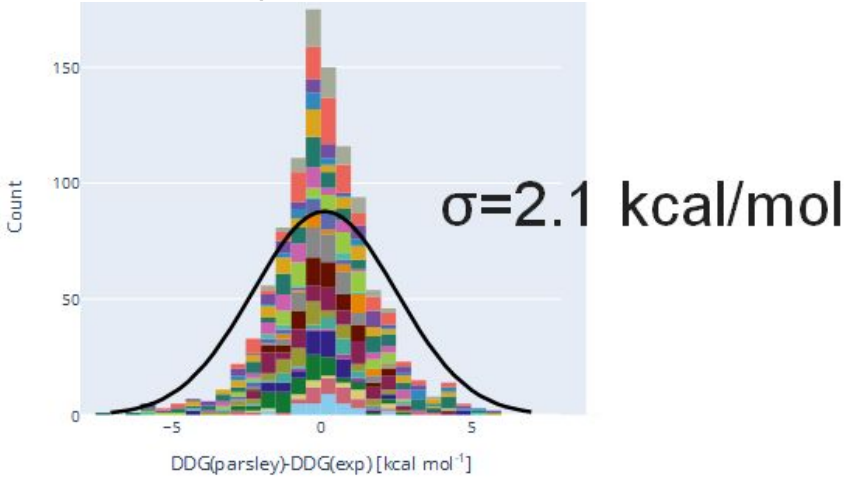
Abs. Error [kcal/mol]	# Perturbations	% of total
< 0.5	322	29
<1.0	592	52
<2.0	911	79
<3.0	1052	92
total	1149	100

Origin of errors:

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

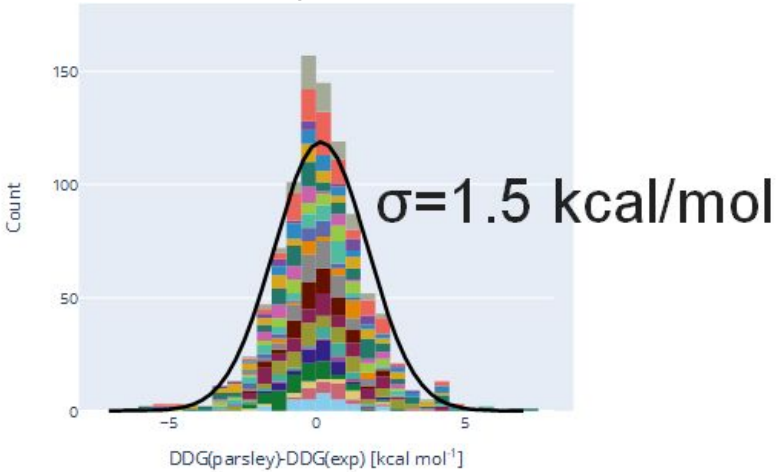
Results can be filtered based on different convergence metrics

All perturbations



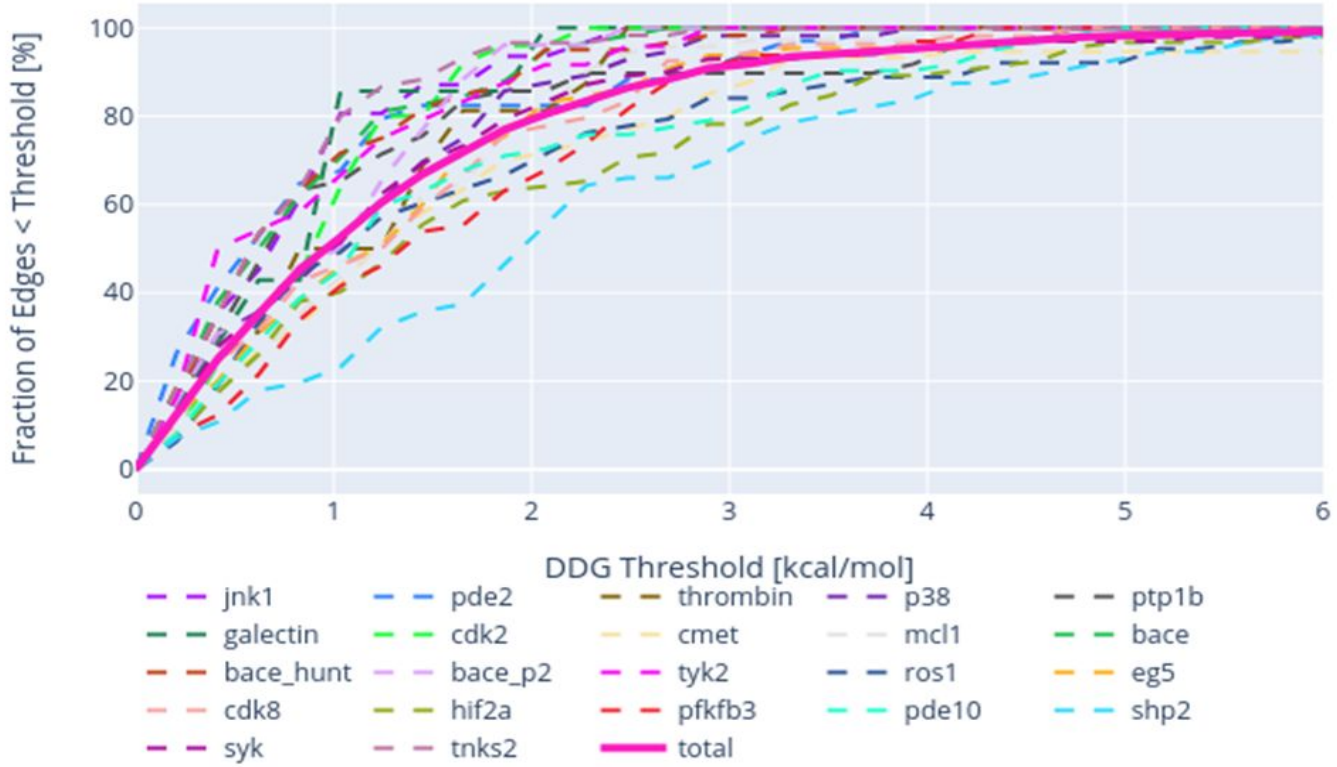
Abs. Error [kcal/mol]	# Perturbations	% of total
< 0.5	322	29
<1.0	592	52
<2.0	911	79
<3.0	1052	92
total	1149	100

Filtered perturbations



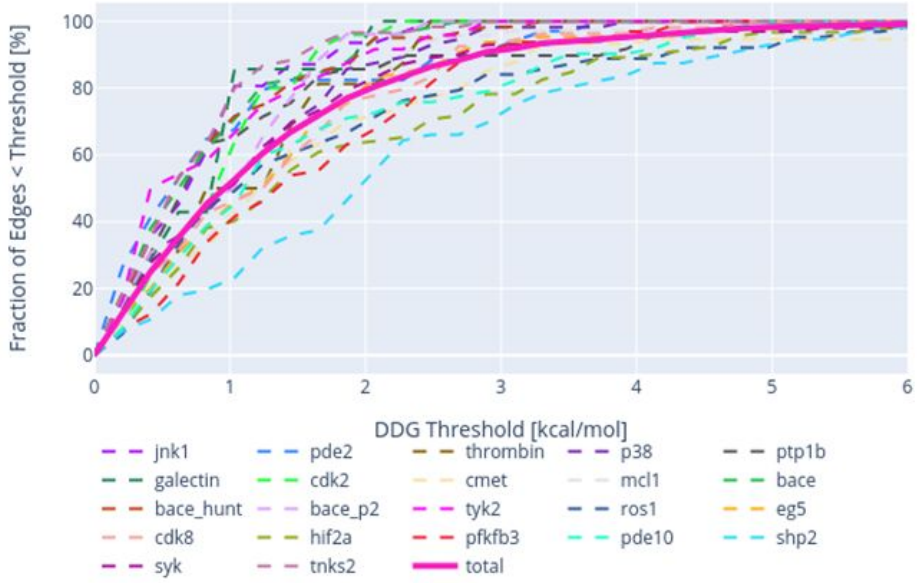
Abs. Error [kcal/mol]	# Perturbations	% of total
< 0.5	383	32
<1.0	508	57
<2.0	748	85
<3.0	835	94
total	885	100

Number of outliers vary strongly per target

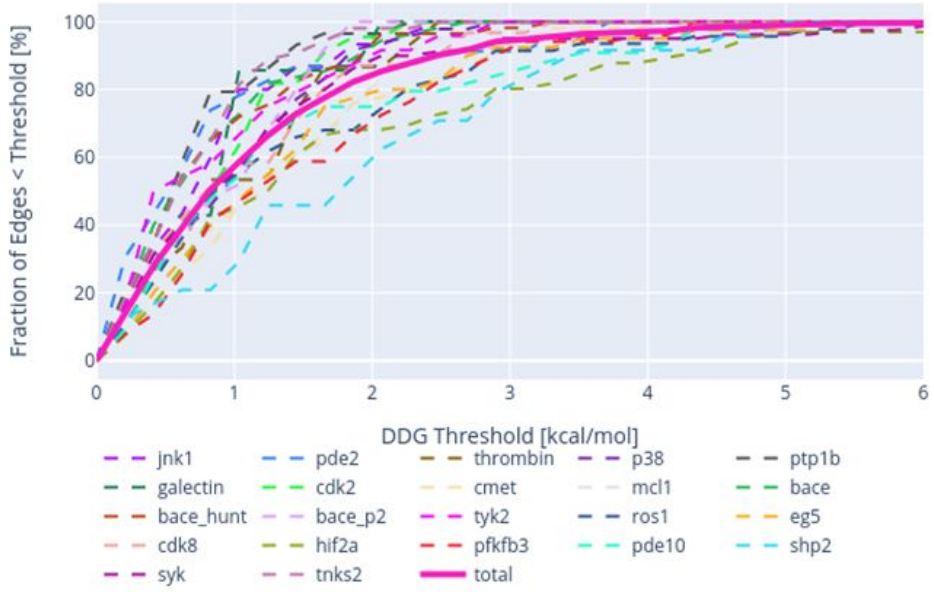


Filtering reduces outliers and are a good diagnostic tool

All perturbations

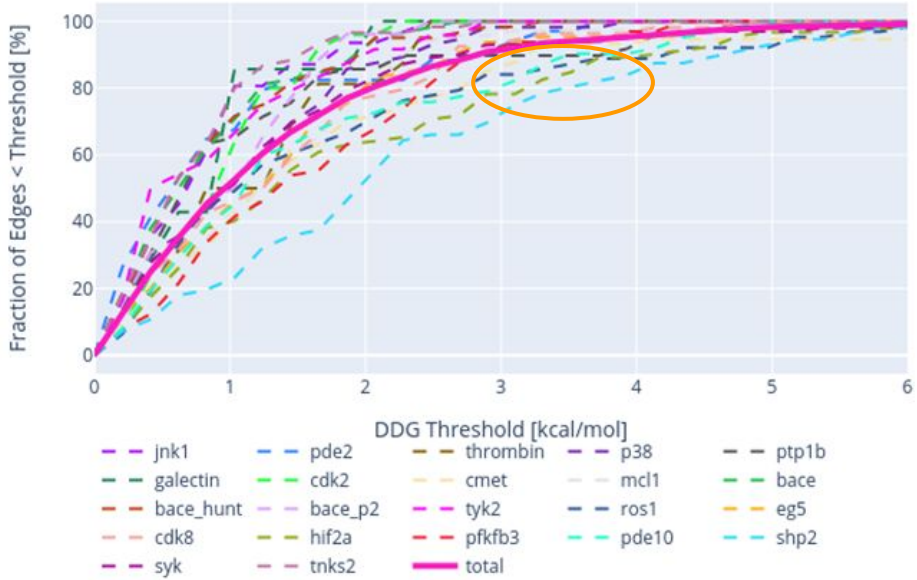


Filtered perturbations

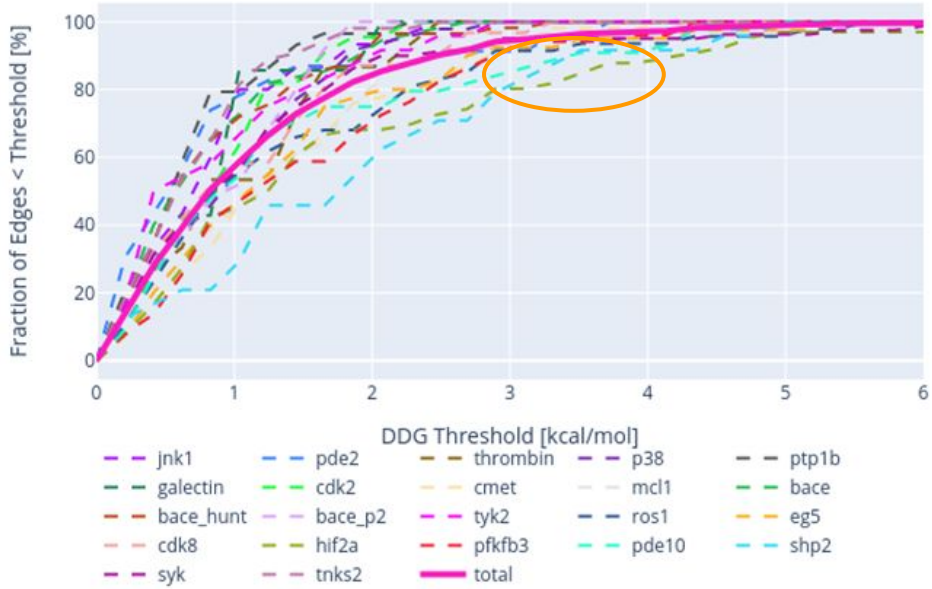


Filtering reduces outliers and are a good diagnostic tool

All perturbations



Filtered perturbations



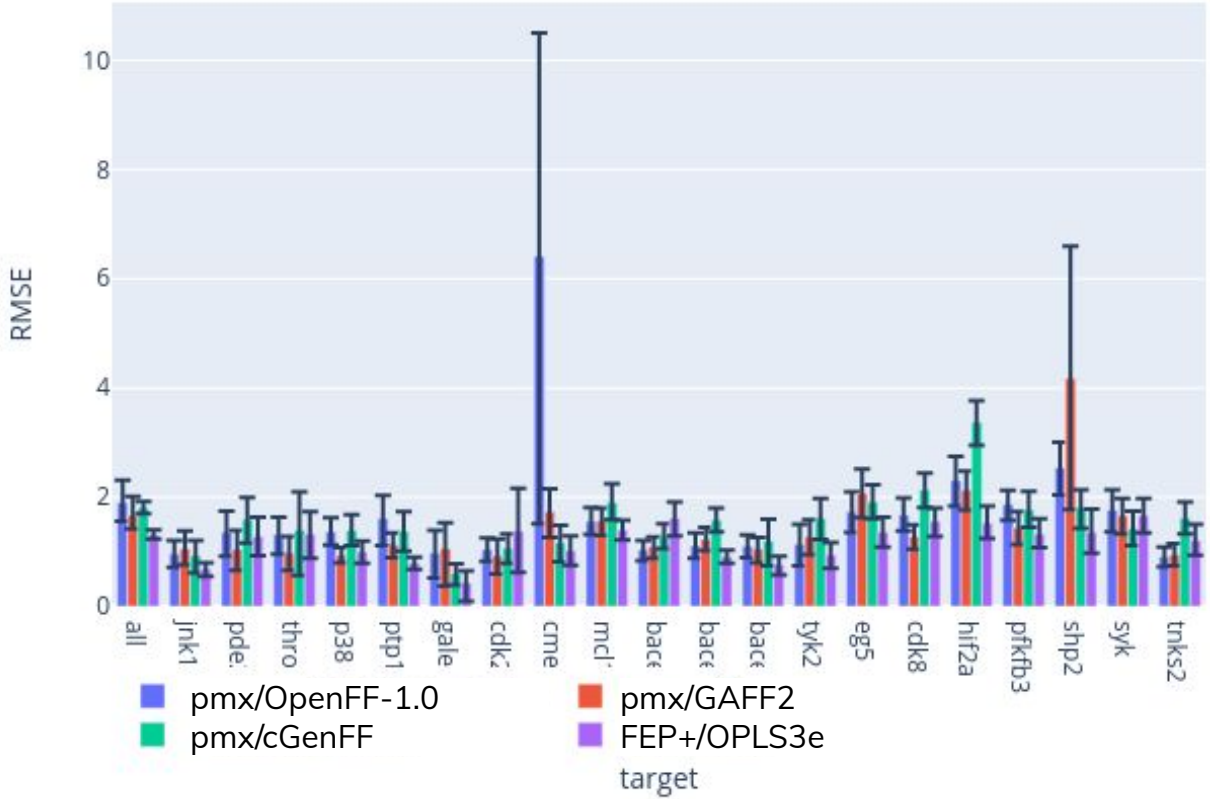
HIF2a: set-up error (wrong poses, tautomers, charges, FF parameters)

SHP2: convergence error (too short simulation time, too difficult perturbations)

OpenFF-1.0 offers similar performance as other force fields

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

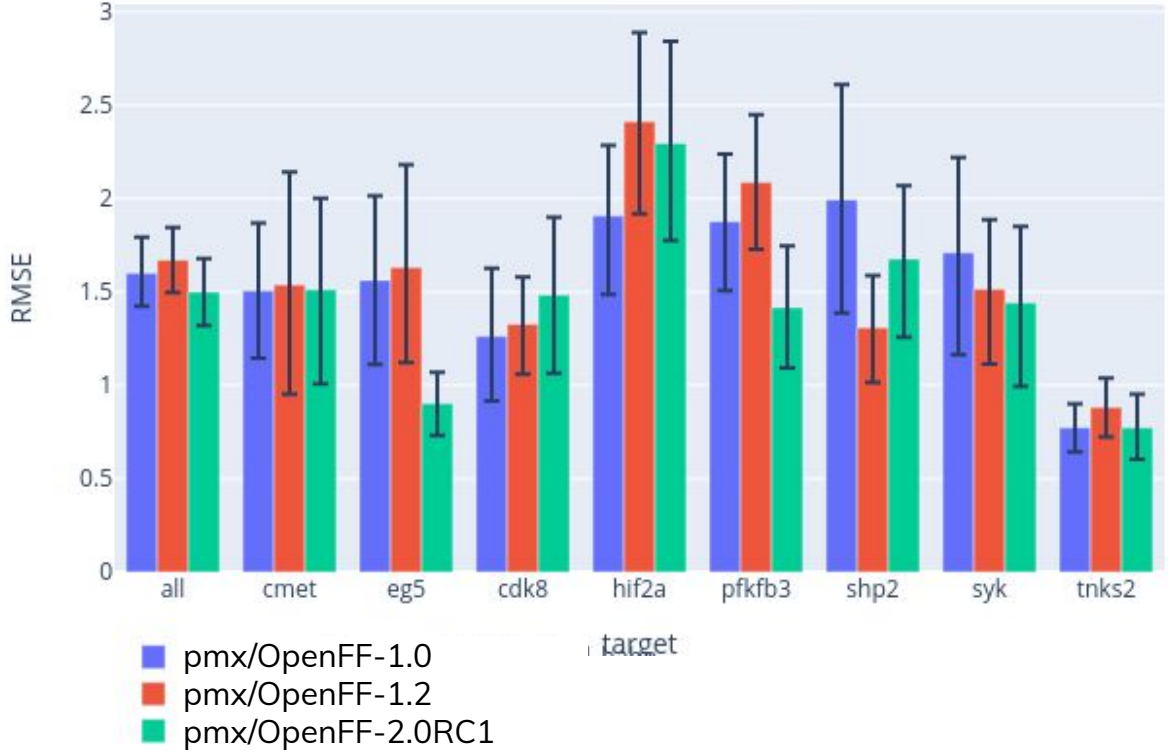
Vytas Gapsys



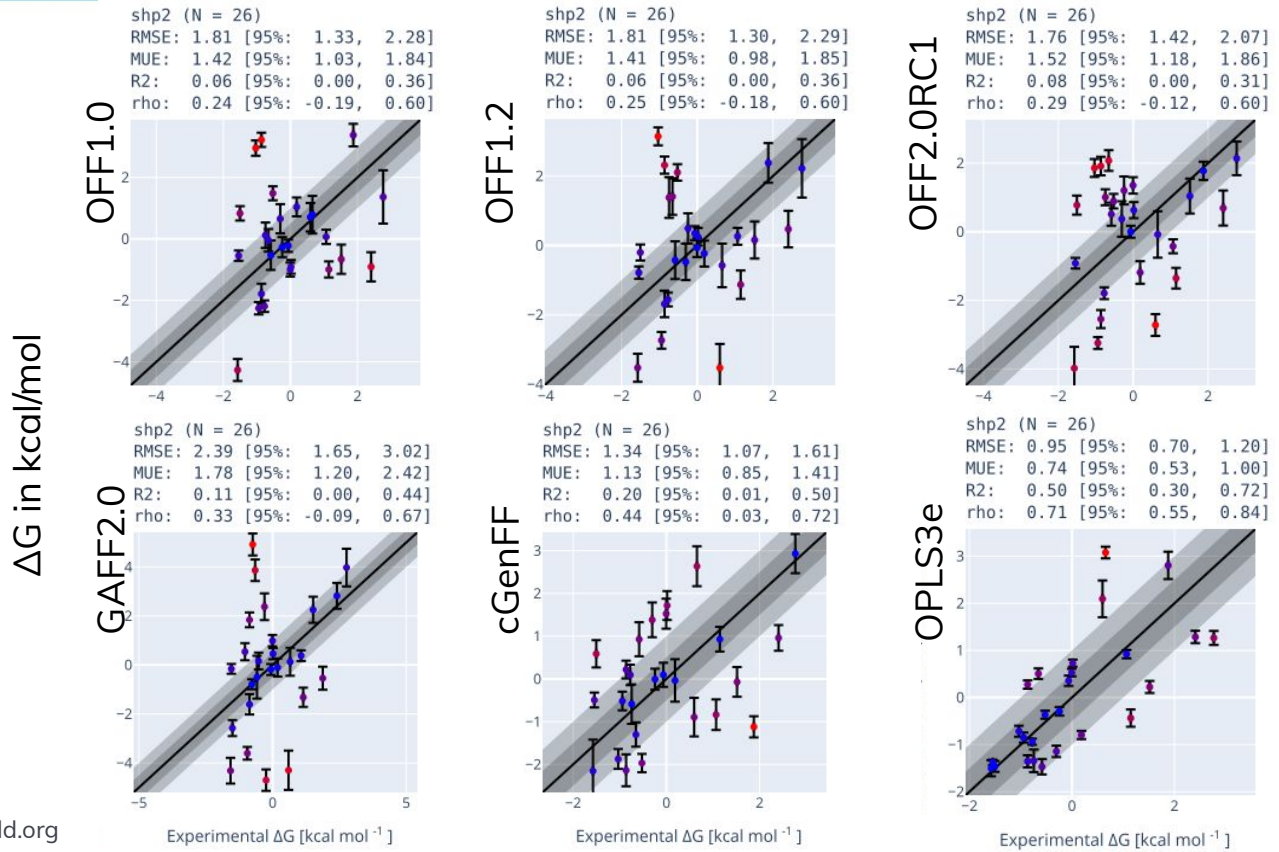
OpenFF-2.0-RC1 improved results over OpenFF-1.0

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

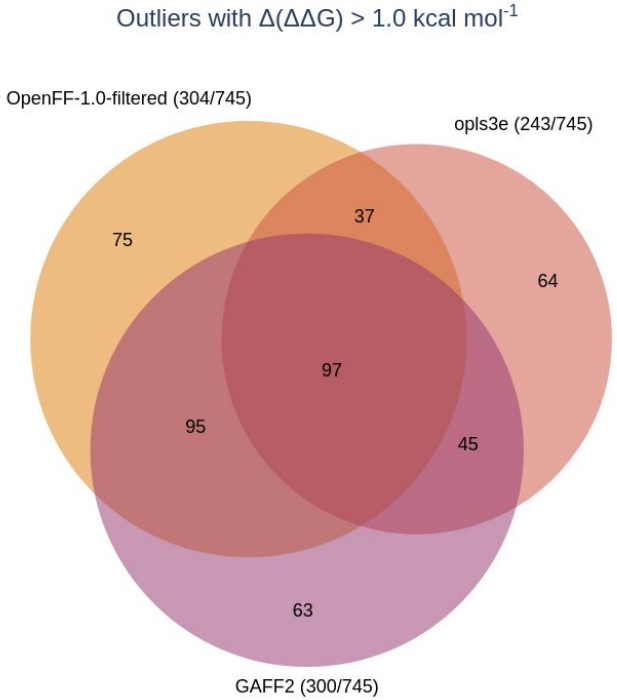
Vytas Gapsys

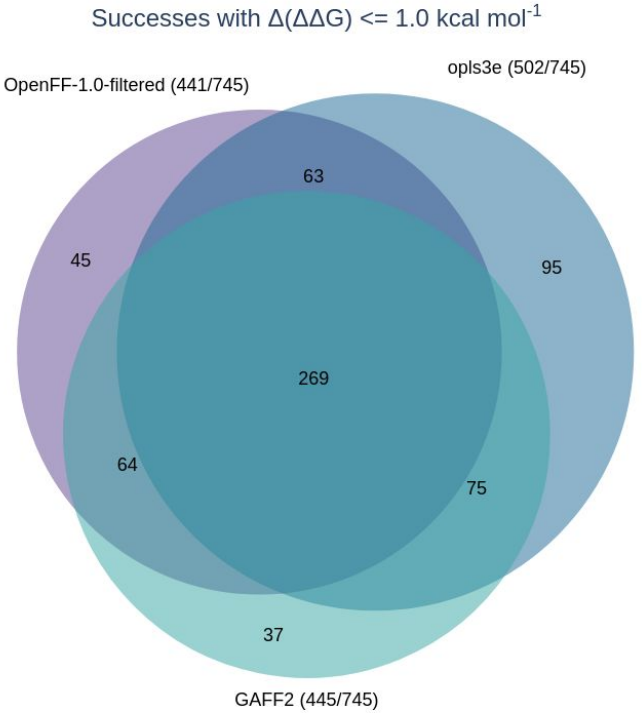


ΔG statistics might give a different perspective

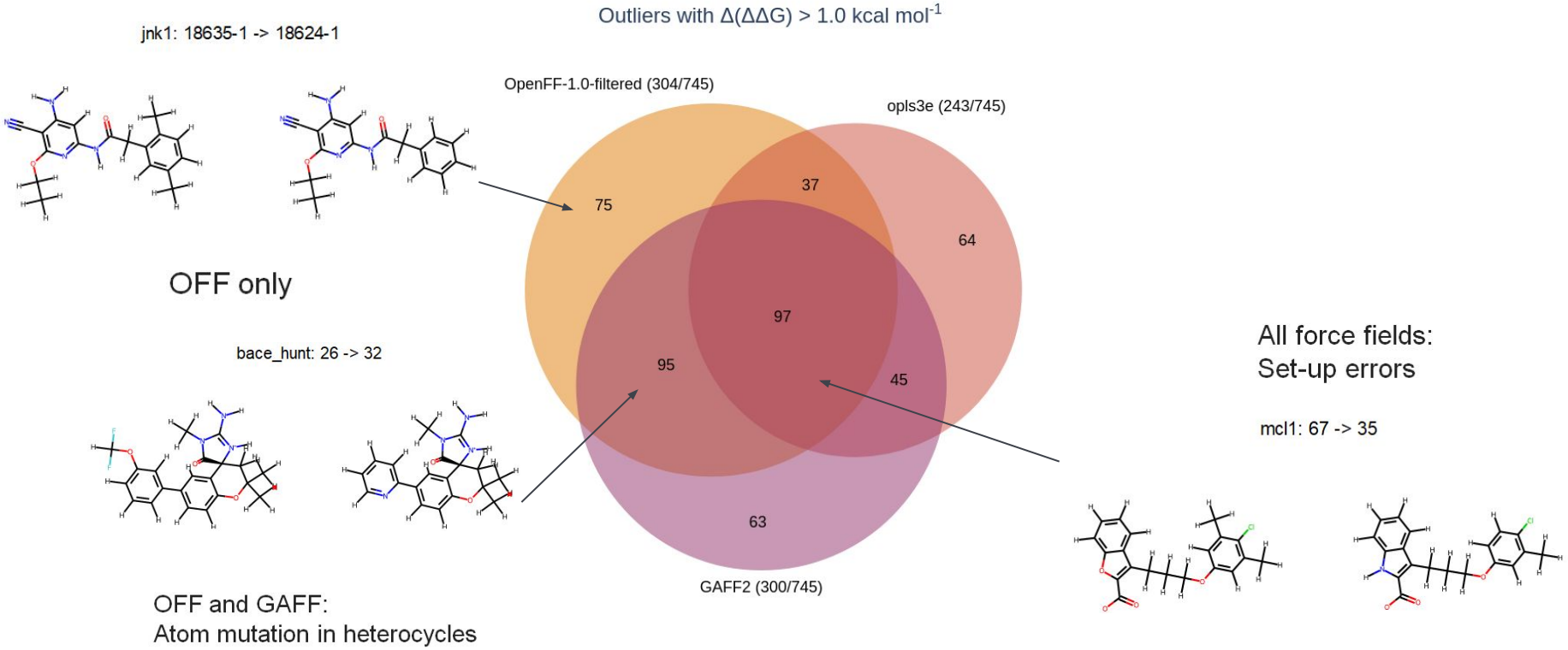


Comparing the calculated sets to experiments shows that outliers are specific to force fields





Comparing the calculated sets to experiments can help finding issues

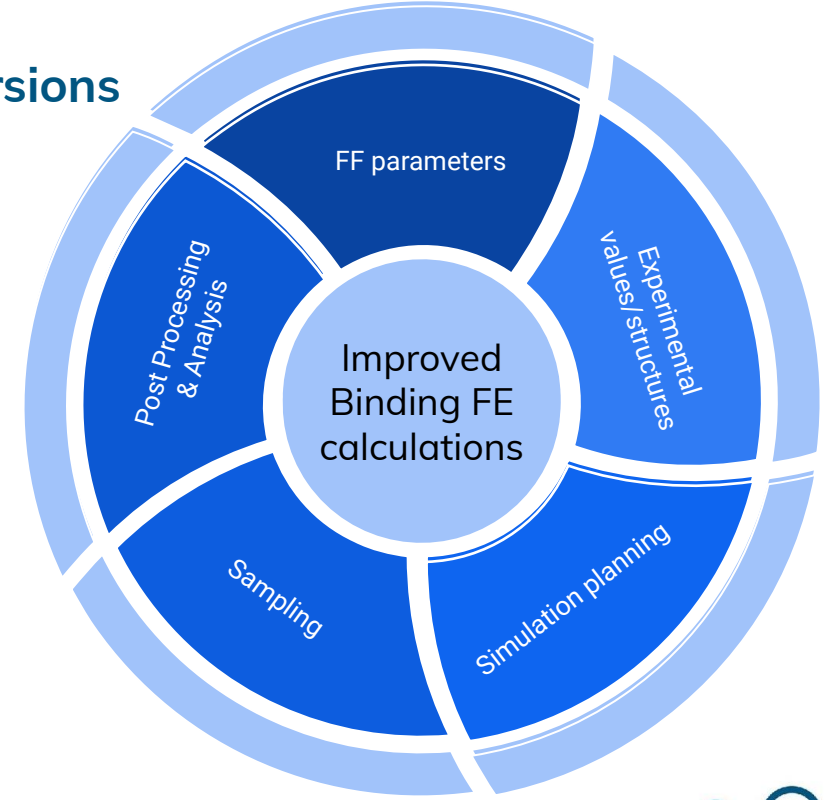


Conclusions/Discussion

Encouraging OpenFF results across all versions

- FF differences are often in the noise of calculations
- Analysis led to more observations and detection of issues (e.g. structure preparation)
- Further work is needed to automatize and increase reproducibility of RBFE calculations
- FF development should be alongside development of other aspects

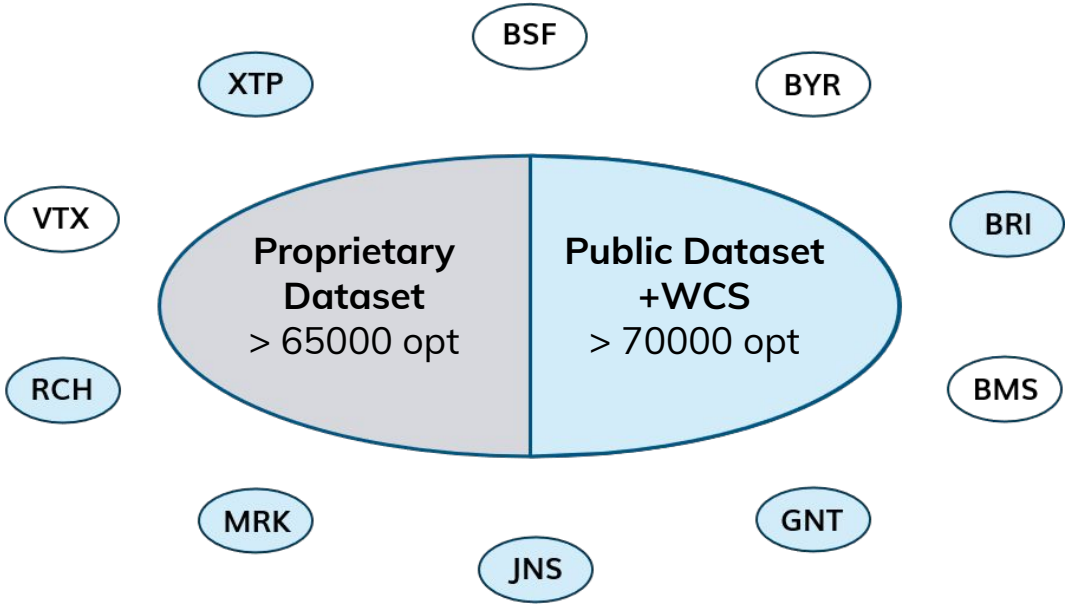
➔ *What other analyses/calculations would you like to see in future work?*



Pharma Partners Small Molecule Benchmarking Season 1

Geometry Optimization Benchmarking Season 1 is complete

Geometry Optimization Benchmark - Season 1 has been officially kicked of on Jan. 22



4 developers

- David Dotson (OpenFF)
- David Hahn (OpenFF, Janssen)
- Jeff Wagner (OpenFF)
- Josh Horton (OpenFF, Cole Lab)

4 subject-matter experts

- Josh Horton (OpenFF, Cole Lab)
- Bill Swope (representing Genentech)
- Lee-Ping Wang (OpenFF, Wang Lab)
- Lorenzo D'Amore (OpenFF, Janssen)

OpenFF-2.0.0 (Sage) - the next generation of OpenFF force field

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

2016

SMIRNOFF99Frosst

Initial SMIRNOFF port of the parm99Frosst force field

2019

Parsley

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

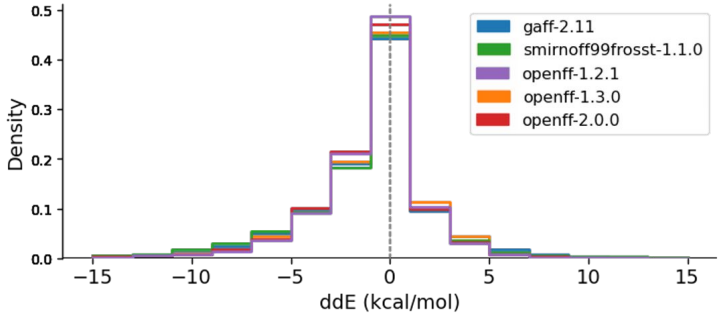
2021

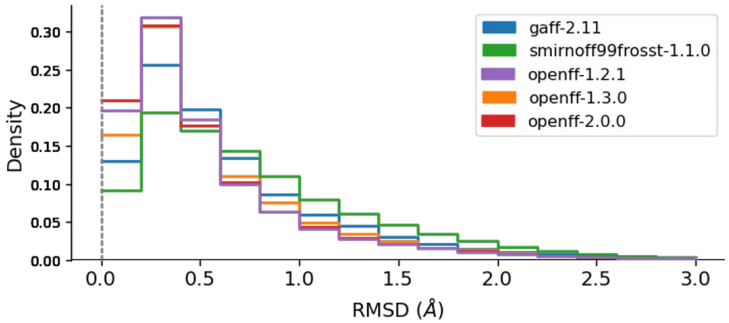
Sage

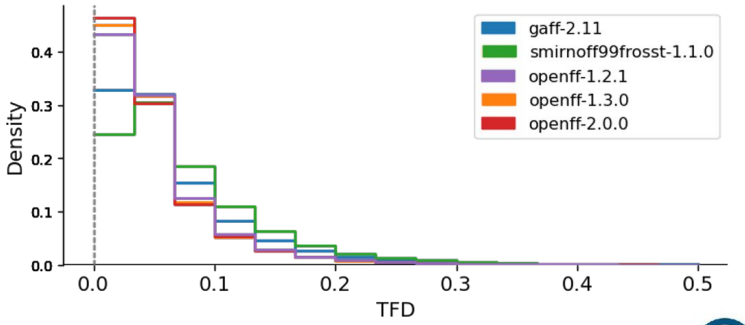
Retrained **vdW** parameters against physical property data + retrained **valence** parameters

Public Dataset Aggregated Results

- **OpenFF-2.0.0 (SAGE)** showed excellent performance when benchmarked against the **Public OpenFF Industry Dataset**

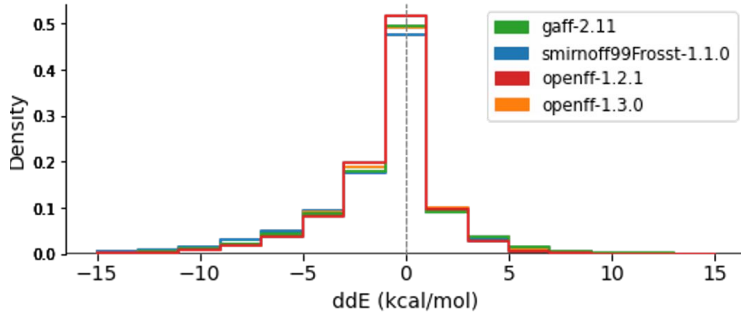


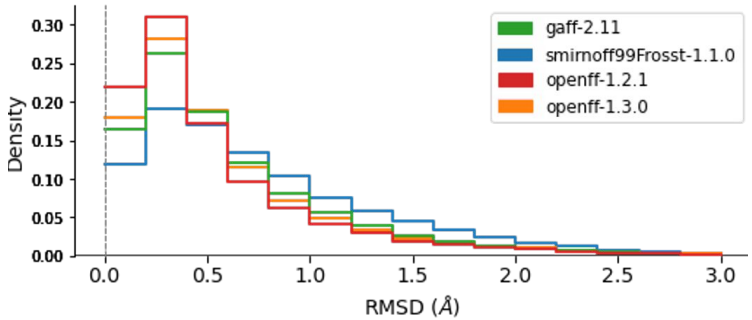


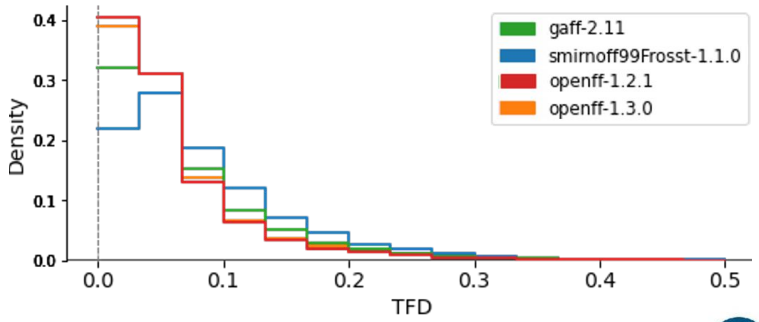


Parsley has been a significant improvement over Smirnoff

- **OpenFF-1.2.1** showed good performance when benchmarked against the **Proprietary OpenFF Industry Dataset**







The workflow now supports optimizations with OPLS

During Season 1 thanks to David Hahn we rolled out the ability to execute optimization benchmarks with OPLS.

- The **`openff-benchmark schrodinger`** command tree allows the execution of optimizations using:
  - OPLS4 (Schrödinger release 2021-1 and later)
  - OPLS3e (Schrödinger release 2020-4 and earlier)
- Both custom and default parameters
- Same input/output behavior as **`openff-benchmark optimize execute`**
- **Requires Schrödinger binaries** (ffbuilder, macromodel) and an **active license**

Disclaimer: Every partner/client of Schrödinger has to independently get approval from Schrödinger before publishing the results.

Summary: next steps for pharma partners

Season 1 is complete

Manuscript

- Janssen will lead the manuscript development effort

Requested

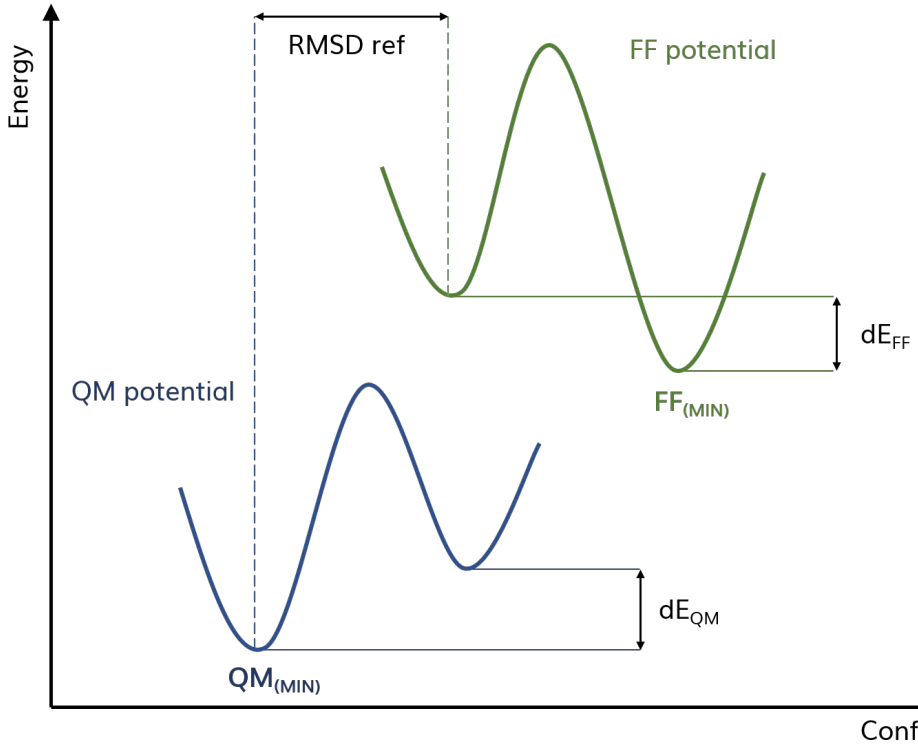
- Optimizations with OpenFF-2.0.0 on proprietary dataset

Strongly encouraged

- OPLS4 (Schrödinger 2021-1) optimizations with both custom and default parameters on proprietary dataset

- **OpenFF-2.0.0 results have been received from 2/10 partners.**
- **OPLS4 results have been received from 1/10 partners (+1 with default parameters)**

Feature proposed were made available during Season 1



Energy comparison

- `swope`

  $dE = (E_{FF,i} - E_{FF,min})$

- `lucas`

  $dE = (E_{FF,ref} - E_{FF,min})$

- `compare-forcefields`

  $ddE = (E_{FF,i} - E_{FF,min}) - (E_{QM,i} - E_{QM,min})$

- `match-minima`

  $ddE = (E_{FFj} - E_{FF,ref}) - (E_{QM,i} - E_{QM,min})$

Feature requested were made available during Season 1

- Code refactor now allows to run the analyses on different methods as separate task with a remarkable improvement in time consumption

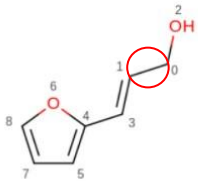
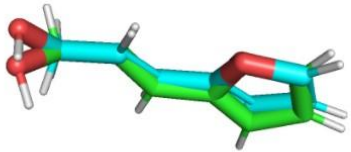
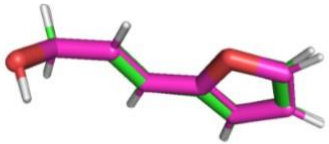
```
for ff_method in `ls -d 4-compute-mm/*`; do
    openff-benchmark report compare-forcefields --input-path 4-compute-qm \
                                                --ref-method b3lyp-d3bj \
                                                --input-path $ff_method \
                                                --output-directory 5-match-minima &
done
```

It ran on the OpenFF Industry Public Dataset approx. **22 hours** for 5 different FFs rather than **22*5 hours**

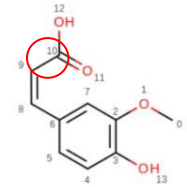
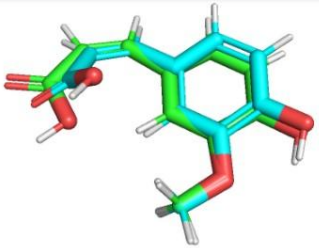
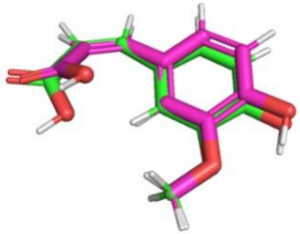
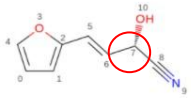
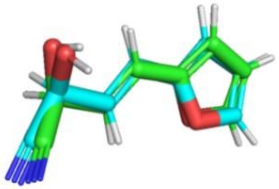
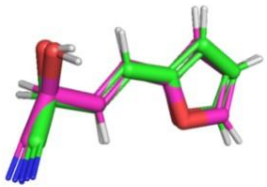
Partners independently developed custom analyses

- Thomas Fox identified structural shortcomings (bonds, angles, torsion) for specific chemical groups optimized with openFF-1.3.0

vinyl-CO

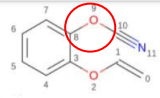
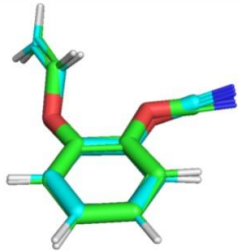
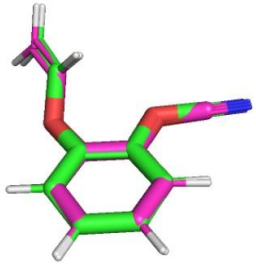
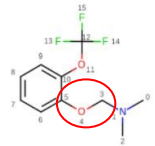
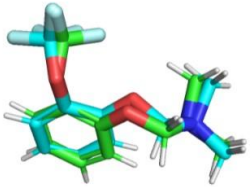
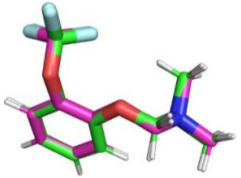
Chemical structure	OpenFF-1.3.0 opt geo (cyan) aligned with QM opt geo (green)	OpenFF-2.0.0 opt geo (magenta) aligned with QM opt geo (green)	param ID in 1.3.0	OpenFF-1.3.0 difference with QM (in degrees)	OpenFF-2.0.0 difference with QM (in degrees)
			t20	Dihedral 2-**0**-**1**-3 48.63	Dihedral 2-**0**-**1**-3 3.25

Partner independently developed custom analyses

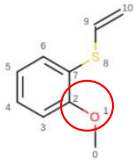
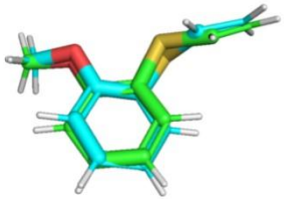
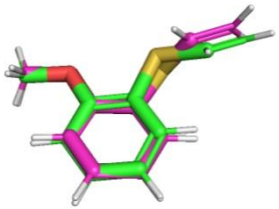
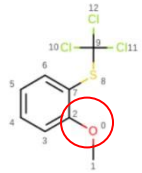
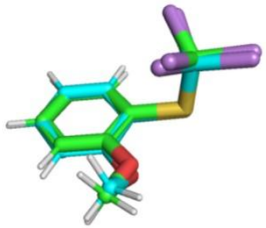
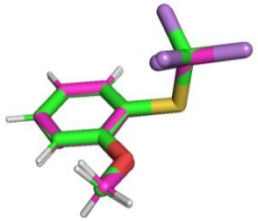
Chemical structure	OpenFF-1.3.0 opt geo (cyan) aligned with QM opt geo (green)	OpenFF-2.0.0 opt geo (magenta) aligned with QM opt geo (green)	param ID in 1.3.0	OpenFF-1.3.0 difference with QM (in degrees)	OpenFF-2.0.0 difference with QM (in degrees)
			t47, t48	Dihedral 8-**9-10**-12 35.56 Dihedral 8-**9-10**-11 34.50	Dihedral 8-**9-10**-12 18.26 Dihedral 8-**9-10**-11 15.36
			t20	Dihedral 5-**6-7**-8 34.92 Dihedral 5-**6-7**-10 39.45	Dihedral 5-**6-7**-8 19.48 Dihedral 5-**6-7**-10 24.30

Partner independently developed custom analyses

aryl-methoxy

Chemical structure	OpenFF-1.3.0 opt geo (cyan) aligned with QM opt geo (green)	OpenFF-2.0.0 opt geo (magenta) aligned with QM opt geo (green)	param ID in 1.3.0	OpenFF-1.3.0 difference with QM (in degrees)	OpenFF-2.0.0 difference with QM (in degrees)
			t96	Dihedral 3-**8-9**-10 23.39 Dihedral 7-**8-9**-10 24.53	Dihedral 3-**8-9**-10 4.96 Dihedral 7-**8-9**-10 5.33
			t86, t96	Dihedral 1-**3-4**-5 31.07 Dihedral 3-**4-5**-6 44.78 Dihedral 3-**4-5**-10 42.26	Dihedral 1-**3-4**-5 10.59 Dihedral 3-**4-5**-6 10.45 Dihedral 3-**4-5**-10 8.77

Partner independently developed custom analyses

Chemical structure	OpenFF-1.3.0 opt geo (cyan) aligned with QM opt geo (green)	OpenFF-2.0.0 opt geo (magenta) aligned with QM opt geo (green)	param ID in 1.3.0	OpenFF-1.3.0 difference with QM (in degrees)	OpenFF-2.0.0 difference with QM (in degrees)
			t86, t96	Dihedral 0-**1**-**2**-3 49.55 Dihedral 0-**1**-**2**-7 47.50	Dihedral 0-**1**-**2**-3 4.66 Dihedral 0-**1**-**2**-7 3.90
			t86, t96	Dihedral 1-**0**-**2**-3 39.20 Dihedral 1-**0**-**2**-7 38.11	Dihedral 1-**0**-**2**-3 12.94 Dihedral 1-**0**-**2**-7 12.65

General conclusions

- Benchmarking is an essential part within the OpenFF infrastructure
- Geometry optimization benchmarking almost fully automated
- OpenFF-2.0.0 Sage shows promising results
- Important insights shared by the OpenFF community

Q&A

- Q&A Time
- Break (5 min)
- Interactive session (45 minutes)
- Q&A Time

open
forcefield

@openforcefield

www.openforcefield.org

5th open force field follow-up workshop

Sept. 1, 2021 | Interactive session

Interactive session

- Download the material: https://github.com/openforcefield/2021-benchmarking-workshop
- Follow the installation instructions:

1. Clone the repository

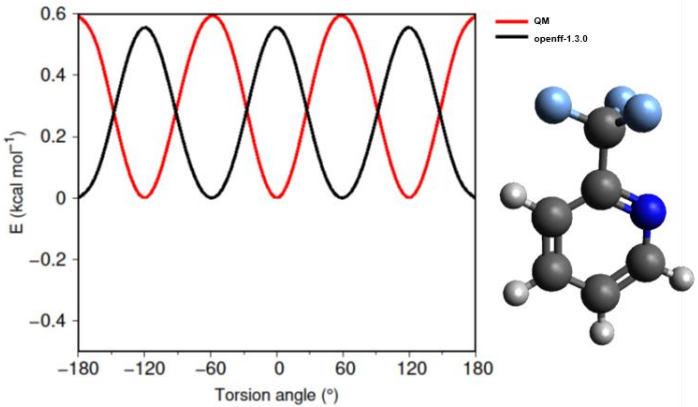
```
git clone git@github.com:openforcefield/2021-benchmarking-workshop.git
cd 2021-benchmarking-workshop
```

2. Install and activate the conda environment

```
conda env create -f env.yml
conda activate 2021-benchmarking-workshop
```

3. Start up the jupyter notebook

```
jupyter notebook workshop.ipynb
```



FE calculations with the pmx workflow

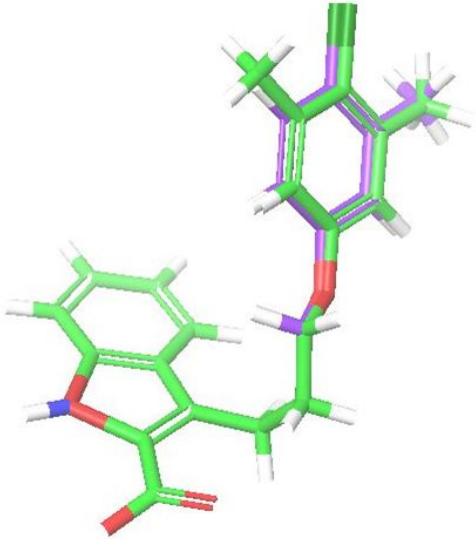
Currently still a home-brewed version: github.com/dfhahn/pmx/tree/py3
It misses protein/ligand preparation and docking/alignment
The OpenFF & OpenFE communities are working towards a industry-standard solution for the community

- OpenFF-1.X.X parameterization and conversion to GROMACS topologies
- Creation of hybrid molecules/topologies with pmx package
- Protein FF: AMBER ff99sb*ILDN [1]
- Solvation / 150 nM NaCl with GROMACS gmx
- Run times:
- 6 ns per end state in water/complex
- Extraction of 80 frames per end state and 50 ps non-equilibrium runs from A->B and B->A (total 8 ns in water/complex)
- 3 repeats (20 ns each) of the above = 60 ns sampling per perturbation
- Analysis of work distributions (MBAR) with pmx

[1] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. Proteins 2006, 65 (3), 712–725. https://doi.org/10.1002/prot.21123.

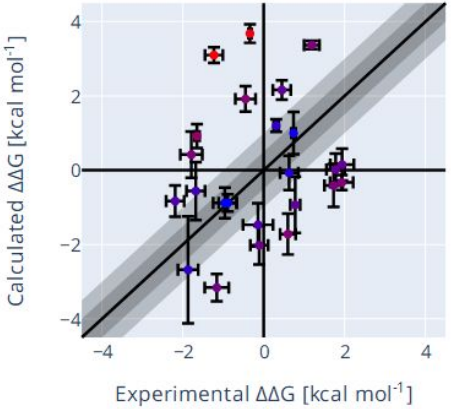
Inspection leads to better agreement with experiment: MCL-1

Ambiguous poses:
meta-substituted phenyl rings



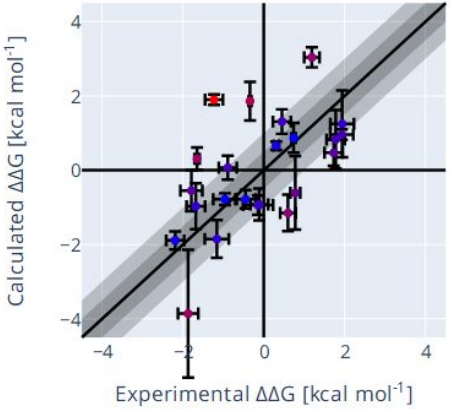
Original

mcl1 (N = 24)
RMSE: 2.03 [95%: 1.59, 2.45]
MUE: 1.74 [95%: 1.35, 2.18]



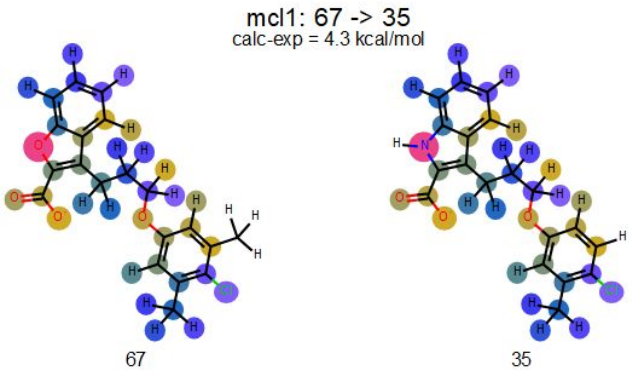
Flipped rings

mcl1_alt (N = 23)
RMSE: 1.33 [95%: 0.98, 1.67]
MUE: 1.11 [95%: 0.83, 1.42]



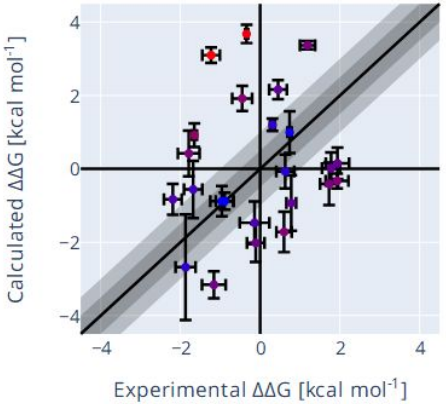
Inspection leads to better agreement with experiment: MCL-1

Ambiguous poses:
meta-substituted phenyl rings



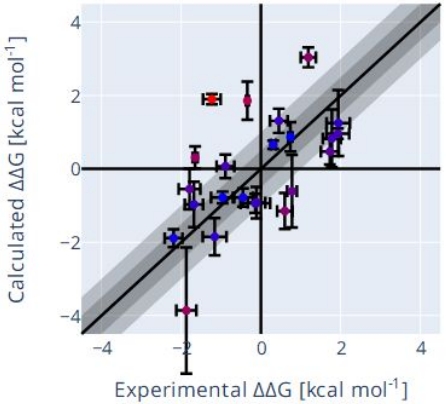
Original

mcl1 (N = 24)
RMSE: 2.03 [95%: 1.59, 2.45]
MUE: 1.74 [95%: 1.35, 2.18]



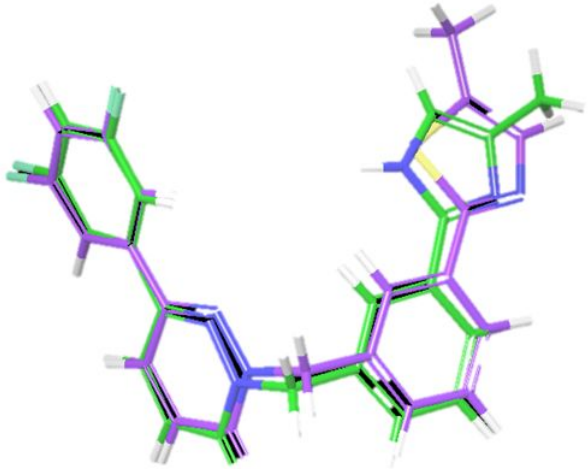
Flipped rings

mcl1_alt (N = 23)
RMSE: 1.33 [95%: 0.98, 1.67]
MUE: 1.11 [95%: 0.83, 1.42]



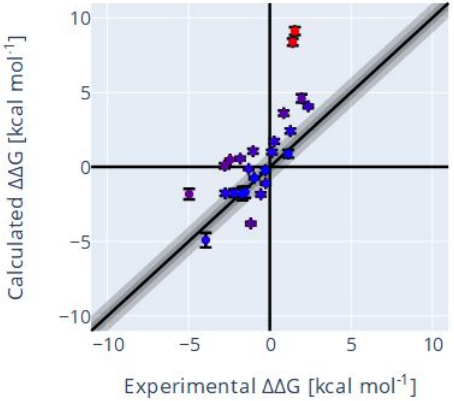
Inspection leads to better agreement with experiment: c-MET

Ambiguous poses:
rotation of 5-ring/tautomeric state



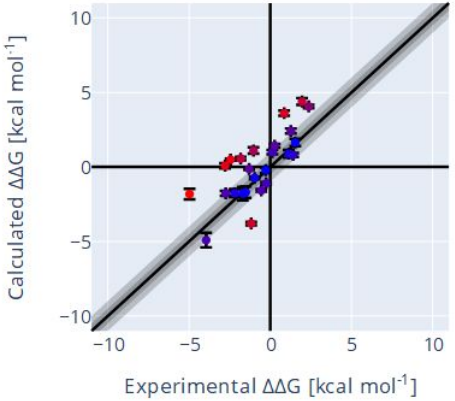
Original

cmet (N = 25)
RMSE: 2.66 [95%: 1.60, 3.73]
MUE: 1.90 [95%: 1.21, 2.73]



Different tautomer and ring flip

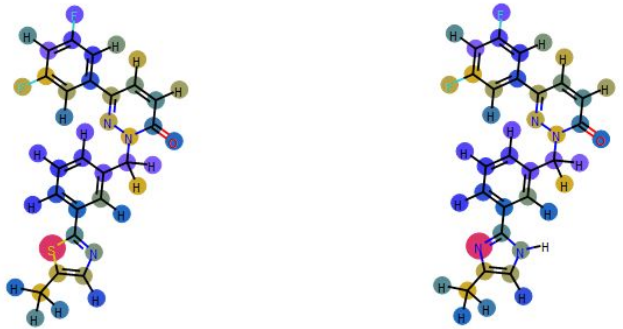
cmet (N = 25)
RMSE: 1.66 [95%: 1.27, 1.99]
MUE: 1.32 [95%: 0.93, 1.72]



Inspection leads to better agreement with experiment: c-MET

Ambiguous poses:
rotation of 5-ring/tautomeric state

cmet: CHEMBL3402750_400_10 -> CHEMBL3402748_5300_8

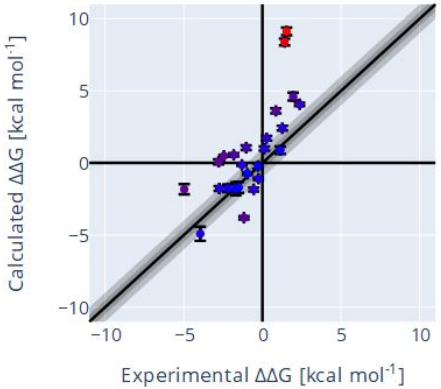


CHEMBL3402750_400_10 CHEMBL3402748_5300_8

ΔΔG(exp)=1.5 kcal/mol
ΔΔG(original)=9.1 kcal/mol
ΔΔG(improved)=1.7 kcal/mol

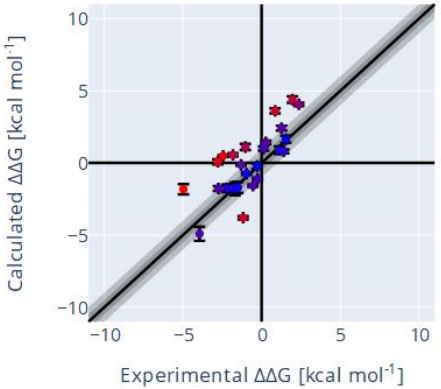
Original

cmet (N = 25)
RMSE: 2.66 [95%: 1.60, 3.73]
MUE: 1.90 [95%: 1.21, 2.73]

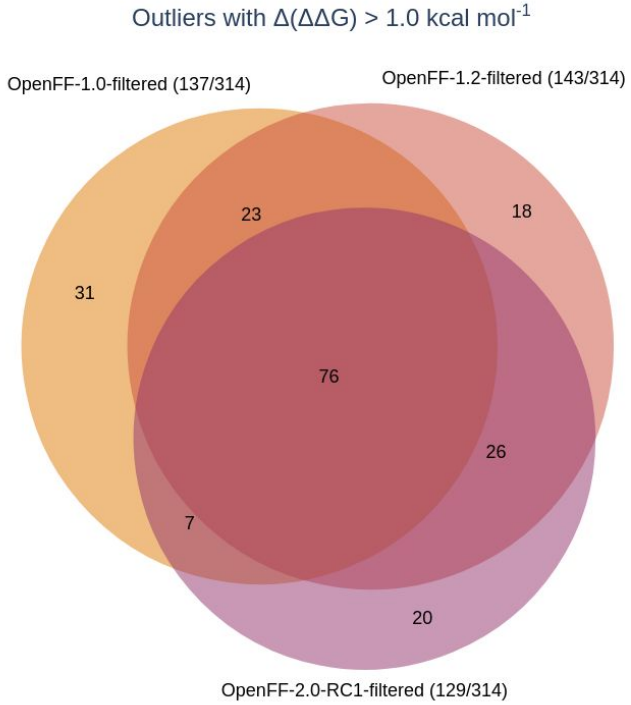


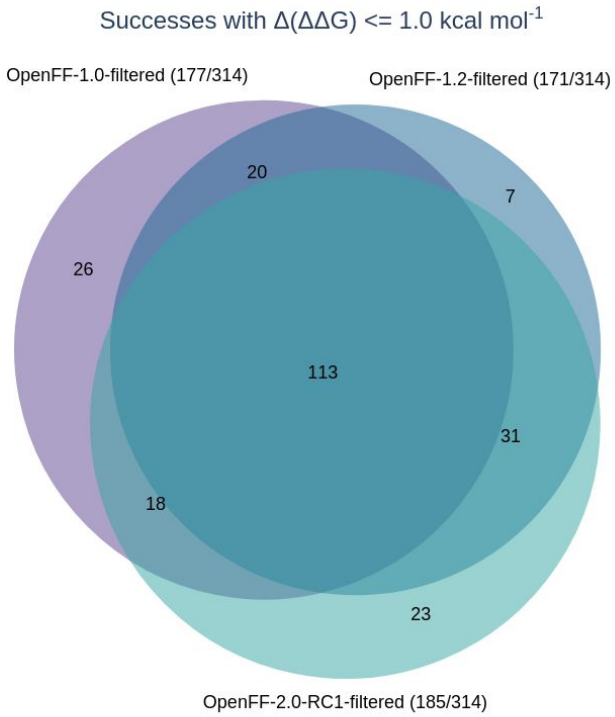
Different tautomer and ring flip

cmet (N = 25)
RMSE: 1.66 [95%: 1.27, 1.99]
MUE: 1.32 [95%: 0.93, 1.72]



OpenFF versions have mostly same outliers and successes





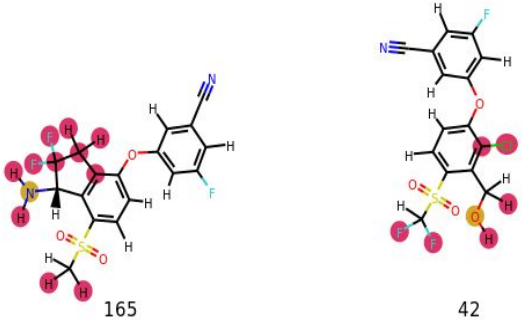
Differences between OpenFF-1.0 and OpenFF-2.0RC1

Exp.	0.3 kcal/mol
OpenFF-1.0	-3.3 kcal/mol
OpenFF-2.0RC1	0.3 kcal/mol

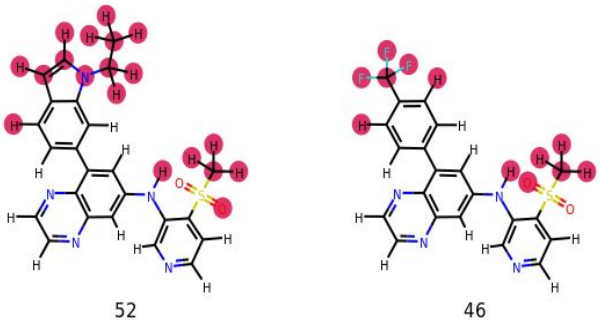
Exp.	2.6 kcal/mol
OpenFF-1.0	0.0 kcal/mol
OpenFF-2.0RC1	3.6 kcal/mol

Exp.	-1.6 kcal/mol
OpenFF-1.0	2.7 kcal/mol
OpenFF-2.0RC1	-0.8 kcal/mol

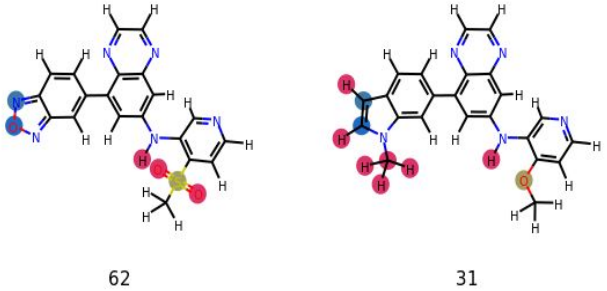
165 -> 42



52 -> 46

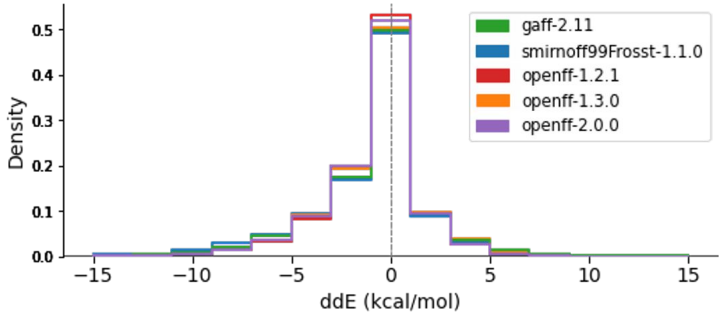


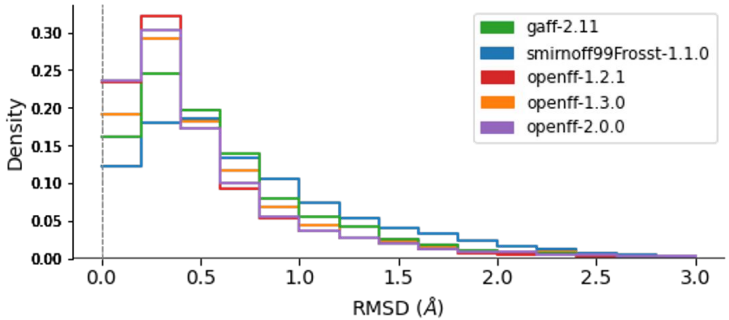
62 -> 31

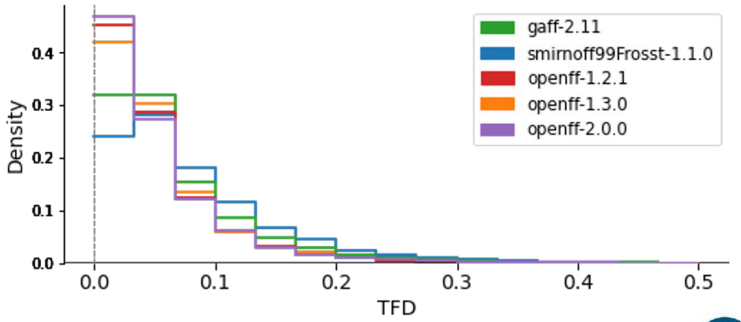


Proprietary JNS+RCH Dataset including OpenFF-2.0.0

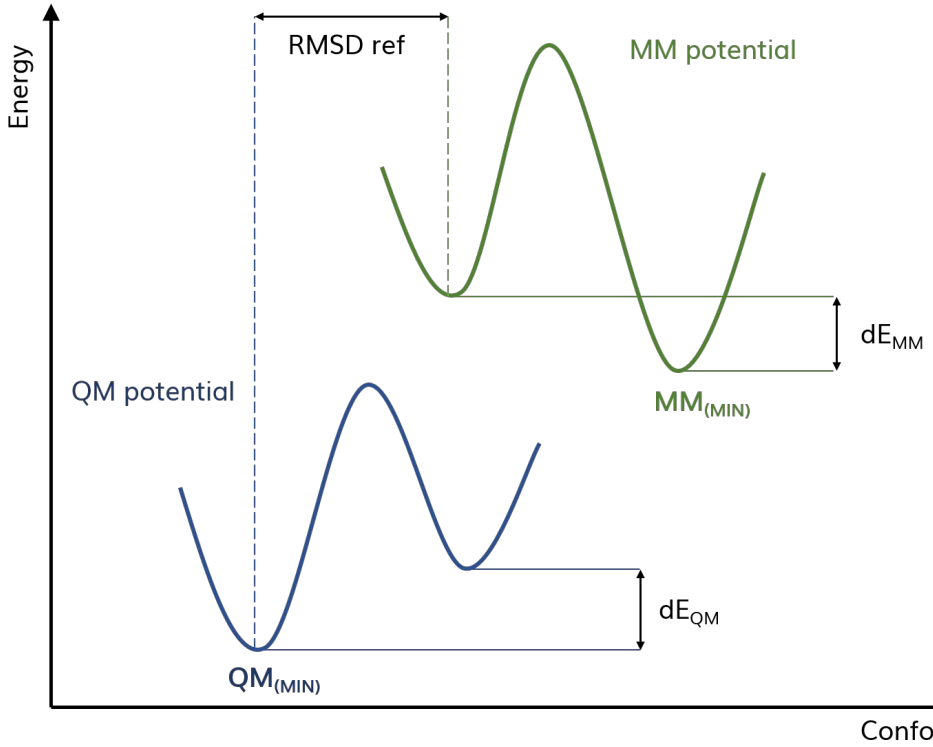
- **OpenFF-2.0.0** shows similar performance compared to benchmarks against the **Public Openff Industry Dataset**







Feature requested were made available during Season 1



Geometry comparison

- `compare-forcefields`

  RMSD, TFD [FF$_i$ / QM$_{min}$]

- `match-minima`

  RMSD, TDF [FF$_j$ / QM$_{min}$]

- `swope`

  RMSD [FF$_i$ / QM$_{min}$]

- `lucas`

  RMSD [FF$_{ref}$ / FF$_{min}$]