COVID MOONSHOT

DOCKING AND FREE ENERGY CALCULATIONS

John D. Chodera
MSKCC Computational and Systems Biology Program
http://www.choderalab.org

FOLDING @HOME

DISCLOSURES:

- Scientific Advisory Board, OpenEye Scientific
  All funding: http://choderalab.org/funding

9 Apr 2020 - BioSolveIT webinar

ENSEMBLE DOCKING WITH SHAPE OVERLAY WAS USED TO TRIAGE COMPOUNDS WITH POOR FITS

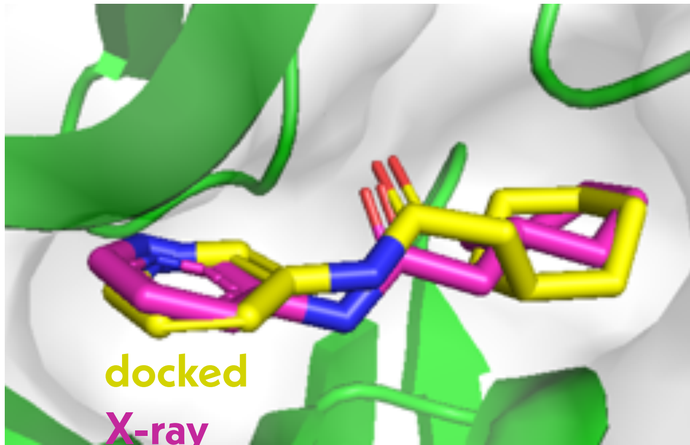
Goal: **Filter out problematic designs** that **can't fit or recapitulate interactions**

* Initial poses selected via **shape and color overlay** with **inspiration fragments**
* Docked to all crystal structures listed for **inspiration fragments**
* Minimized poses scored with **Chemgauss4** scoring function
* Best-scoring pose from all structures selected
* Very poor scores/poses triaged

Focus was on noncovalent complex, rather than covalent.
Bringing in more folks to help with covalent docking.

Implemented in Python with the OpenEye OEDocking Toolkit*

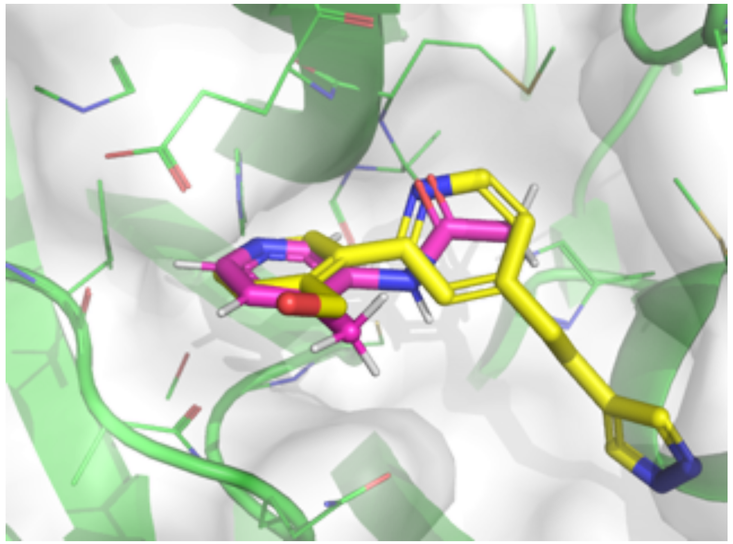
x0678 redocked into all structures



* Free for academics engaged in open science! https://www.eyesopen.com/academic-licensing
All scripts and output files: http://github.com/foldingathome/covid-moonshot

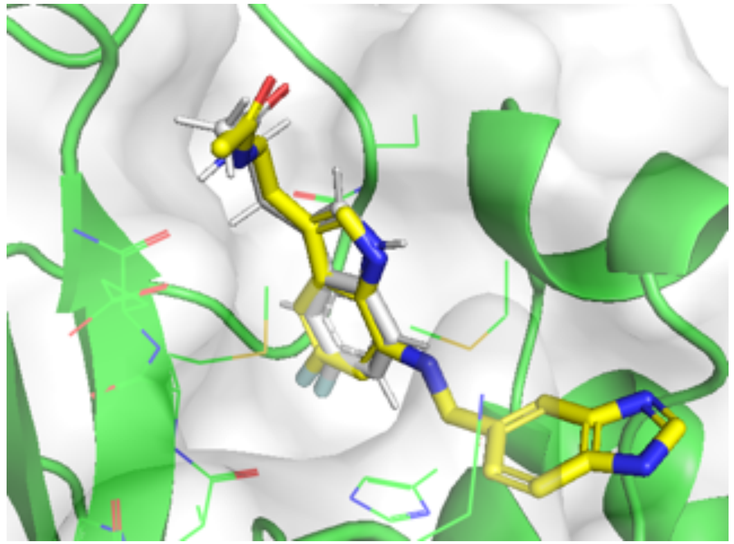
HYBRID ENSEMBLE DOCKING AIMS TO IDENTIFY COMPOUNDS THAT BUILD ON FRAGMENTS

isosteres



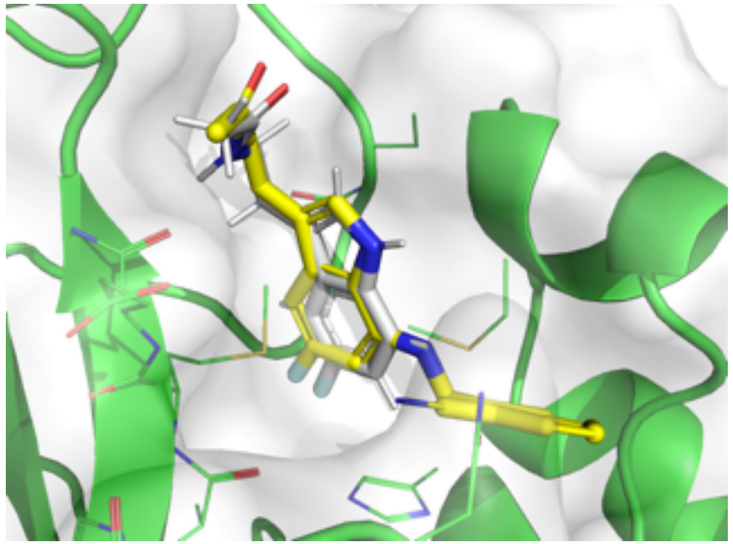
MUS-SCH-c2f-4

fragment extensions



GAB-REV-4a4-20

fragment extensions



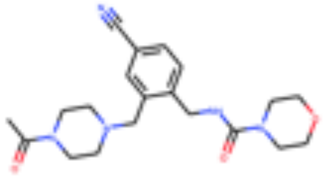
GAB-REV-4a4-13

* Free for academics engaged in open science! https://www.eyesopen.com/academic-licensing
All scripts and output files: http://github.com/foldingathome/covid-moonshot

DIVERSE SCAFFOLDS OF SUBMITTED COMPOUNDS PRESENTS A CHALLENGE TO ACCURATE SCORING

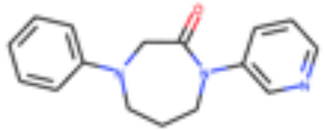
Submissions

JAN-GHE-fd8



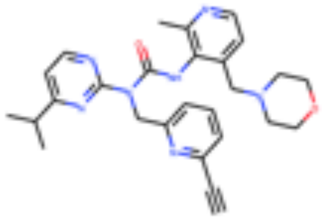
View

DAR-DIA-fc9



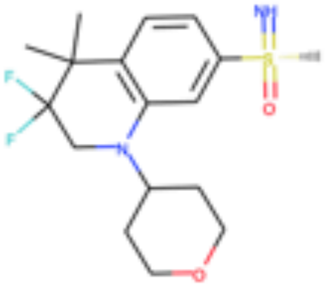
View

AGN-NEW-fad



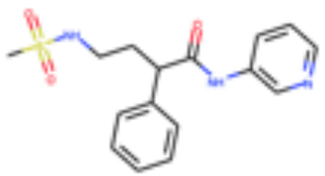
View

DAV-AUT-fa2



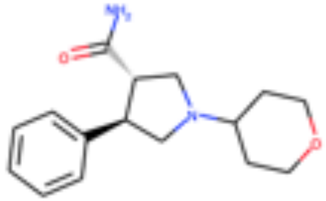
View

ADA-UNI-f8e



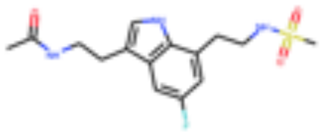
View

DUN-NEW-f8c



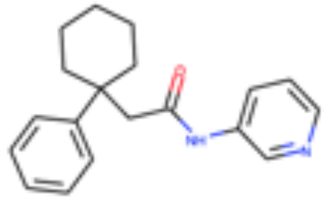
View

PET-SGC-f81



View

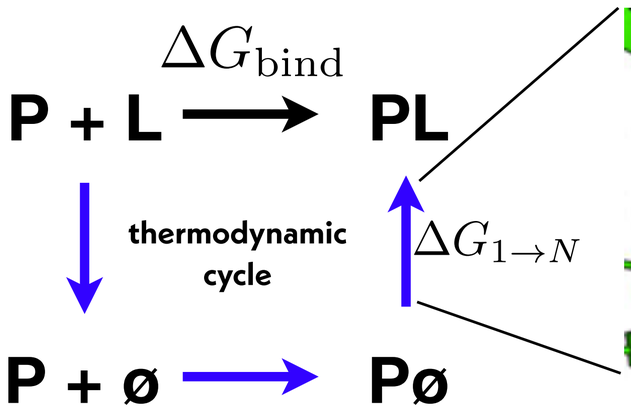
CHR-SOS-f73



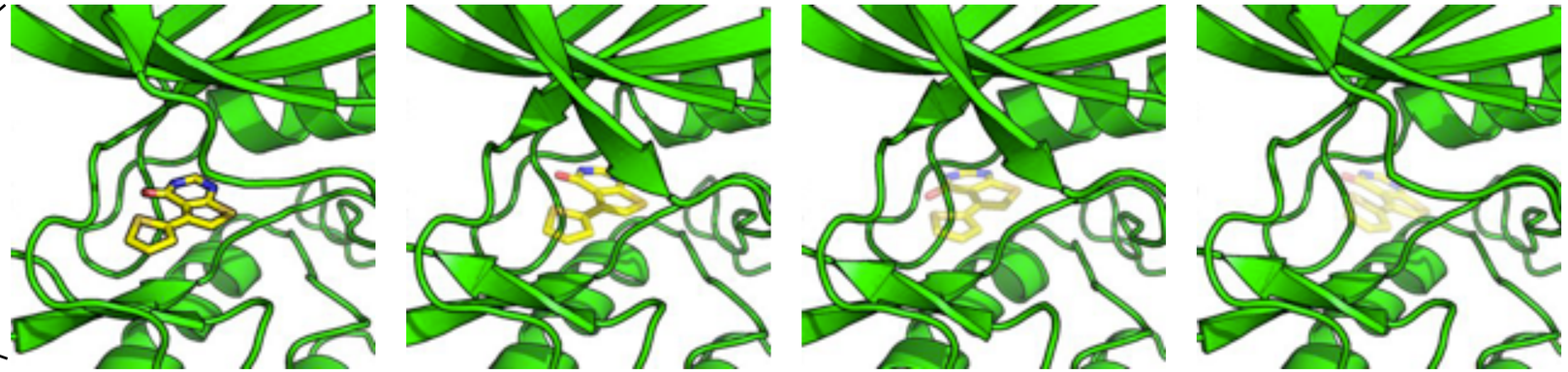
View

https://covid.postera.ai/covid/submissions

ABSOLUTE ALCHEMICAL FREE ENERGY CALCULATIONS PROVIDE A STRUCTURE-ENABLED WAY TO COMPUTE BINDING AFFINITIES



multiple simulations of alchemical intermediates



restraint imposition | discharging | steric decoupling | noninteracting

By breaking the problem into statistically easily computable pieces, calculation can be completed in just hours

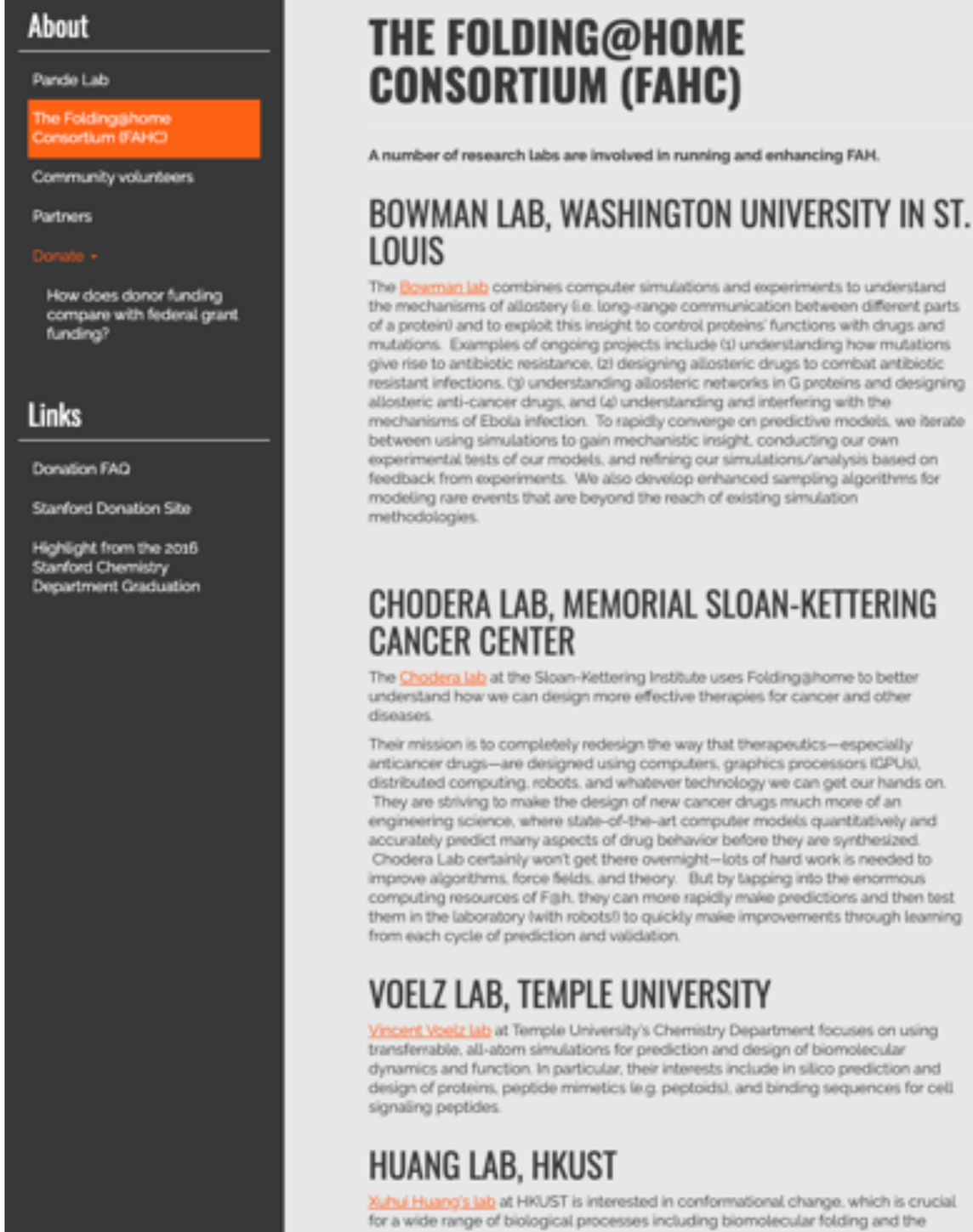
$$\Delta G_{1\to N} = -\beta^{-1} \ln \frac{Z_N}{Z_1} = -\beta^{-1} \ln \frac{Z_2}{Z_1} \cdot \frac{Z_3}{Z_2} \cdots \frac{Z_N}{Z_{N-1}}$$

$$Z_n = \int dx\, e^{-\beta U_n(x)}$$ partition function

Pioneering work from many: McCammon, van Gunsteren, Kollman, Jorgensen, Chipot, Roux, Boresch, Fujitani, Pande, Shirts, Swope, Christ, Mobley, Schrödinger, and many more

WE MOBILIZED THE FOLDING@HOME CONSORTIUM TO FOCUS ON COVID-19

* **generating structural ensembles** to enable small molecule drug discovery
* **identifying cryptic pockets** with the potential for allosteric inhibition
* **prioritizing compound synthesis** with alchemical free energy calculations
* **resolving key steps in the viral life cycle** in atomistic detail with Markov state models



About

Pande Lab

The Folding@home Consortium (FAHC)

Community volunteers

Partners

Donate

How does donor funding compare with federal grant funding?

Links

Donation FAQ

Stanford Donation Site

Highlight from the 2016 Stanford Chemistry Department Graduation

THE FOLDING@HOME CONSORTIUM (FAHC)

A number of research labs are involved in running and enhancing FAH.

BOWMAN LAB, WASHINGTON UNIVERSITY IN ST. LOUIS

The Bowman lab combines computer simulations and experiments to understand the mechanisms of allostery (i.e. long-range communication between different parts of a protein) and to exploit this insight to control proteins' functions with drugs and mutations. Examples of ongoing projects include (1) understanding how mutations give rise to antibiotic resistance, (2) designing allosteric drugs to combat antibiotic resistant infections, (3) understanding allosteric networks in G proteins and designing allosteric anti-cancer drugs, and (4) understanding and interfering with the mechanisms of Ebola infection. To rapidly converge on predictive models, we iterate between using simulations to gain mechanistic insight, conducting our own experimental tests of our models, and refining our simulations/analysis based on feedback from experiments. We also develop enhanced sampling algorithms for modeling rare events that are beyond the reach of existing simulation methodologies.

CHODERA LAB, MEMORIAL SLOAN-KETTERING CANCER CENTER

The Chodera lab at the Sloan-Kettering Institute uses Folding@home to better understand how we can design more effective therapies for cancer and other diseases.

Their mission is to completely redesign the way that therapeutics—especially anticancer drugs—are designed using computers, graphics processors (GPUs), distributed computing, robots, and whatever technology we can get our hands on. They are striving to make the design of new cancer drugs much more of an engineering science, where state-of-the-art computer models quantitatively and accurately predict many aspects of drug behavior before they are synthesized. Chodera Lab certainly won't get there overnight—lots of hard work is needed to improve algorithms, force fields, and theory. But by tapping into the enormous computing resources of F@h, they can more rapidly make predictions and then test them in the laboratory (with robots!) to quickly make improvements through learning from each cycle of prediction and validation.

VOELZ LAB, TEMPLE UNIVERSITY

Vincent Voelz lab at Temple University's Chemistry Department focuses on using transferrable, all-atom simulations for prediction and design of biomolecular dynamics and function. In particular, their interests include in silico prediction and design of proteins, peptide mimetics (e.g. peptoids), and binding sequences for cell signaling peptides.

HUANG LAB, HKUST

Xuhui Huang's lab at HKUST is interested in conformational change, which is crucial for a wide range of biological processes including biomolecular folding and the

THANKS TO OUR DONORS, FOLDING@HOME PROVIDES SIGNIFICANT COMPUTATIONAL RESOURCES TO ACCELERATE COVID-19 RESEARCH

a few weeks ago

Client statistics by OS

OS Type	Native TFLOPS*	x86 TFLOPS*	Active CPUs	Active Cores	Total CPUs
Windows	857	857	67,467	187,104	5,857,235
Mac OS X	91	91	8,083	85,382	217,033
Linux	87	87	6,383	26,457	882,200
NVIDIA GPU	1	2	4	4	348,371
ATI GPU	10,243	21,613	7,178	7,178	426,335
NVIDAI Fermi GPU	36,065	76,097	21,570	21,587	624,822
Total	47,344	98,747	110,685	327,712	8,355,996

1924085 people have non-anonymously contributed to Folding@home.

Table last updated at Sat, 19 Oct 2019 18:23:11

DB date 2019-10-19 23:22:42

Active CPUS are defined as those which have returned WUs within 50 days. The FLOPS per core was last updated based on a FAH core performance report run on Wed May 11 11:56:35 PDT 2016.

*TFLOPS is the actual teraflops from the software cores, not the peak values from CPU/GPU specs. Please see our FAQ and FLOPS FAQ.

~100 pflop/s

now

Folding@home

Team Monthly | Team | Donor | OS Stats

Active CPUs & GPUs by OS

OS	AMD GPUs	NVidia GPUs	CPUs	CPU cores	TFLOPS	x86 TFLOPS
Windows	110,331	429,105	771,877	5,371,818	943,128	1,914,273
Linux	6,766	106,664	358,737	3,472,940	235,229	443,906
macOSX	6	0	74,836	389,031	4,366	4,375
Totals	117,103	535,769	1,205,450	9,233,789	1,182,723	2,362,554

CPUs and GPUs which have returned Work Units within the last 50 days are listed by OS. FLOPS per core is estimated.

TFLOPS is Tera Floating-point OPerations per Second or trillions of math operations per second. Please see our FLOPS FAQ for more information.

Reported on Wed, 08 Apr 2020 15:22:05 GMT

~2.4 exaflop/s

SOME SAMPLING SCHEMES ARE MORE CLOUD-FRIENDLY THAN OTHERS

Independent simulations

Easy to parallelize, but sampling problems at any λ can make calculations unreliable

simple but dangerous

AMBER18 TI
Song, Lee, Zhu, York, Merz 2019
https://doi.org/10.1021/acs.jcim.9b00105

Hamiltonian replica exchange ★

Good sampling at any λ can rescue problems at other λ if good λ overlap

reliable but complex and costly

Schrödinger FEP+
Wang, Wu, Deng, Kim, ... Abel 2015
https://doi.org/10.1021/ja512751q

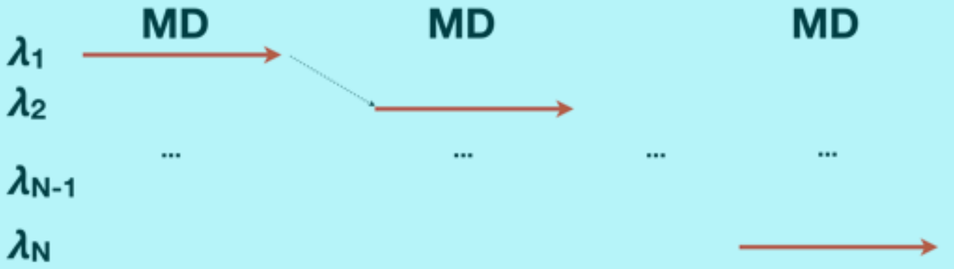
NAMD
Jiang, Thirman, Jo, Roux 2018
http://doi.org/10.1021/acs.jpcb.8b03277

also **OpenMM** (via **openmmtools**)

Single-replica methods

For certainly problems, can converge extremely quickly in a fraction of computer effort; tricky to make reliable

promising but relatively immature

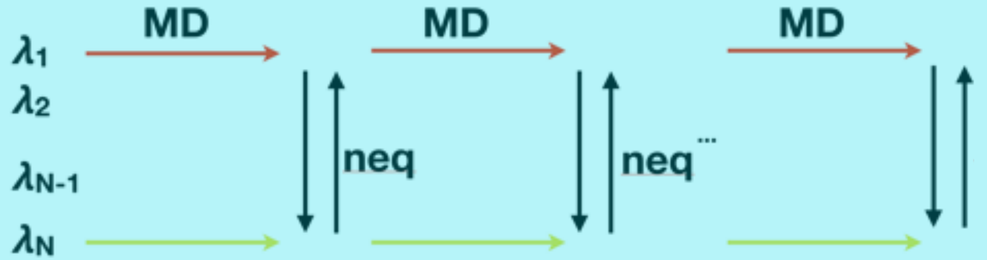


Hongzhi, Fayer, Wang 2006
https://doi.org/10.1063/1.2424700
Tan 2017
https://doi.org/10.1080/10618600.2015.1113

Nonequilibrium methods

Less efficient than equilibrium calculations, but can work robustly and scalably if properly tuned

promising but relatively immature



pmx / gromacs
Aldeghi, Gapsys, de Groot 2018
https://doi.org/10.1021/acscentsci.8b00717

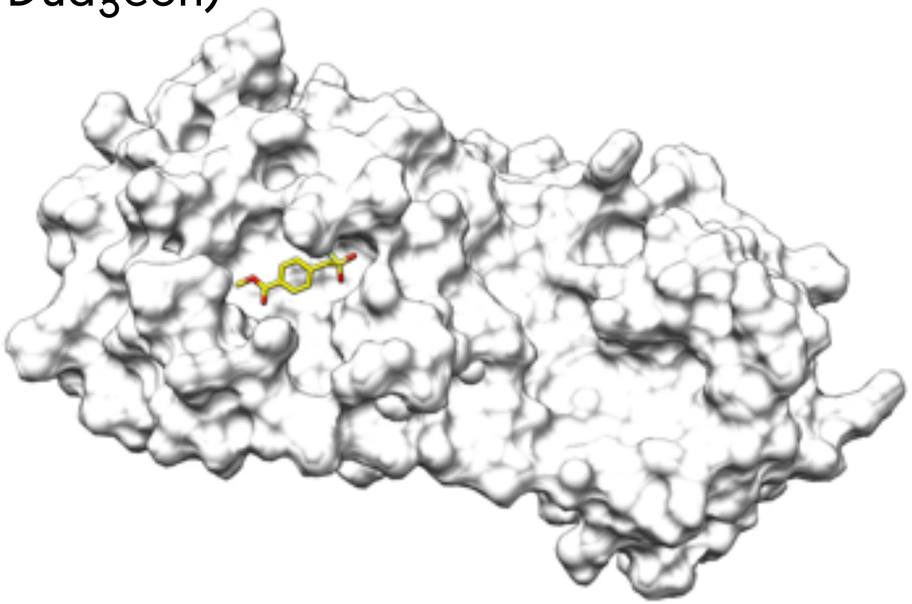
★ current best practice

FOLDING@HOME ALLOWS FREE ENERGY CALCULATIONS AT MASSIVE SCALE

- **5 988 free energy calculations on Moonshot compounds**
- **50 336+ Astex compound graph enumerated compounds** (from Tim Dudgeon)

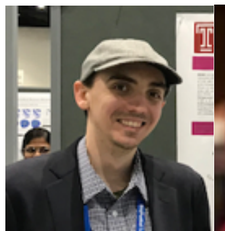
- Amber14SB + OpenFF 1.1.0 “Parsley” + TIP3P water
- Built / minimized / equilibrated with CUDA-accelerated OpenMM
- 4 fs timesteps using 4 amu hydrogens

- Single-replica (self-adjusted mixtures sampling)
- 40 alchemical states per thermodynamic leg
- Weak harmonic ligand restraint to binding pocket
- Five independent replicates for each complex
- **563 240 total trajectories running on Folding@Home**



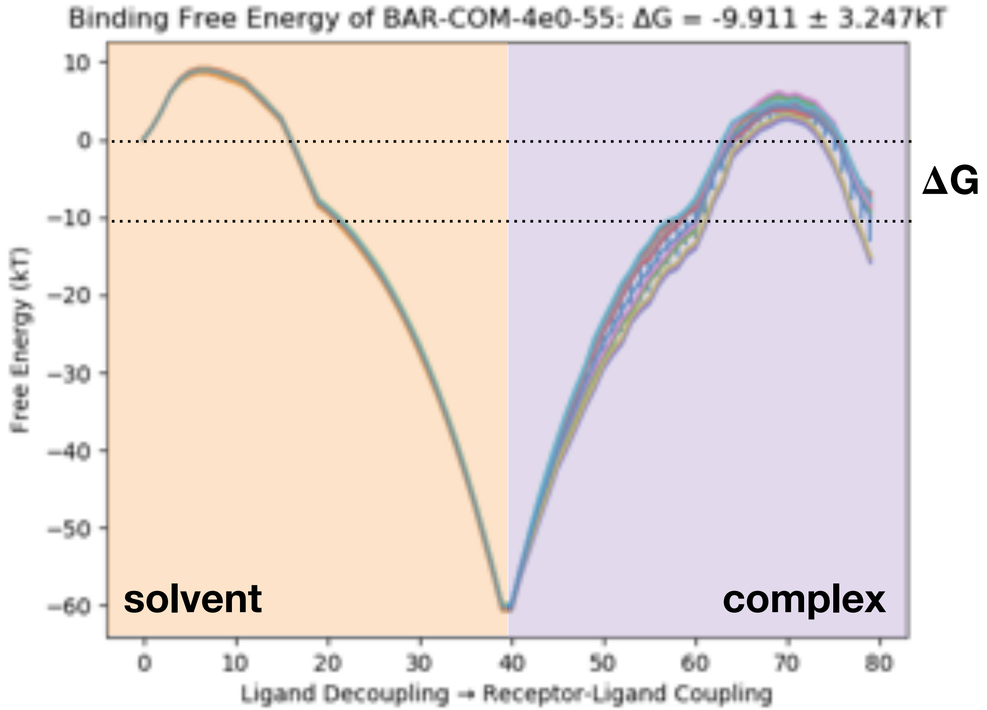
@voelzlab
voelzlab.org

Matt Hurley
Vincent Voelz
Voelz Lab
Temple University



ABSOLUTE FREE ENERGY CALCULATIONS ARE CHALLENGING TO CONVERGE EVEN AT MASSIVE SCALES

- Moonshot compounds have been running ~6 days
  - solvated complex: **448 µs MD (~70µs/day)**!
  - solvated ligand: **4.454 ms MD aggregate (~700µs/day)**!
- Extract complexes that appear to have converged:
  - Looking for sufficient complex sampling
  - Want systems with "small" Wang-Landau increments
- Initial results show need for more receptor-ligand sampling, but also already reveals ligands that may be promising



@voelzlab
voelzlab.org

Matt Hurley
Vincent Voelz
Voelz Lab
Temple University

Open Force Field Initiative

An open source, open science, and open data approach to better force fields

Download the Toolkit | Read the Docs | Get the Force Fields | View the Source

Join the Consortium as an Industry Partner to support high-quality biomolecular force fields and receive prioritized support.

View slides and presentations from our most recent Consortium Workshop held at UC San Diego on August 30-31, 2019.

Open Source

Software permissively licensed under the MIT License and developed openly on GitHub.

Open Science

Scientific reports on open access preprint servers bioRxiv and chemRxiv.

Open Data

Curated physical property and quantum chemical datasets for building high-quality force fields.

Developed in coordination with the NSF-funded Molecular Sciences Software Institute

MolSSI

http://openforcefield.org

The Open Force Field 1.0 small molecule force field, our first optimized force field (codename "Parsley")

At the end of our first year, the Open Force Field Consortium releases its first optimized force field: the Open Force Field 1.0 (codename "Parsley") small molecule force field

35 minute read, Published: 10 Oct, 2019

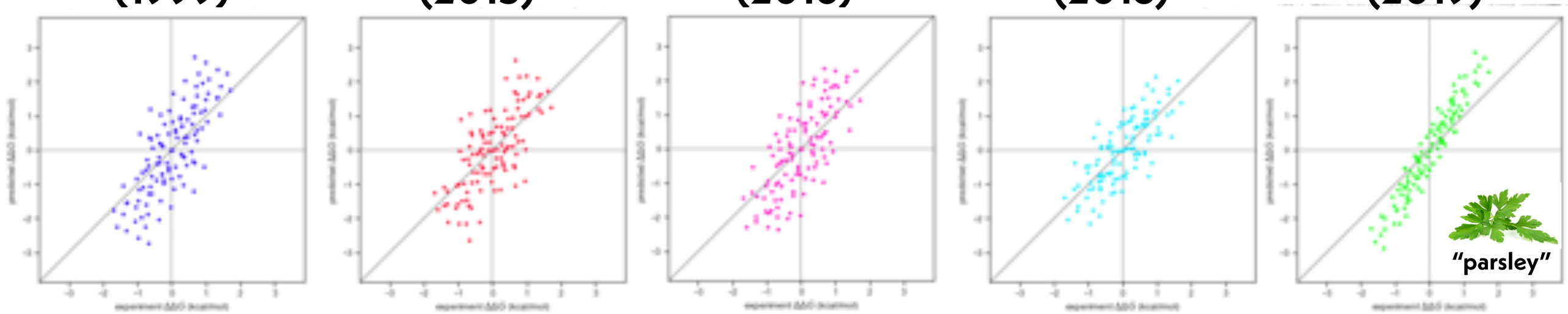
We're delighted to announce the release of "Parsley", the Open Force Field 1.0 small molecule force field---the first in a series of iteratively-improved small molecule force fields for biomolecular simulation funded in part by the Open Force Field Consortium. This is the first optimized force field to use the SMIRNOFF force field specification for atom type-free direct chemical perception, and provides substantially improved valence (bond, angle, and torsion) parameters relative to its predecessor, the AMBER-lineage SMIRNOFF99Frosst. This force field was optimized to improve agreement with quantum chemical geometries, energetics, and vibrational frequencies, and will likely provide improved accuracy (relative to its predecessor) for a wide variety of properties, especially energetics and geometries relative to gas phase quantum chemical calculations

https://openforcefield.org/news/introducing-openforcefield-1.0/

“PARSLEY” PROVIDES SIGNIFICANT ACCURACY IMPROVEMENTS

Open Force Field Initiative

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



thrombin
PDB101: 1PPB

HANNAH BRUCE MACDONALD
MSKCC

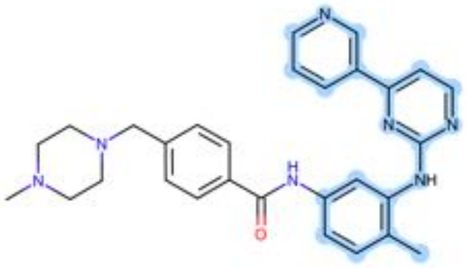
http://github.com/choderalab/perses

DOMINIC RUFA

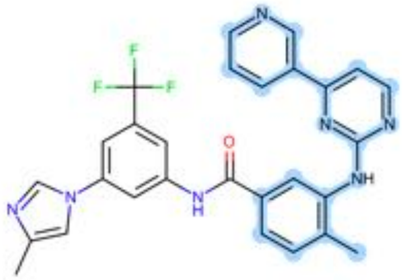
PERSES LITE: HYBRID TOPOLOGY ALCHEMICAL FREE ENERGY CALCULATIONS

Propose new molecules with common scaffold via MCSS

Imatinib

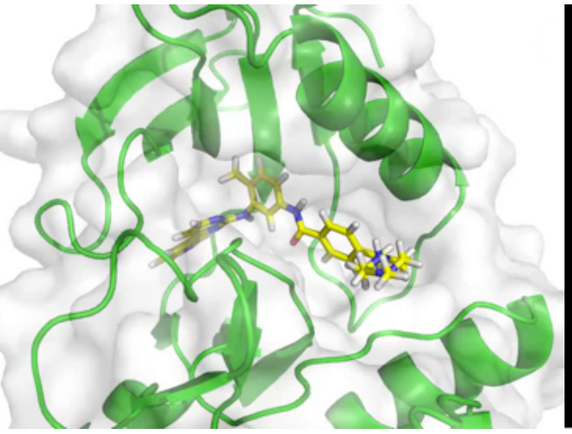


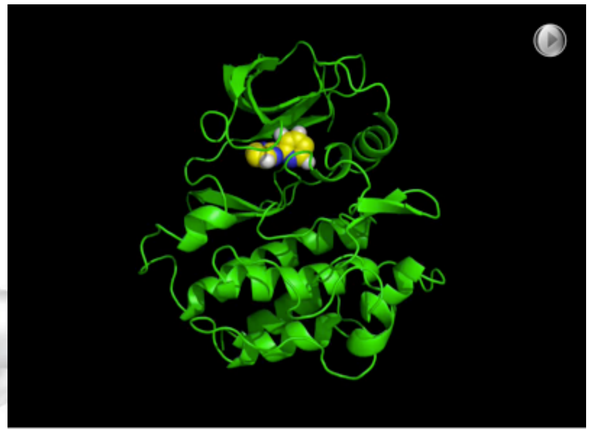
Nilotinib



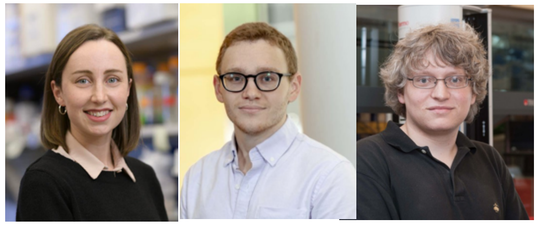
nonequilibrium switching

Build in new atoms with reversible-jump Monte Carlo





HANNAH BRUCE MACDONALD
DOMINIC RUFA
PATRICK GRINAWAY



SILICON Therapeutics

BRYCE ALLEN
WOODY SHERMAN

CHODERA LAB

National Institutes of Health

NSF

SCHRÖDINGER.

OpenEye SCIENTIFIC

SILICON Therapeutics

PARKER INSTITUTE for CANCER IMMUNOTHERAPY

Gerstner FAMILY FOUNDATION

STARR CANCER CONSORTIUM

Open Force Field Consortium

XtalPi

CYCLE FOR SURVIVAL

- Scientific Advisory Board, OpenEye Scientific

All funding: http://choderalab.org/funding