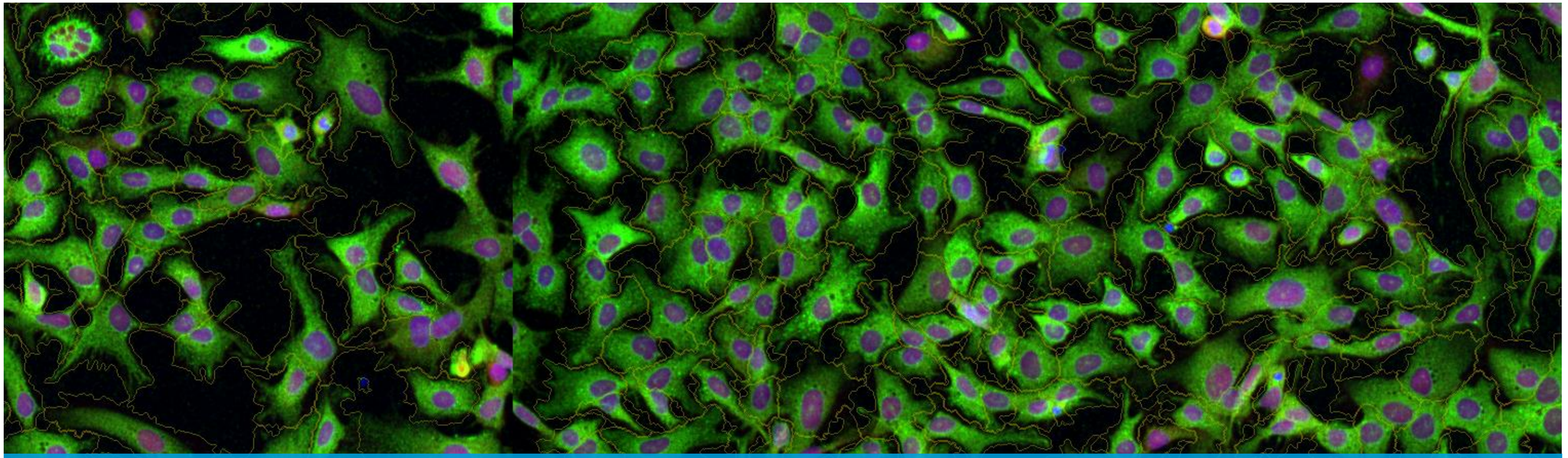


Large scale benchmarking the prospective accuracy of protein-ligand free energy calculations

David Hahn

GCC 2020
Virtual Conference
November 3, 2020

Discovery Sciences

HepG2 hepatocellular carcinoma cells treated with tunicamycin. Nuclei in magenta, and PDI in green.

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Joining forces to advance computational drug discovery

Janssen & Open Force Field

- Evaluate new OpenFF parameters in retrospective drug design projects
  → Give feedback about shortcomings and successes

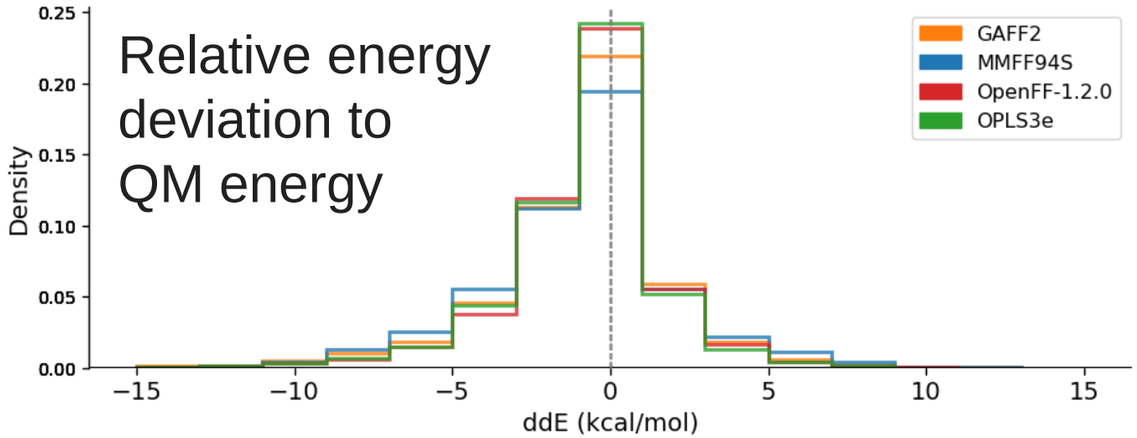
- Generate an open protein-ligand benchmark dataset
  → To be used for systematically comparing methods and parameters in the community (avoid variations in input data)

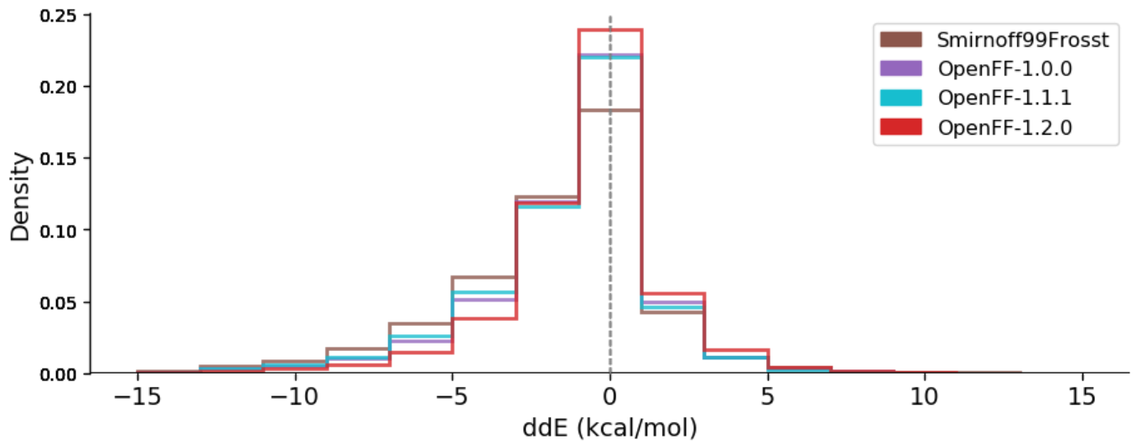
https://openforcefield.org/

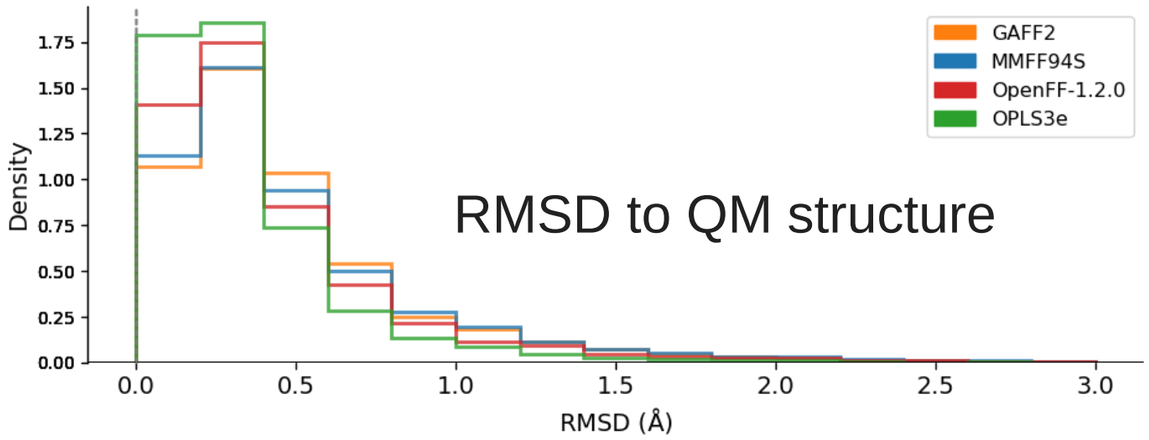
Benchmark assessment of molecular geometries and energies from small molecule force fields

doi.org/10.26434/chemrxiv.12551867.v2

- 23k molecular structures of 3k different small molecules
- OpenFF-1.2.0 ranks second after OPLS3e
- Performance improved with every OpenFF release







Binding Free Energy Calculations are increasingly used in Pharma Industry

ChemMedChem
Communications
doi.org/10.1002/cmdc.201900727

Computationally Empowered Workflow Identifies Novel Covalent Allosteric Binders for KRAS^{G12C}

Jérémie Mortier,*[a] Anders Friberg,[a] Volker Badock,[a] Dieter Moosmayer,[a] Jens Schro
Patrick Steigemann,[a] Franziska Siegel,[a] Stefan Gradl,[a] Marcus Bauser,[a] Roman C. Hil
Hans Briem,[a] Knut Eis,[a] Benjamin Bader,[a] Duy Nguyen,*[a] and Clara D. Christ*[a]

Journal of Medicinal Chemistry
Article
pubs.acs.org/jmc

Prospective Evaluation of Free Energy Calculations for the Prioritization of Cathepsin L Inhibitors

Bernd Kuhn,[†,#] Michal Tichý,[‡,#] Lingle Wang,[§,#] Shaughnessy Robinson,[§] Rainer E. Martin,[†] Andreas Kuglstatter,[†] Jörg Benz,[†] Maude Giroud,[‡] Tanja Schirmeister,[∥] Robert Abel,*[,§] François Diederich,*[,‡] and Jérôme Hert*[,†]

[†]Roche Pharmaceutical Research and Early Development (pRED), Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd., Grenzacherstrasse 124, 4070 Basel, Switzerland
[‡]Laboratorium für Organische Chemie, ETH Zurich, Vladimir-Prelog-Weg 3, 8093 Zurich, Switzerland
[§]Schrödinger, Inc., 120 West 45th Street, New York, New York 10036, United States
[∥]Institut für Pharmazie und Biochemie, Johannes Gutenberg-Universität Mainz, Staudinger Weg 5, 55128 Mainz, Germany

Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects

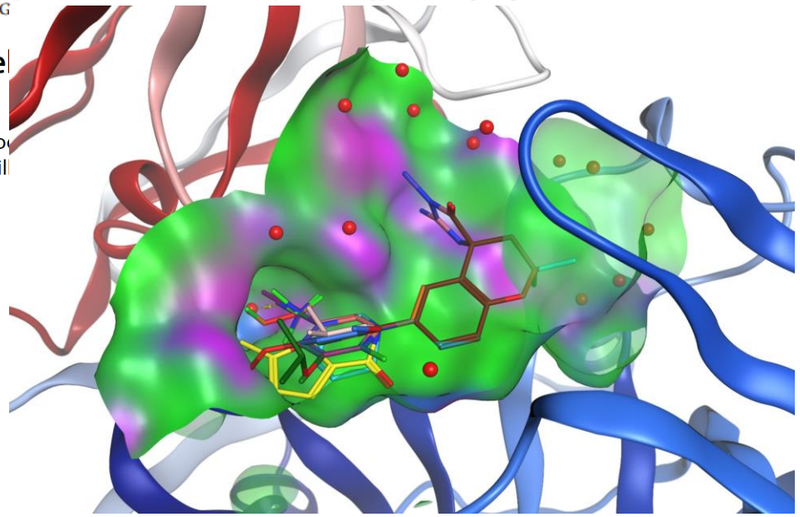
Version 2 ∨ Preprint revised on 25.01.2020, 08:04 and posted on 27.01.2020, 10:02 by Christina Schindler, Hannah Baumann, Andreas Blum, Dietrich Böse, Hans-Peter Buchstaller, Lars Burgdorf, Daniel Cappel, Eugene Chekler, Paul Czodrowski, Dieter Dorsch, Merveille Eguida, Bruce Follows, Thomas Fuchß, Ulrich Grädler, Jakub Gunera, Theresa Johnson, Catherine Jorand Lebrun, Srinivasa Karra, Markus Klein, Lisa Kötzner, Tim Knehans, Mireille Krier, Matthias Leiendecker, Birgitta Leuthner,

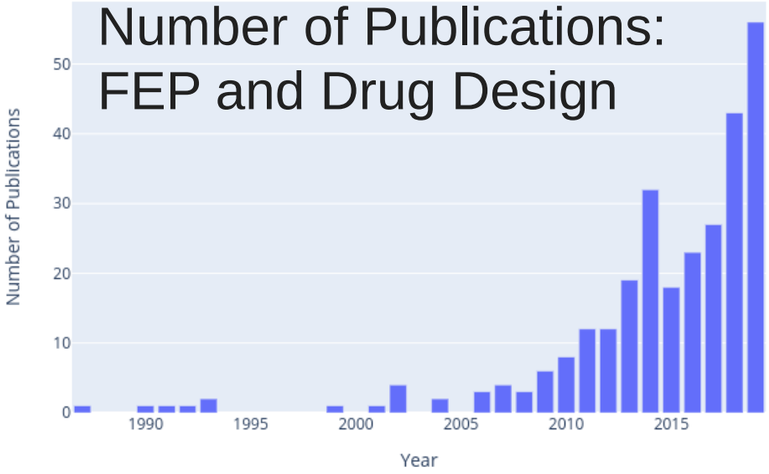
Journal of Chemical Information and Modeling
Article
pubs.acs.org/jcim

Accuracy Assessment and Automation of Free Energy Calculations for Drug Design

Clara D. Christ and Thomas Fox*

Department of Lead Identification and Optimization Support, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, 88397
G





JCTC Journal of Chemical Theory and Computation
Article
Cite This: J. Chem. Theory Comput. 2019, 15, 1884−1895
pubs.acs.org/JCTC

Predicting Activity Cliffs with Free-Energy Perturbation

Laura Pérez-Benito,[†] Nil Casajuana-Martin,[‡] Mireia Jiménez-Rosés,[‡] Herman van Vlijmen,[†] and Gary Tresadern*[,†]

[†]Computational Chemistry, Janssen Research & Development, Janssen Pharmaceutica N. V., Turnhoutseweg 30, Beerse B-2340, Belgium
[‡]Laboratori de Medicina Computacional, Unitat de Bioestadistica, Facultat de Medicina, Universitat Autonoma de Barcelona, Bellaterra 08193, Spain

DOI: 10.1039/C9SC03754C (Edge Article) Chem. Sci., 2020, 11, 1140-1152

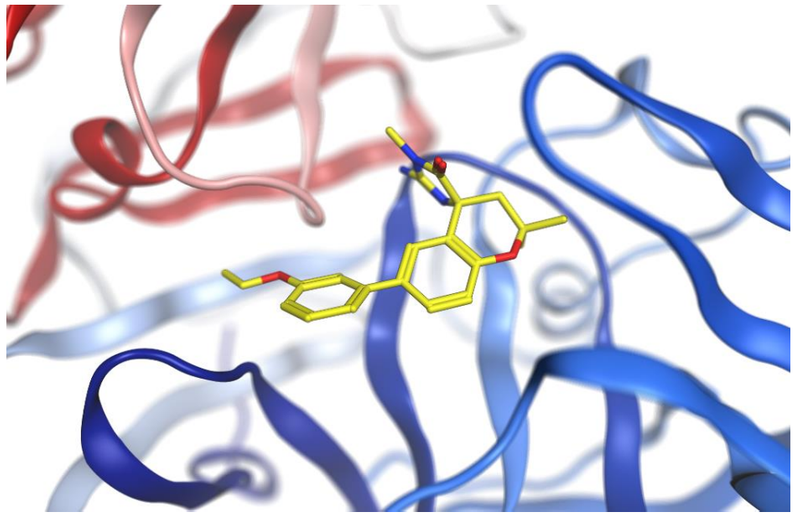
Large scale relative protein ligand binding affinities using non-equilibrium alchemy†‡

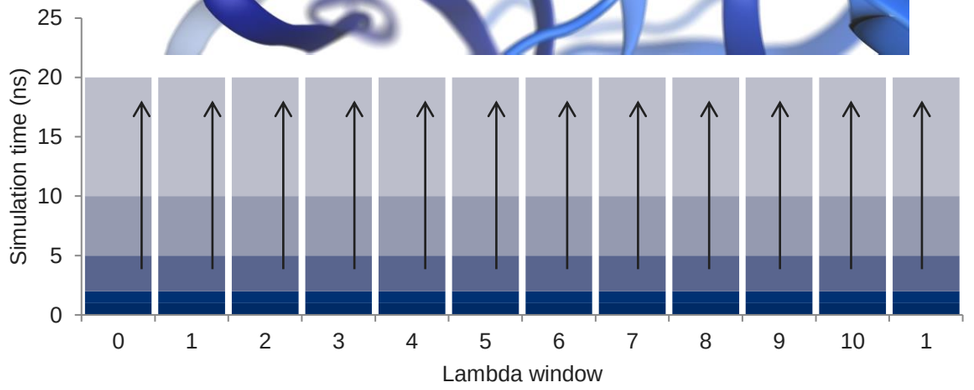
Vytautas Gapsys§ [a], Laura Pérez-Benito§ [b], Matteo Aldeghi [a], Daniel Seeliger [c], Herman van Vlijmen [b], Gary Tresadern [b]*[,b] and Bert L. de Groot [a]*

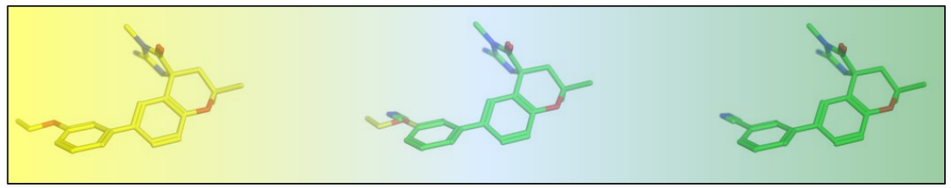
[a]Computational Biomolecular Dynamics Group, Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, D-37077 Göttingen, Germany. E-mail: bgroot@gwdg.de
[b]Computational Chemistry, Janssen Research & Development, Janssen Pharmaceutica N. V., Turnhoutseweg 30, B-2340 Beerse, Belgium. E-mail: gtresade@its.jnj.com
[c]Medicinal Chemistry, Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Strasse 65, D-88397 Biberach a.d. Riss, Germany

Binding free energies are typically calculated by alchemical perturbations

- Model a “hybrid molecule”, mixing ligands **A** and **B**
- Molecular dynamics simulation of this molecule to get the free energy difference between states **A** and **B**
- Computationally very demanding, ~1 day per transformation on a GPU

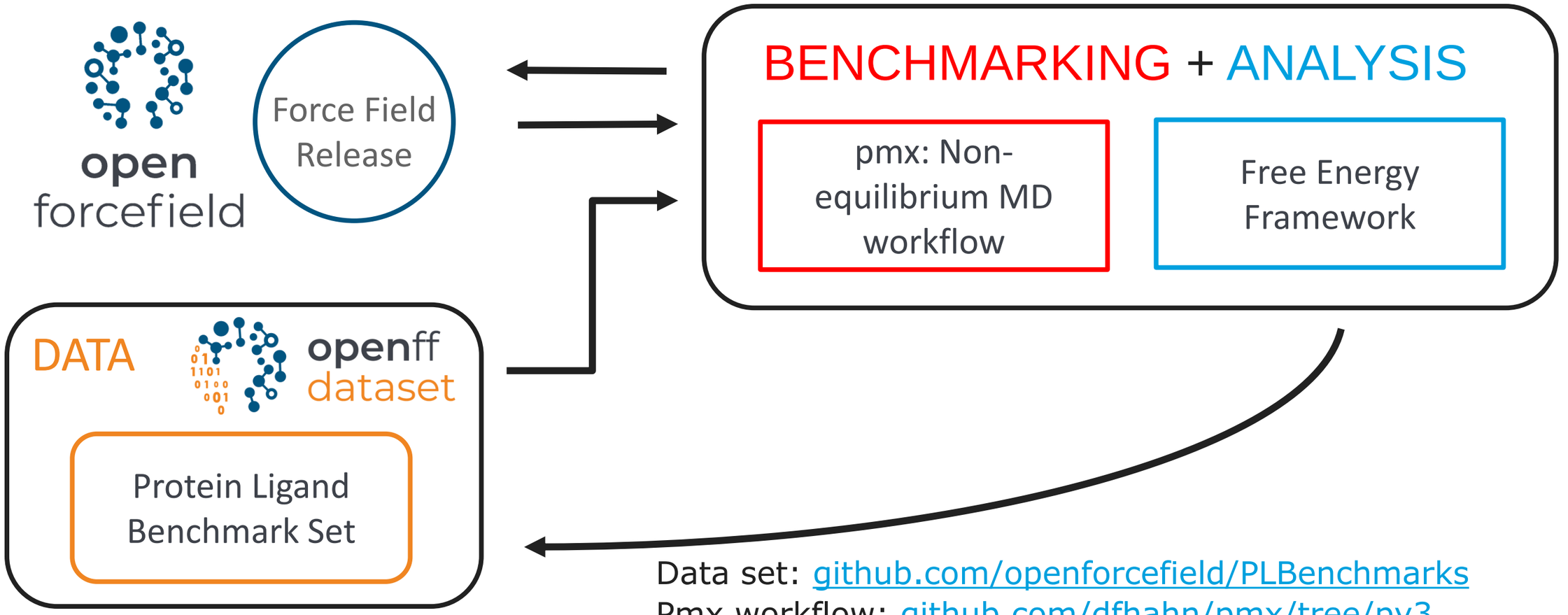






Benchmarking is part of the OpenFF infrastructure

Protein Ligand Binding as validation for OpenFF Parameters



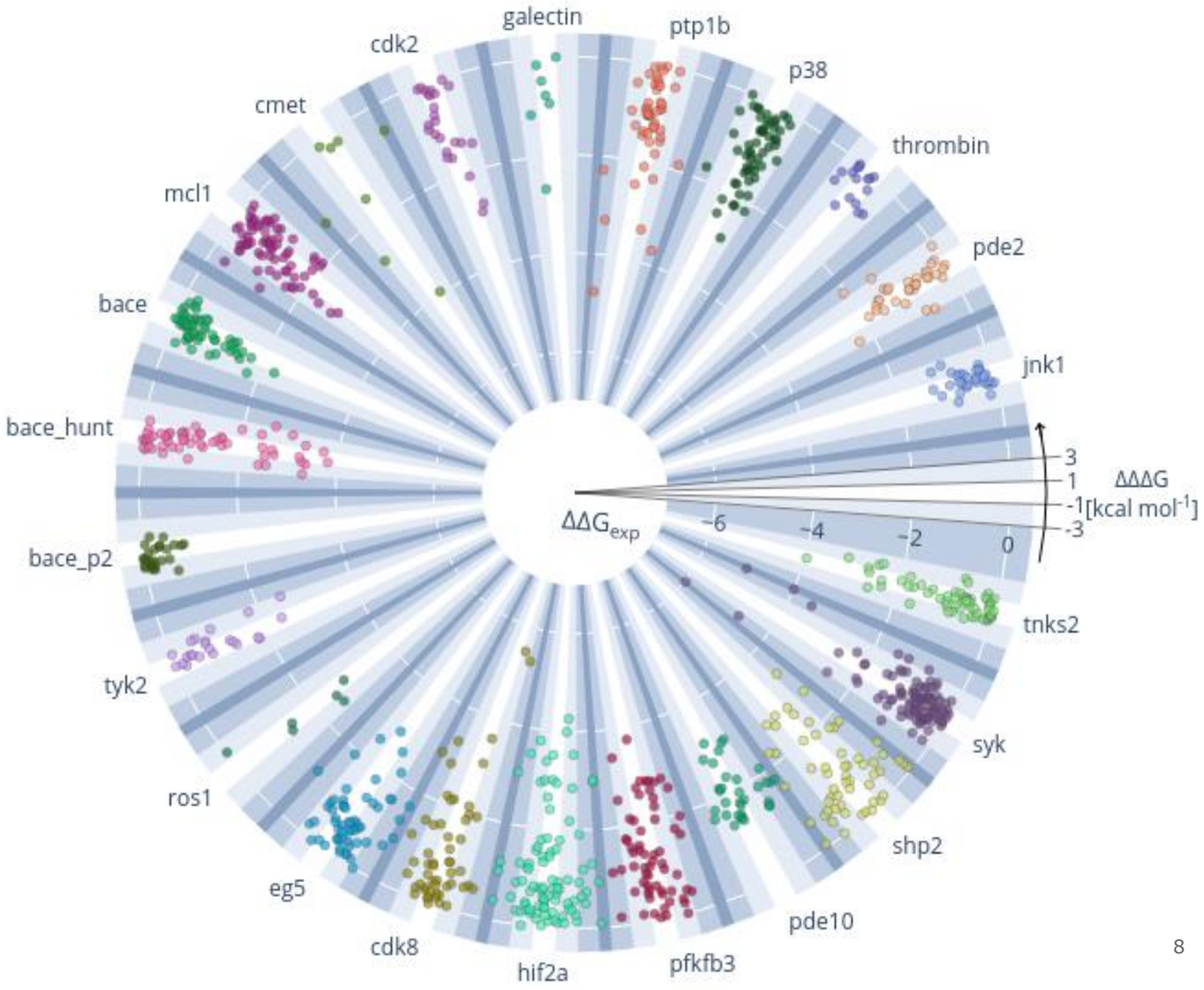
Data set: github.com/openforcefield/PLBenchmarks
Pmx workflow: github.com/dfhahn/pmx/tree/py3
FE Framework: github.com/choderalab/freeenergyframework

Relative free energy calculations with Open Force Field “Parsley” parameters

Skin cells at 20x magnification

Large scale retrospective calculation of relative free energies on 22 targets & 600 ligands

Target	# Ligands	#Perturbations
jnk1	21	31
pde2	21	34
thrombin	11	16
p38	34	56
ptp1b	23	49
galectin	8	7
cdk2	16	25
cmet	24	57
mcl1	42	71
bace	36	58
bace_hunt	32	60
bace_p2	12	26
tyk2	16	24
ros1	28	27
eg5	28	63
cdk8	33	54
hif2a	42	92
pfkfb3	40	66
pde10	35	36
shp2	26	56
syk	44	99
tnks2	27	60
total	**599**	**1067**

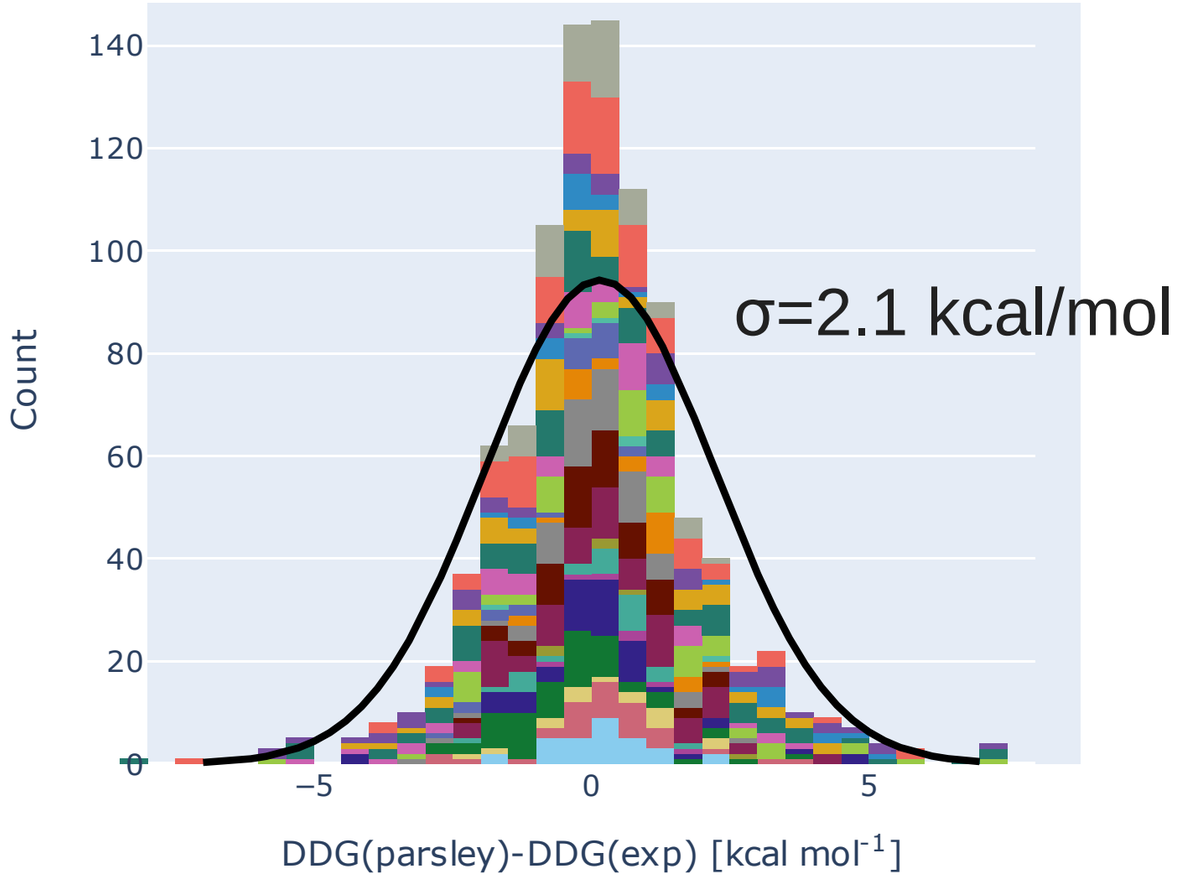


Discovery Sciences (DS)

L. Wang et al. , *J. Am. Chem. Soc.*, 2015, **137** , 2695 —2703.
V. Gapsys et al. *Chem. Sci.*, 2020,**11**, 1140-1152.

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

Different colors denote different targets



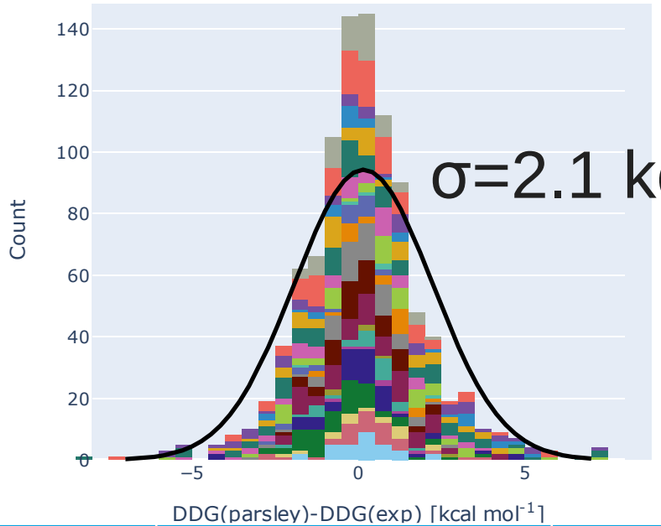
Abs. Error [kcal mol[1]]	# Perturbations	% of total
< 0.5	286	29
<1.0	505	51
<2.0	771	78
<3.0	887	90
total	983	100

Origin of errors:

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

Results can be filtered by convergence metric

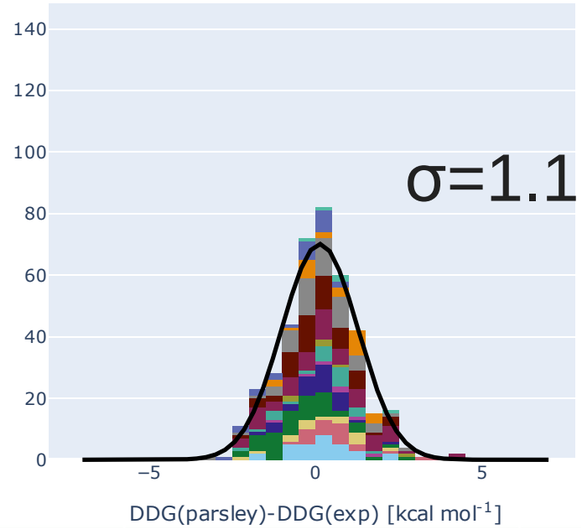
All perturbations



σ=2.1 kcal/mol

Abs. Error [kcal mol[1]]	# Perturbations	% of total
< 0.5	286	29
<1.0	505	51
<2.0	771	78
<3.0	887	90
total	983	100

Converged Perturbations

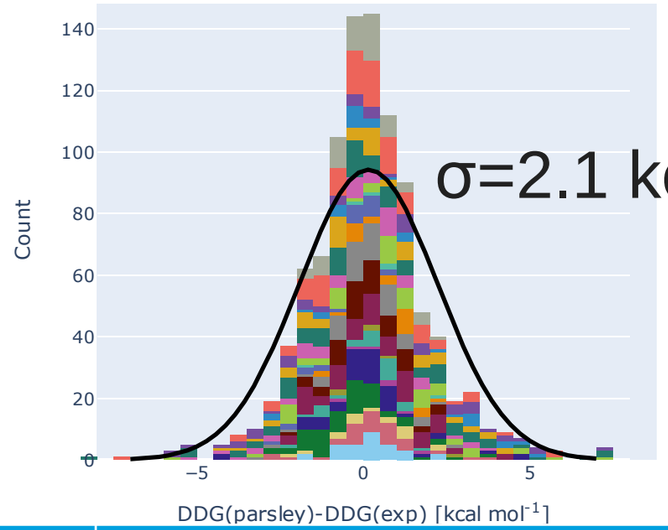


σ=1.1 kcal/mol

Abs. Error [kcal mol[1]]	# Perturbations	% of total
< 0.5	154	38
<1.0	258	64
<2.0	366	91
<3.0	398	99
total	403	100

Outliers accumulate in few targets

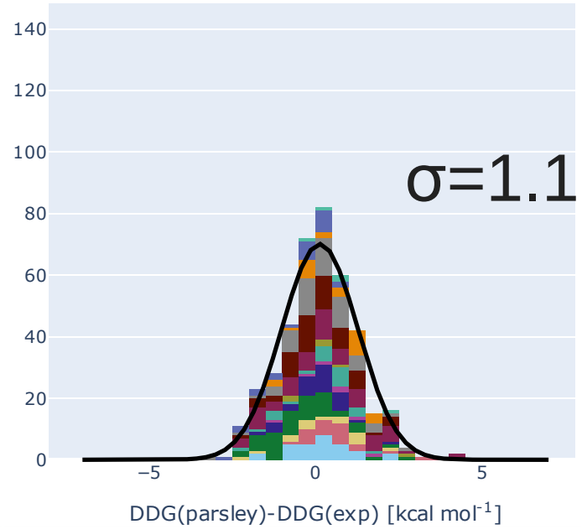
All perturbations



σ=2.1 kcal/mol

Target	# Outliers with AD > 2.0 kcal/mol
Mcl1	13
P38	8
Ptp1b	7
Pde2	6
Bace	4

Converged Perturbations

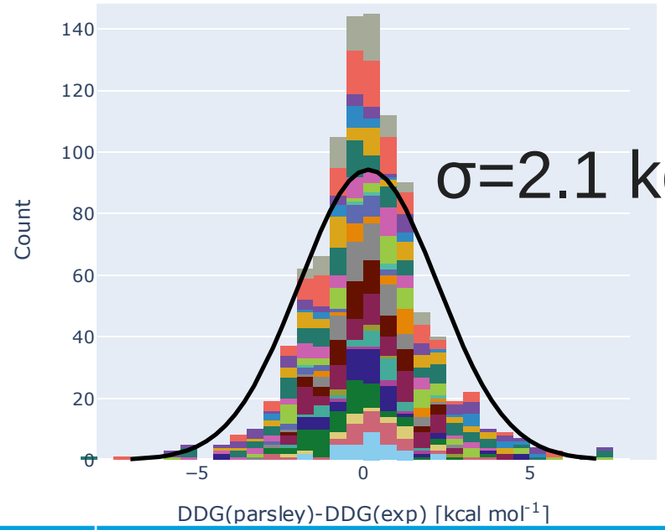


σ=1.1 kcal/mol

Target	# Outliers with AD > 2.0 kcal/mol
Mcl1	11
P38	4
Ptp1b	1
Pde2	3
Bace	4

Convergence criteria are a good diagnosis tool

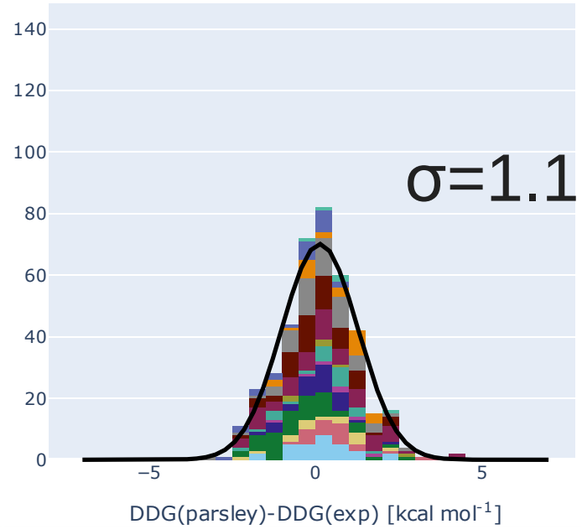
All perturbations



σ=2.1 kcal/mol

Target	# Outliers with AD > 2.0 kcal/mol
Mcl1	**13**
P38	8
ptp1b	***7***
Pde2	6
Bace	4

Converged Perturbations



σ=1.1 kcal/mol

Target	# Outliers with AD > 2.0 kcal/mol
Mcl1	**11**
P38	4
ptp1b	***1***
Pde2	3
Bace	4

mcl1: set-up error (wrong poses, tautomers, charges, FF parameters)

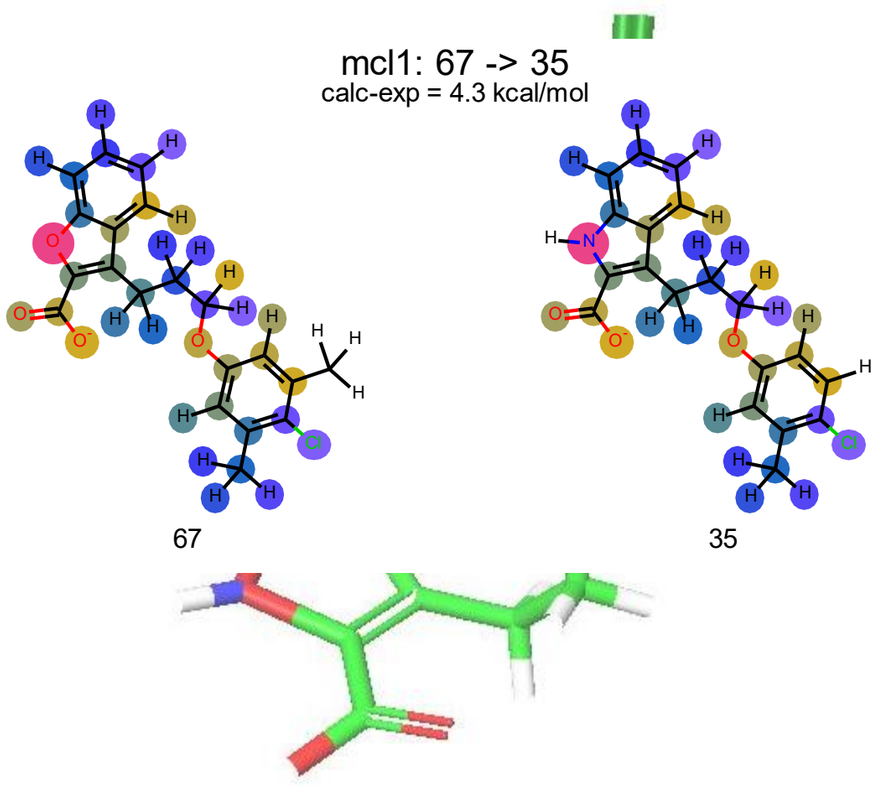
ptp1b: convergence error (too short simulation time, too difficult perturbations)

Improving the set-up

Skin cells at 20x magnification

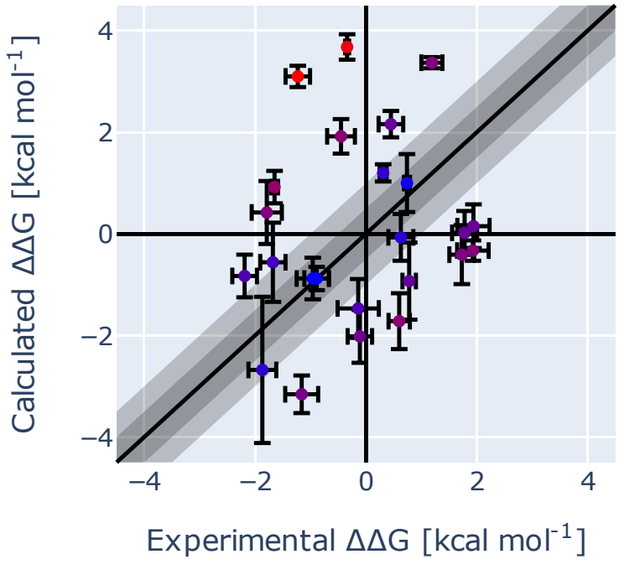
Inspection leads to better agreement with experiment: MCL-1

Ambiguous poses:
meta-substituted phenyl rings



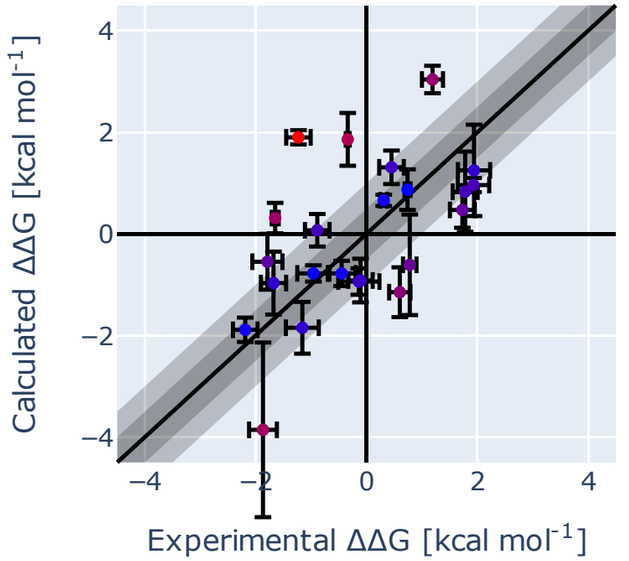
Original

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



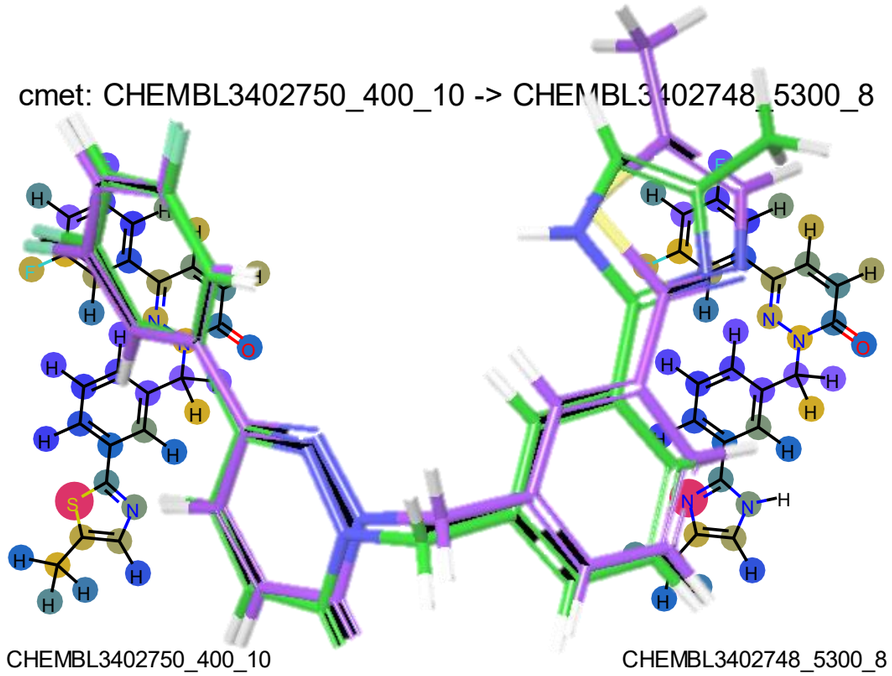
Flipped rings

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



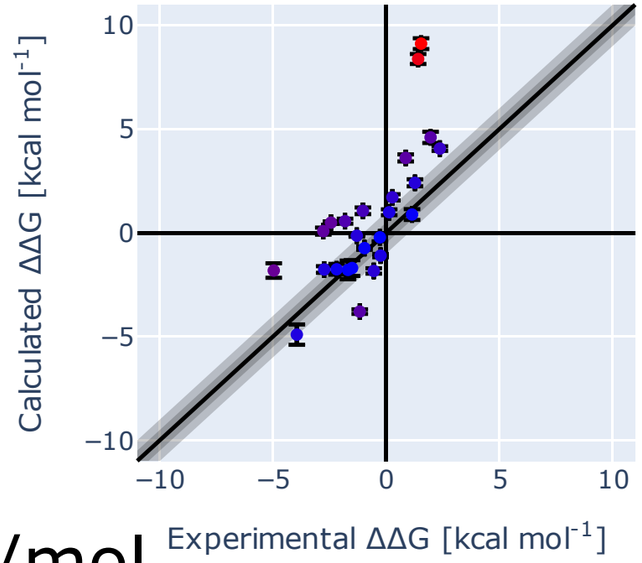
Inspection leads to better agreement with experiment: c-MET

Ambiguous poses:
rotation of 5-ring/tautomeric state



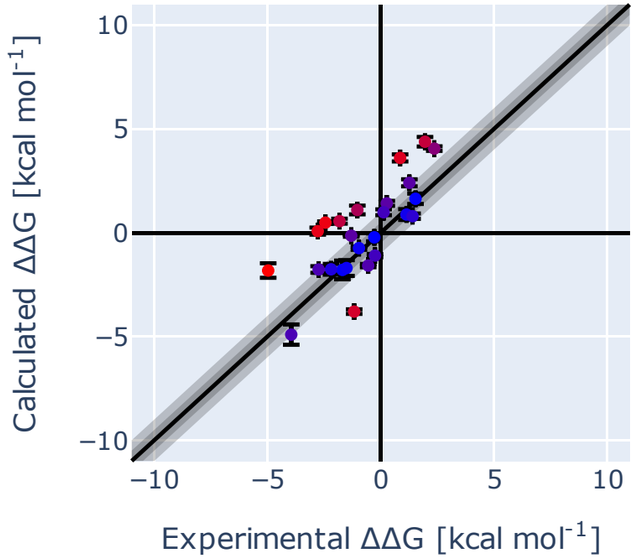
Original

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]



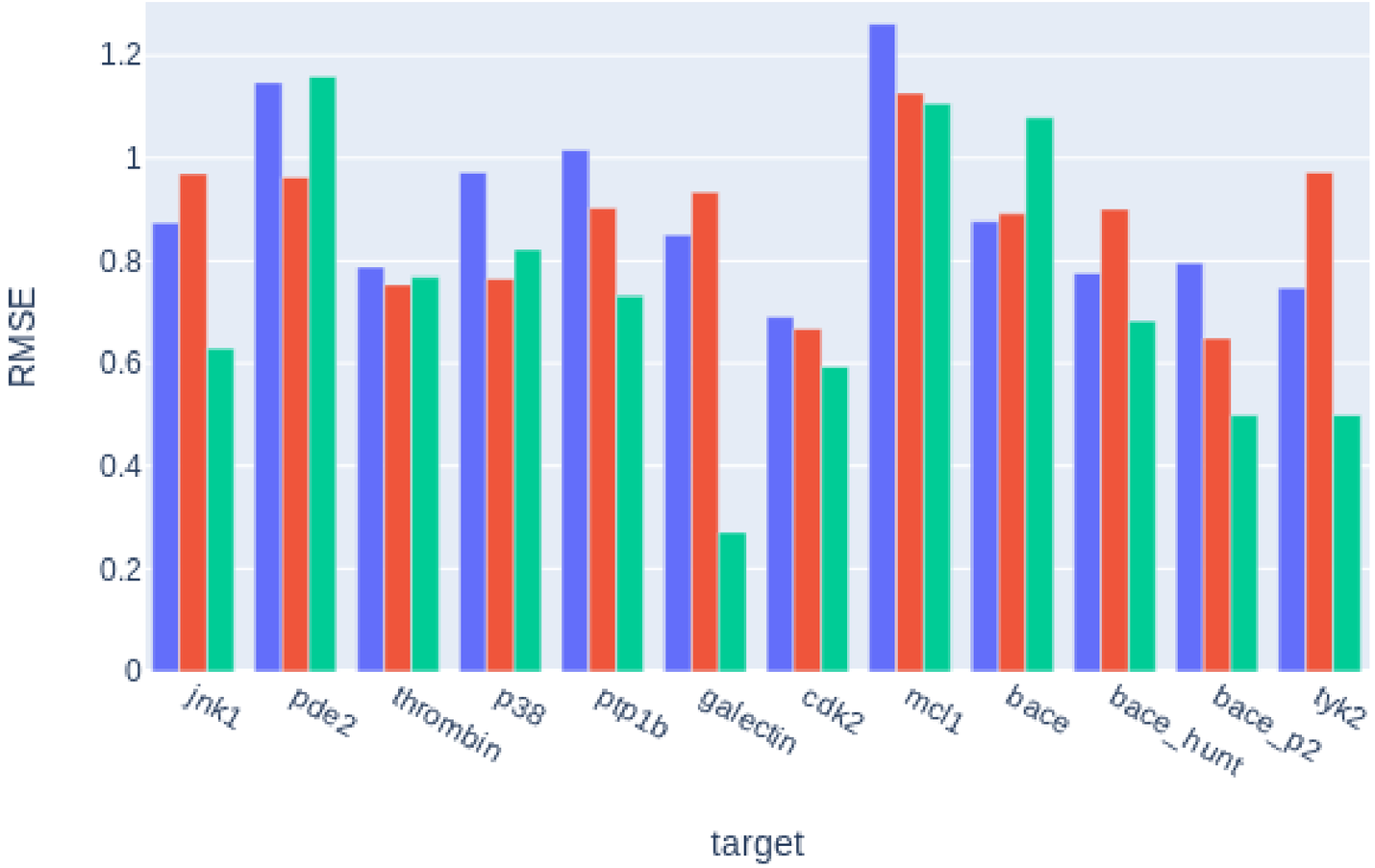
ΔΔG(exp)=1.5 kcal/mol
ΔΔG(original)=9.1 kcal/mol
ΔΔG(improved)=1.7 kcal/mol

Comparison to other force fields

Skin cells at 20x magnification

Different force fields perform comparably

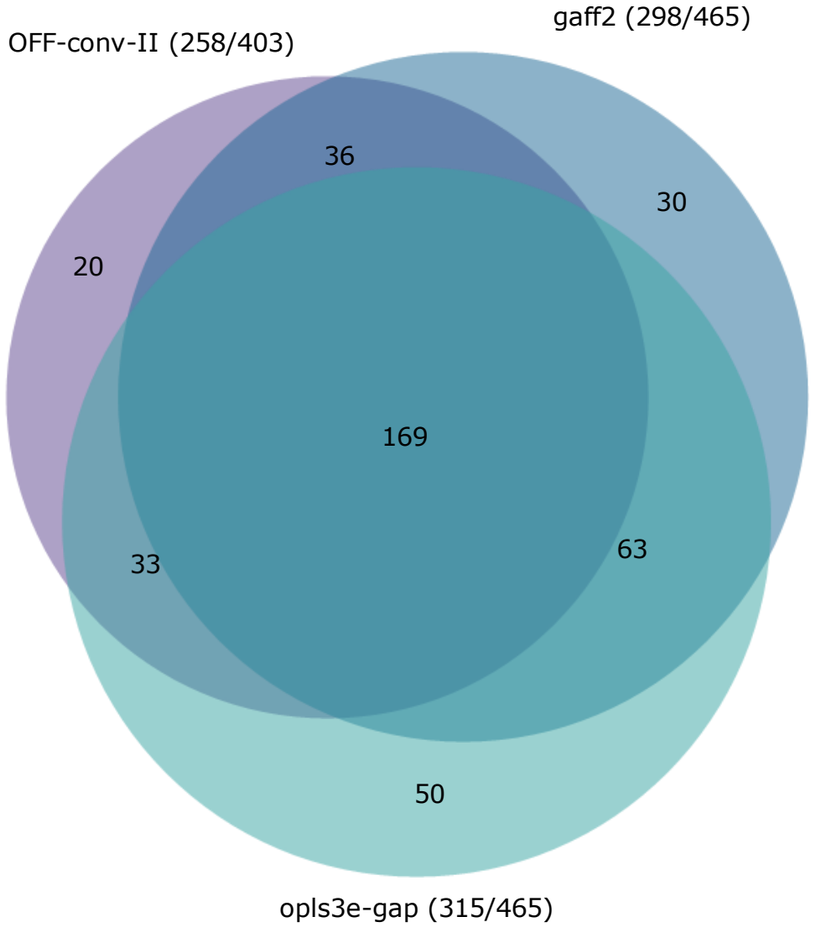
RMSE (kcal/mol) of DG values for
Open Force Field (blue),
GAFF2.1 (red) and
OPLS3e (green)



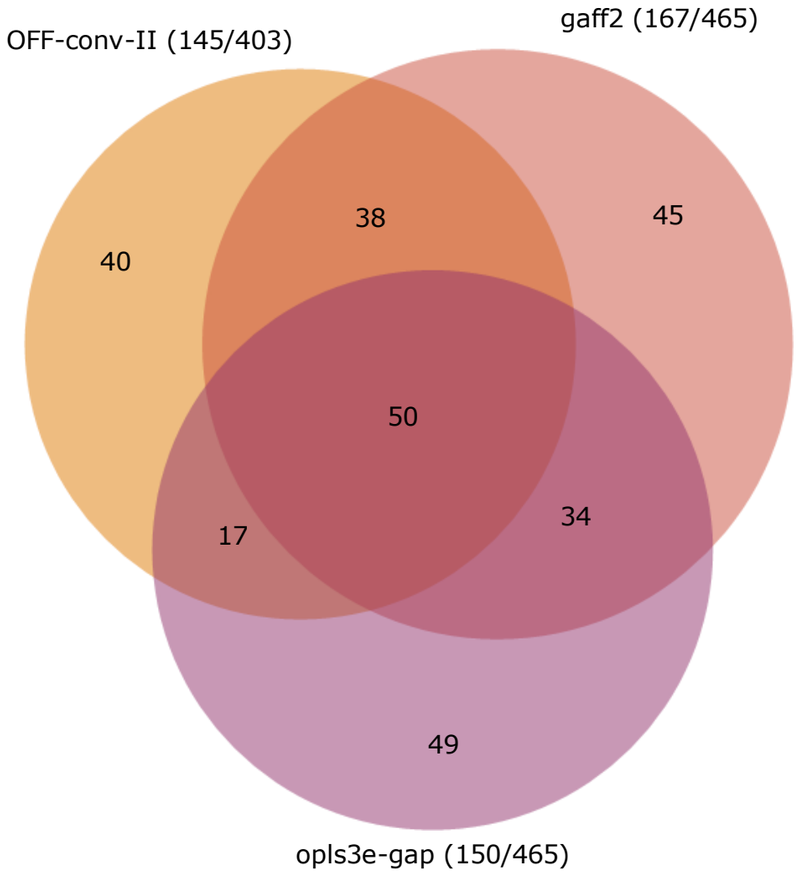
Best performing force fields based on 12 targets
Open Force Field (blue): 1,
GAFF2.1 (red): 3 and
OPLS3e (green): 8

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

Successes with $\Delta(\Delta\Delta G) <= 1.0$ kcal mol^{-1}



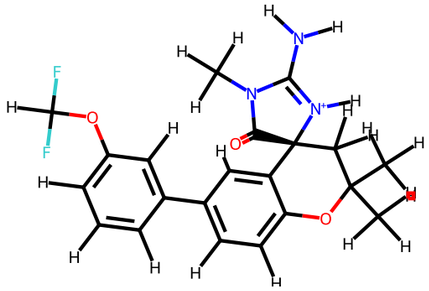
Outliers with $\Delta(\Delta\Delta G) > 1.0$ kcal mol^{-1}

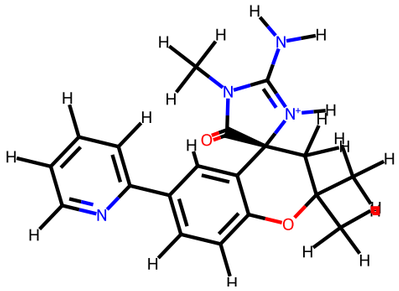


Analyzing the outlier groups can lead to insights about parameters sets

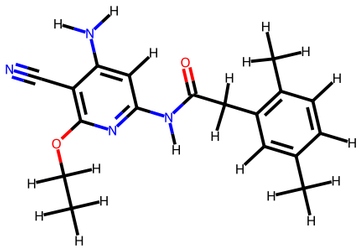
Outliers with $\Delta(\Delta\Delta G) > 2.0$ kcal mol^{-1}

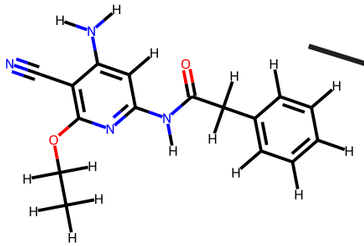
bace_hunt: 26 -> 32

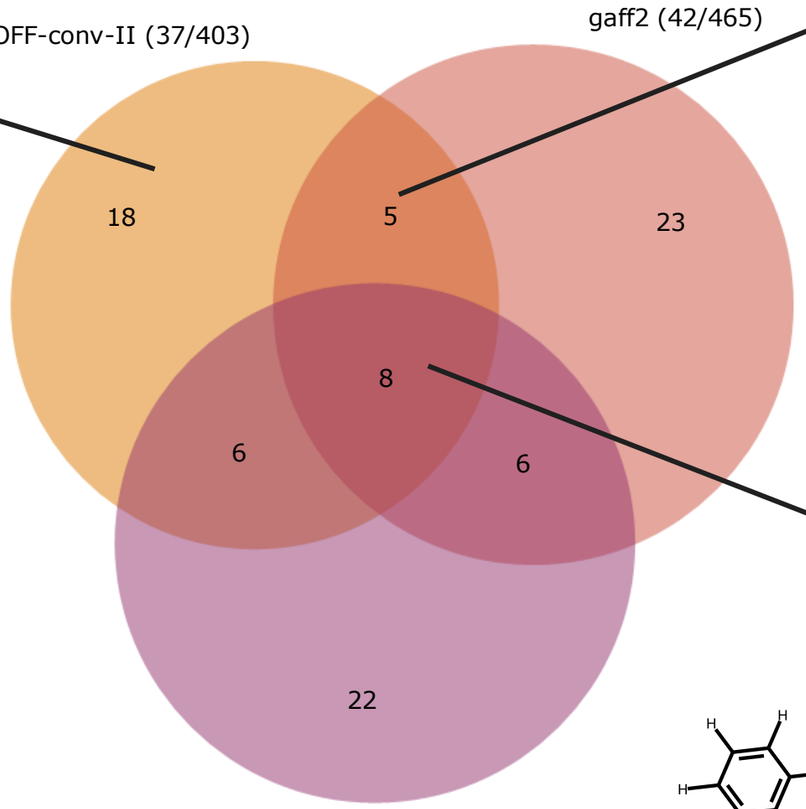




jnk1: 18635-1 -> 18624-1





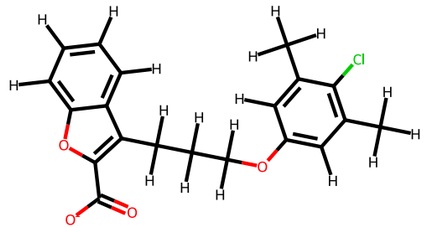


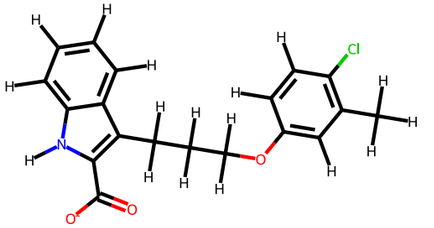
OFF and GAFF:
Atom mutation in heterocycles

OFF only

All force fields:
Set-up errors

mcl1: 67 -> 35



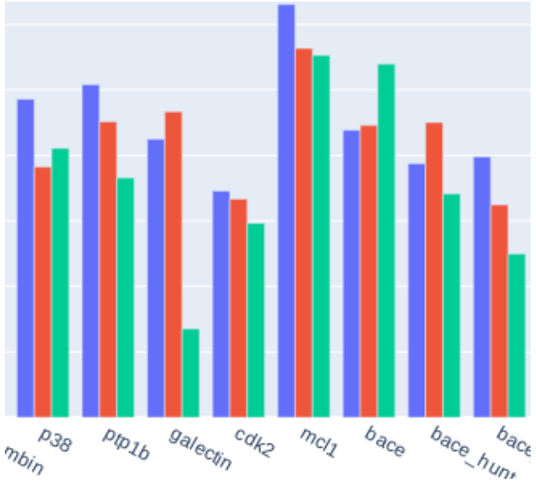


Summary & Outlook

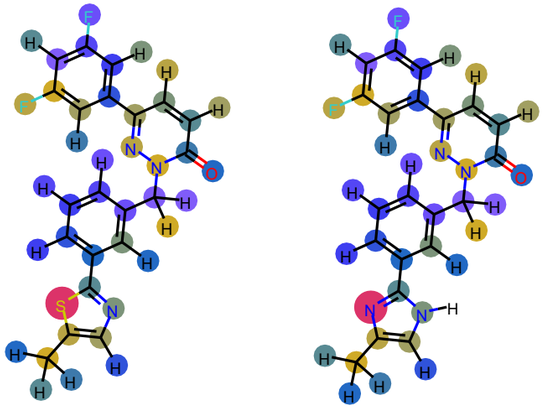
Skin cells at 20x magnification

Summary and Outlook

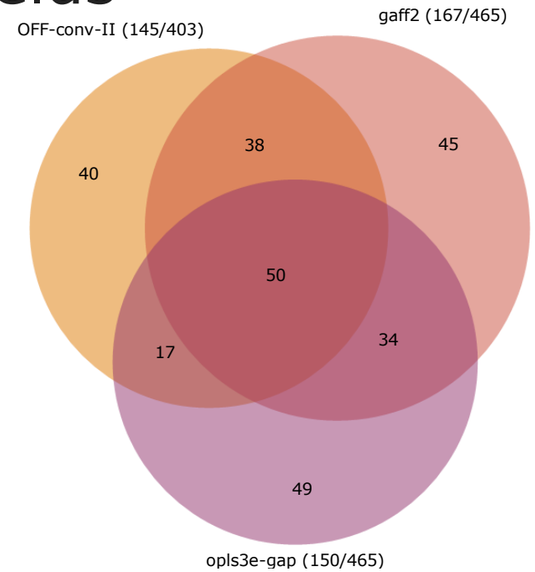
- Encouraging OpenFF results



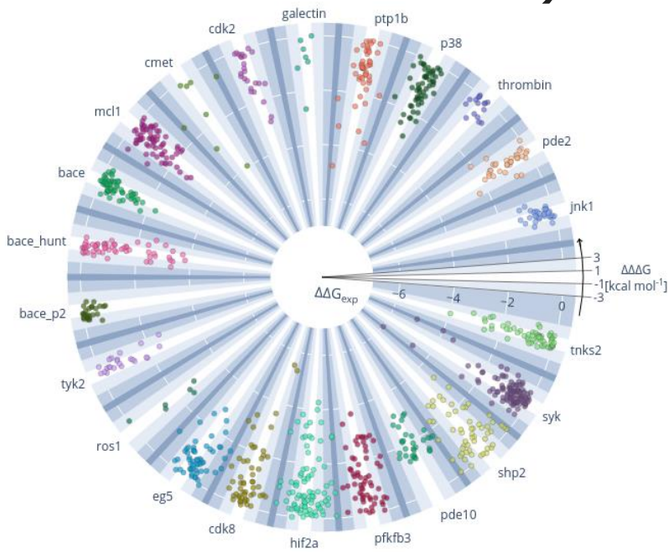
- Detection of issues and improved input



- Further analysis, detect set-up issues and compare force fields



- Further refine benchmark dataset (improve set-up, add and remove data)



GitHub

Data set: github.com/openforcefield/PLBenchmarks
pmx workflow: github.com/dfhahn/pmx/tree/py3
FE Framework: github.com/openforcefield/arsenic
Analysis: github.com/dfhahn/benchmarkpl

Acknowledgements

- Gary Tresadern
- Laura Perez Benito
- Herman van Vlijmen

Resources on AWS

- David Mobley (UC Irvine)
- John Chodera (MSKCC, New York)
- Hannah Bruce MacDonald (MSKCC, New York)
- Jeff Wagner (San Diego)

OpenEye

- Christopher Bayly
- Gaetano Calabro

MPI Goettingen

- Bert de Groot
- Vytas Gapsys
- Yuriy Khalak

Resources by

Thanks for listening!

Skin cells at 20x magnification