ALCHEMICAL FREE ENERGY CALCULATIONS

Differentiating **approaches**, measuring **impact**, and surveying the **open source ecosystem**

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

Talk slides available at http://choderalab.org/news/cadd-grc-2019

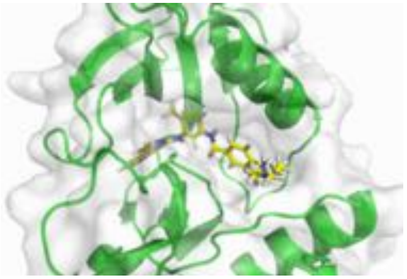
DISCLOSURES:

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

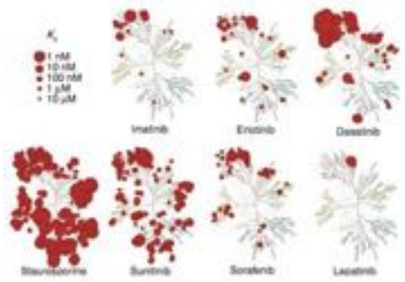
CADD GRC - July 15, 2019 - Mount Snow, VT

CHODERA LAB

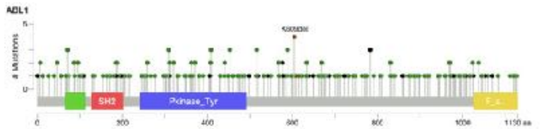
WHERE DO WE NEED NEW PHYSICAL MODELING TOOLS TO SOLVE PROBLEMS THAT CURRENTLY LACK GOOD PREDICTIVE MODELS?



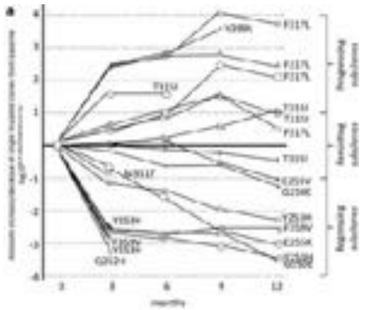
SELECTIVE INHIBITOR DESIGN: TARGETS/ANTITARGETS



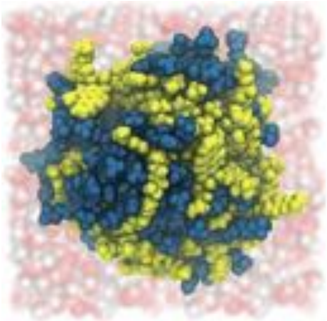
KINASE INHIBITOR SELECTIVITY



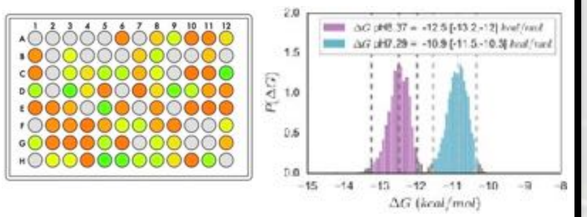
PREDICTING DRUG SENSITIVITY/RESISTANCE



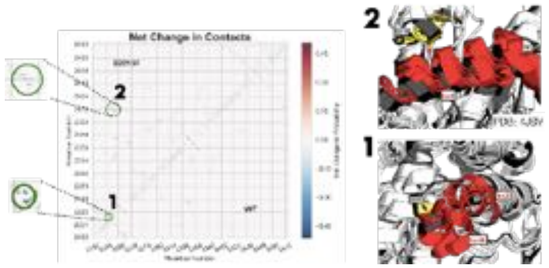
ANTICIPATING DRUG RESISTANCE



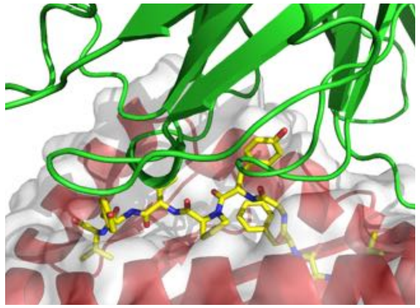
NOVEL DRUG DELIVERY MODALITIES



AUTOMATED BIOPHYSICAL ASSAYS AND INFERENCE

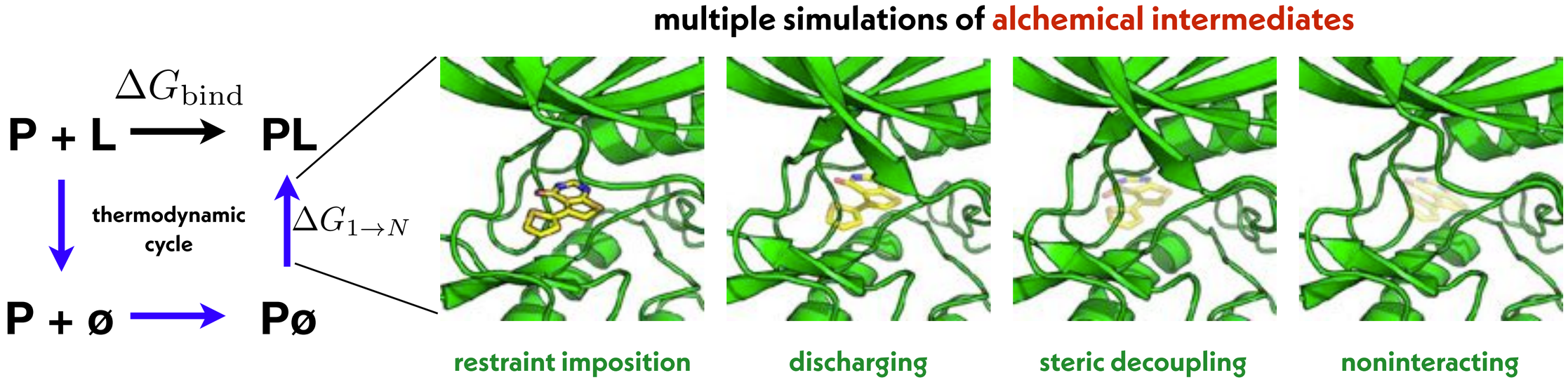


MECHANISMS OF ONCOGENIC ACTIVATION



CANCER IMMUNOTHERAPY

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



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

$$\Delta G_{1\to N} = -\beta^{-1} \ln \frac{Z_N}{Z_1} = -\beta^{-1} \ln \frac{Z_2}{Z_1} \cdot \frac{Z_3}{Z_2} \cdots \frac{Z_N}{Z_{N-1}} \qquad Z_n = \int dx\, e^{-\beta U_n(x)}$$

partition function

Isomorphic with computing Bayes factors or evidence ratios in inferential machine learning

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

WHAT ARE OUR GOALS?

Use free energy calculations as predictive models to impact nearly all structure-based programs

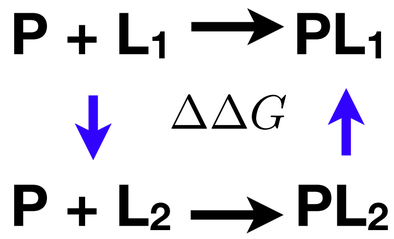
Enable statistically sound decisionmaking

Extend free energy calculations and physical modeling to impact other stages of drug discovery and clinical use

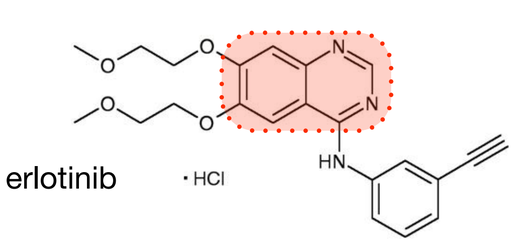
DIFFERENTIATING APPROACHES

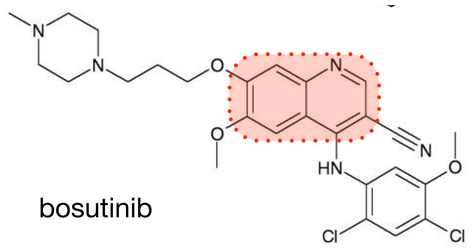
ALCHEMICAL FREE ENERGY CALCULATIONS CAN BE EITHER RELATIVE OR ABSOLUTE

RELATIVE



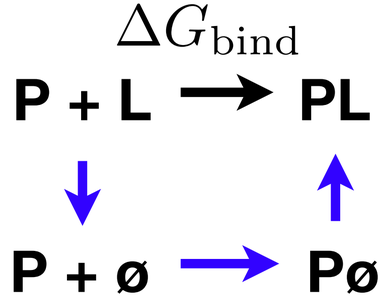
capable of **transforming a few atoms**
good for comparing **similar ligands**
requires same or **similar scaffolds**
requires **common scaffold to anchor series**



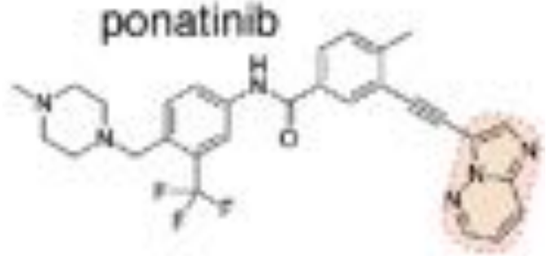


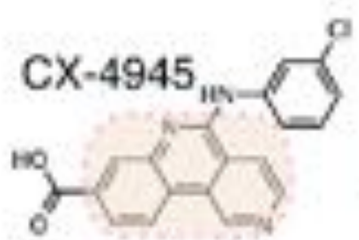
Cournia, Allen, Sherman 2017: http://dx.doi.org/10.1021/acs.jcim.7b00564

ABSOLUTE



capable of **disappearing a few atoms**
good for comparing **dissimilar ligands**
can use entirely **disparate scaffolds**
requires use of **restraints to anchor ligand**





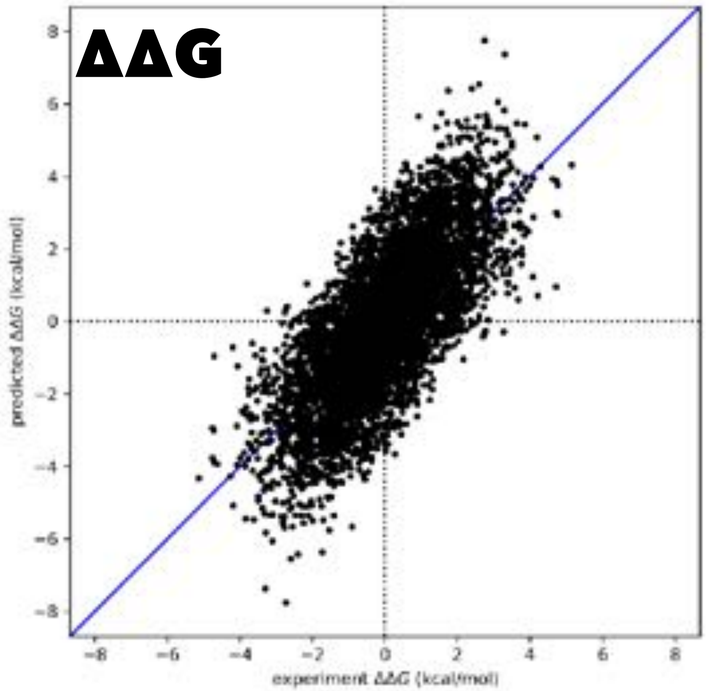
Aldeghi, Bluck, Biggin 2018: https://doi.org/10.1007/978-1-4939-7756-7_11

USEFUL ACCURACY IS SOMETIMES ACHIEVABLE

RELATIVE

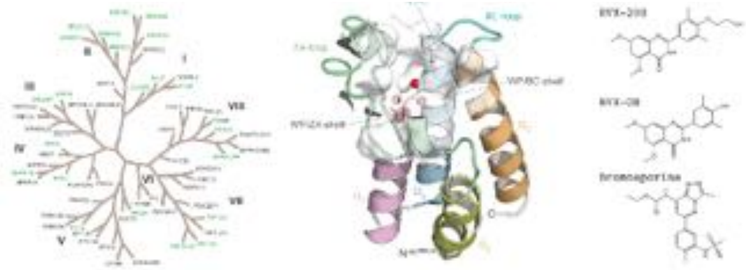
all within-target pairs ΔΔG (N = 5620)

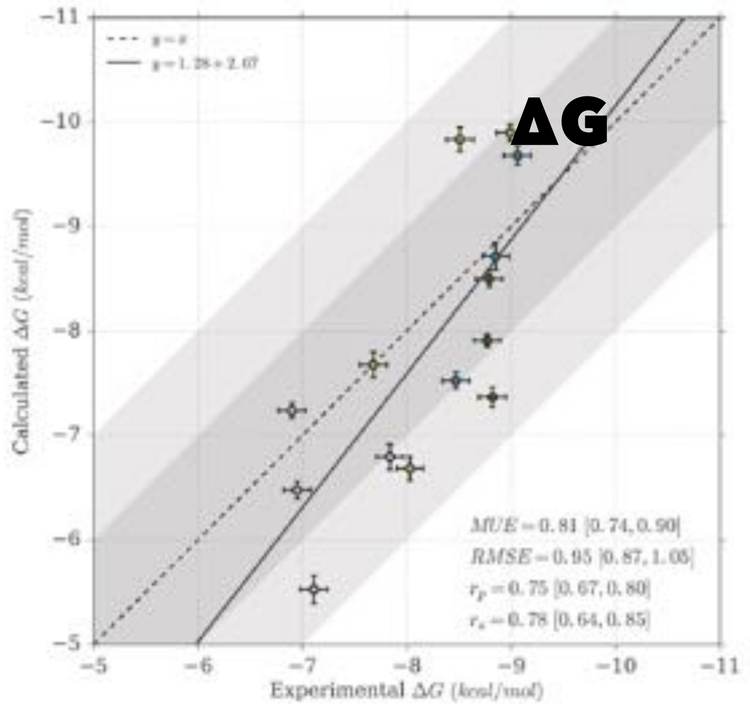
RMSE: OPLS	1.37 [95%:	1.34,	1.39] kcal/mol
MUE : OPLS	1.09 [95%:	1.07,	1.11] kcal/mol
R2 : OPLS	0.10 [95%:	0.06,	0.15] kcal/mol
rho : OPLS	0.73 [95%:	0.72,	0.74] kcal/mol



ΔΔG RMSE ~ 1.4 kcal/mol for well-behaved proteins/chemistries

ABSOLUTE



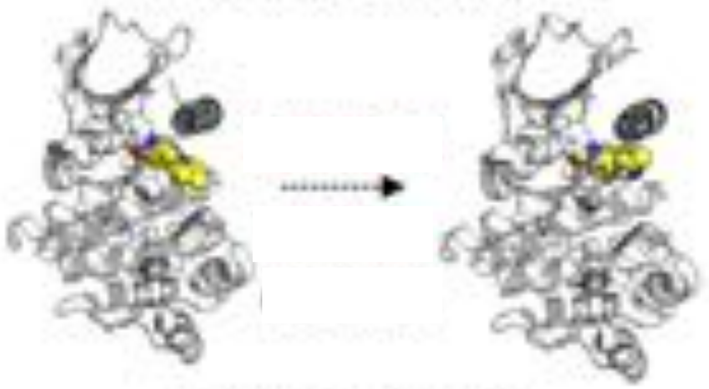


Wang et al. (Schrödinger) JACS 137:2695, 2015
https://doi.org/10.1021/ja512751q
Reanalysis: http://github.com/jchodera/jacs-dataset-analysis

Aldeghi et al. JACS 139:946, 2017.
https://doi.org/10.1021/jacs.6b11467

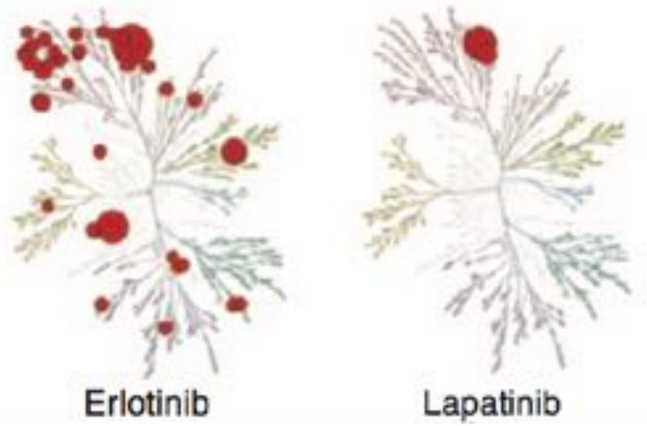
ALCHEMICAL FREE ENERGY CALCULATIONS HAVE MULTIPLE POTENTIAL APPLICATIONS IN DRUG DISCOVERY

driving affinity / potency



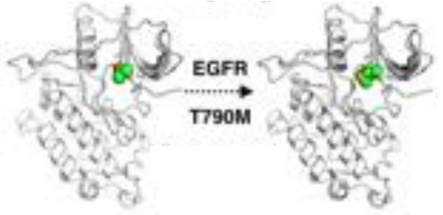
driving selectivity

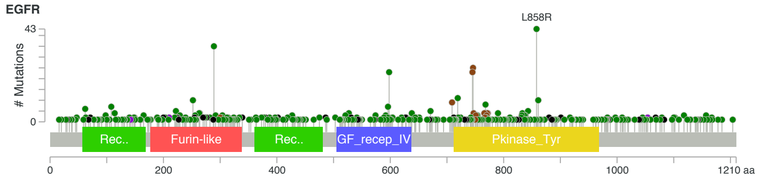
Moraca, Negri, de Olivera, Abel JCIM 2019
https://doi.org/10.1021/acs.jcim.9b00106



predicting clinical drug resistance/sensitivity

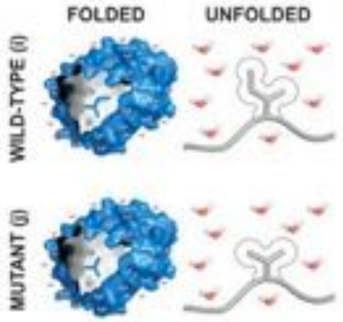
Hauser, Negron, Albanese, Ray, Steinbrecher, Abel, Chodera, Wang.
Communications Biology 1:70, 2018
https://doi.org/10.1038/s42003-018-0075-x

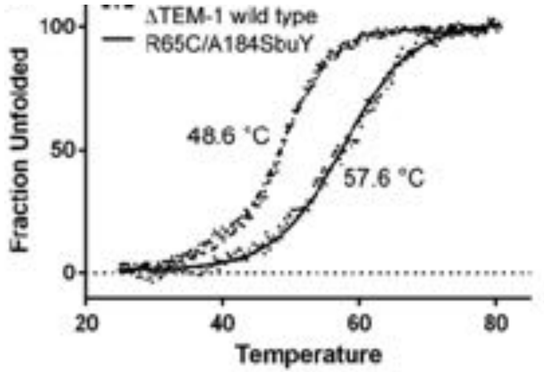




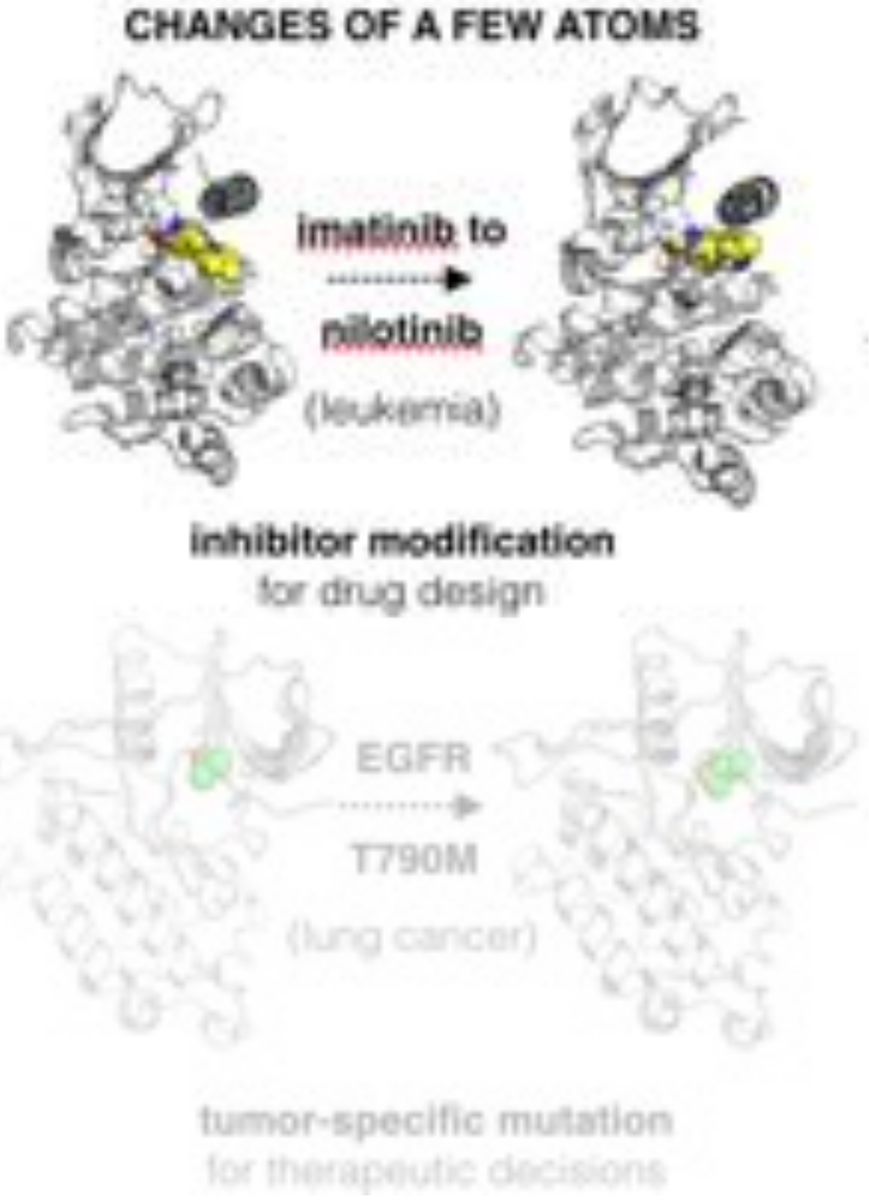
optimizing thermostability

Gapsys, Michielssens, Seeliger, and de Groot. Angew Chem 55:7364, 2016
https://doi.org/10.1002/anie.201510054





CONSIDER THESE QUESTIONS IN THE CONTEXT OF KINASE INHIBITORS FOR CANCER

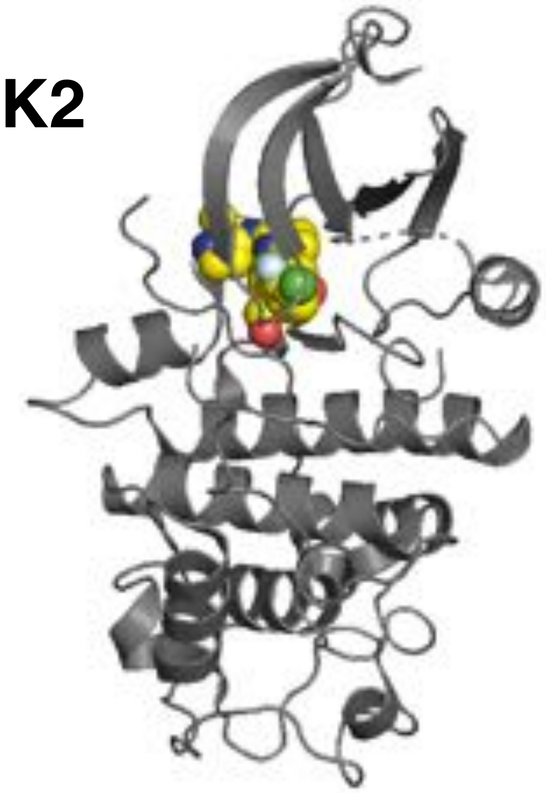


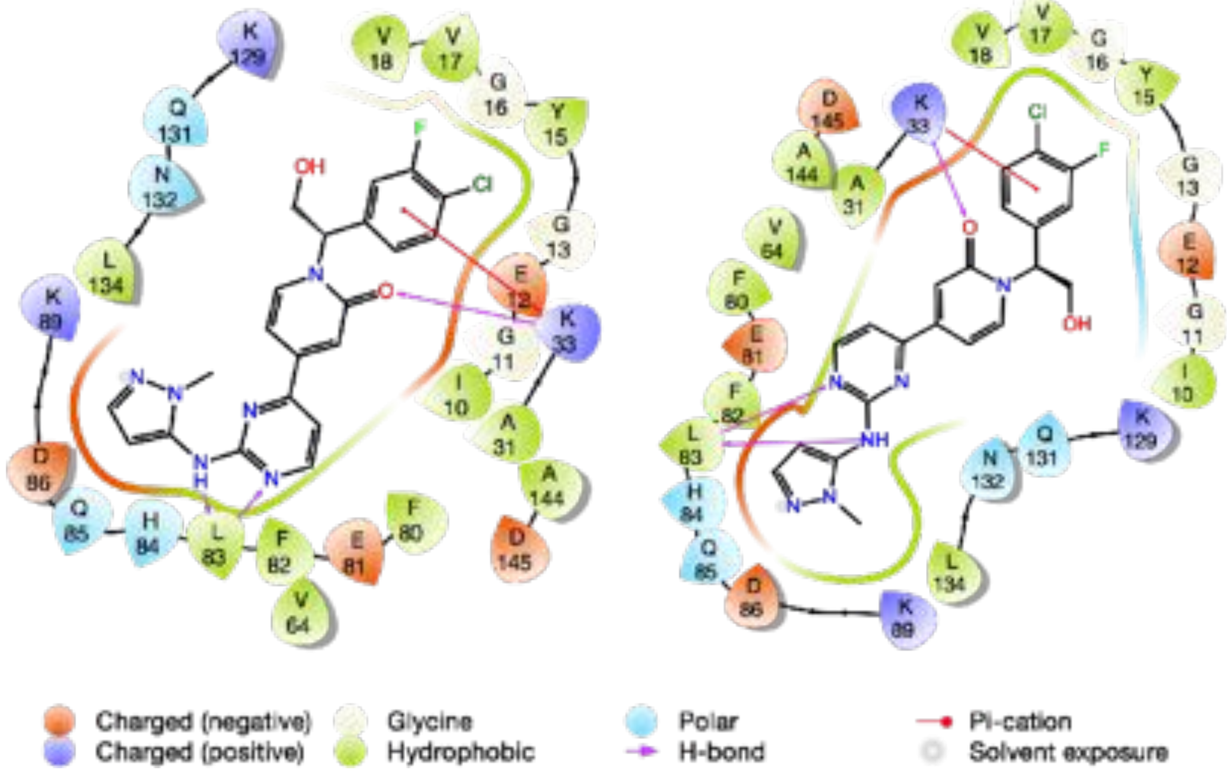
HOW CAN WE DESIGN SPECIFICALLY TARGETED CANCER DRUGS?

HOW CAN WE PREDICT DRUG RESISTANCE AND SUSCEPTIBILITY?

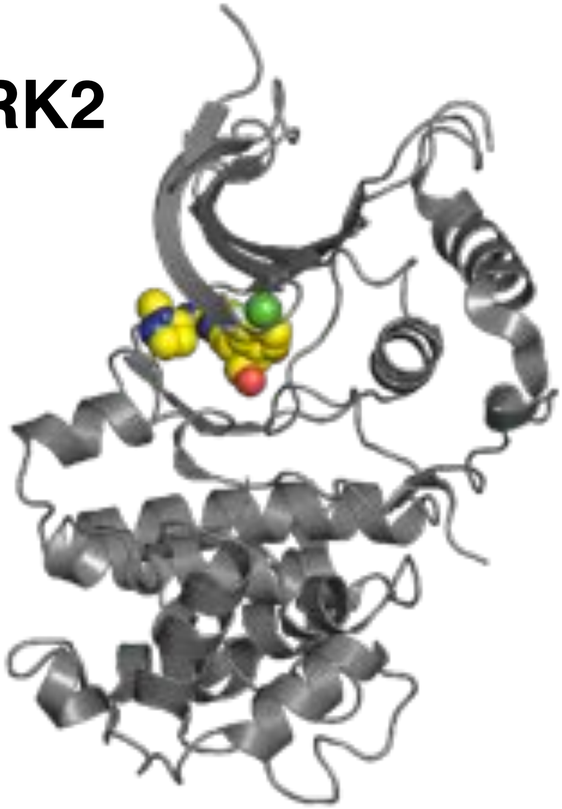
HOW MUCH DOES CANCELLATION OF FORCE FIELD ERROR HELP OR HURT SELECTIVITY PREDICTION?

CDK2





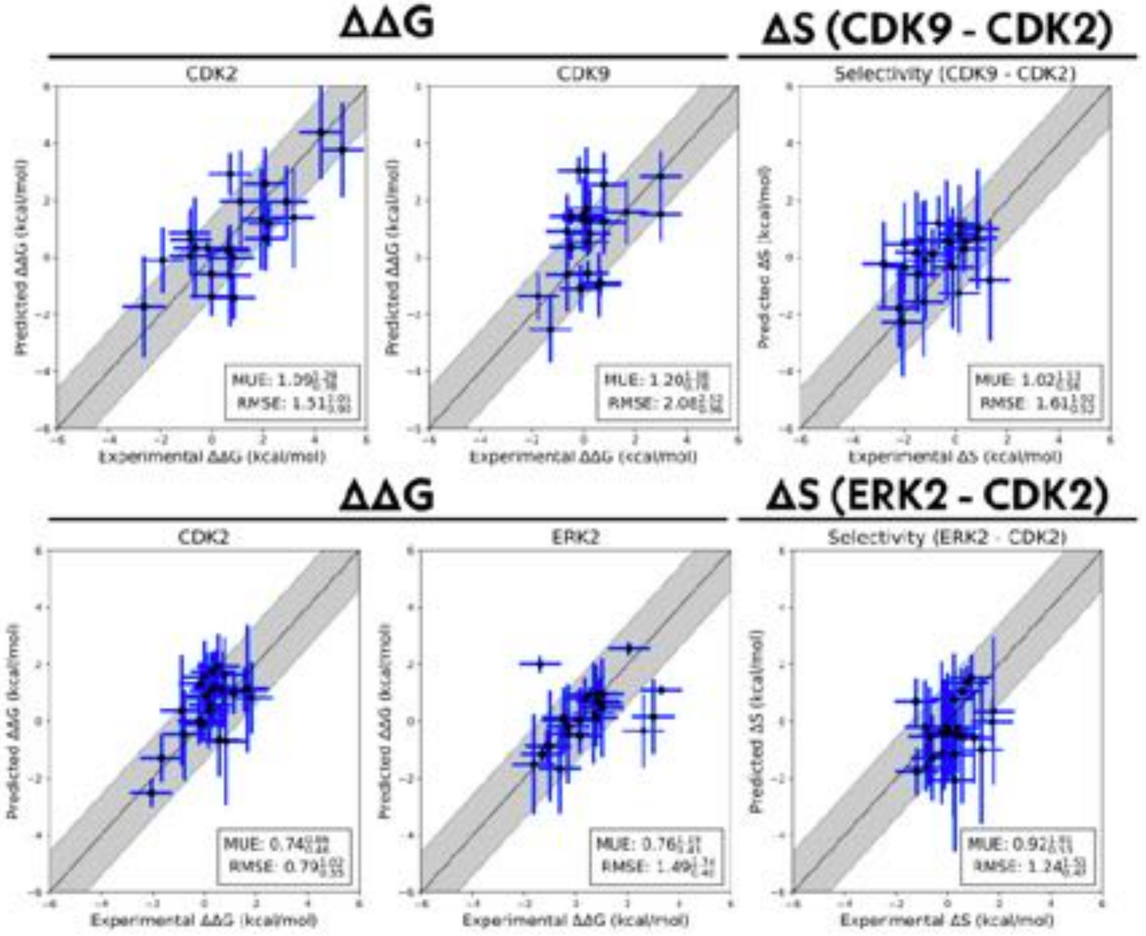
ERK2

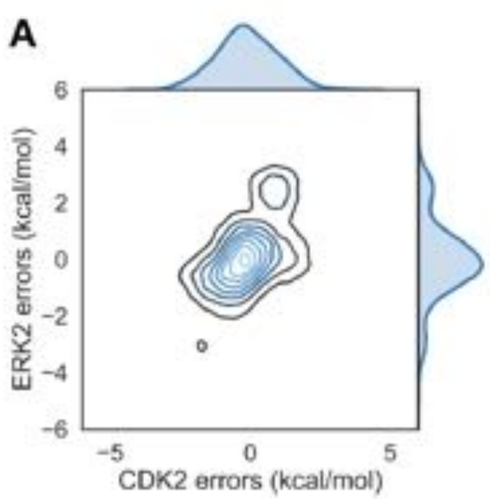


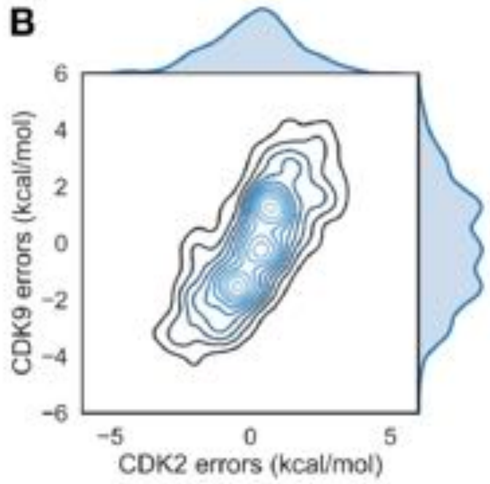
Quantify via the **correlation coefficient**

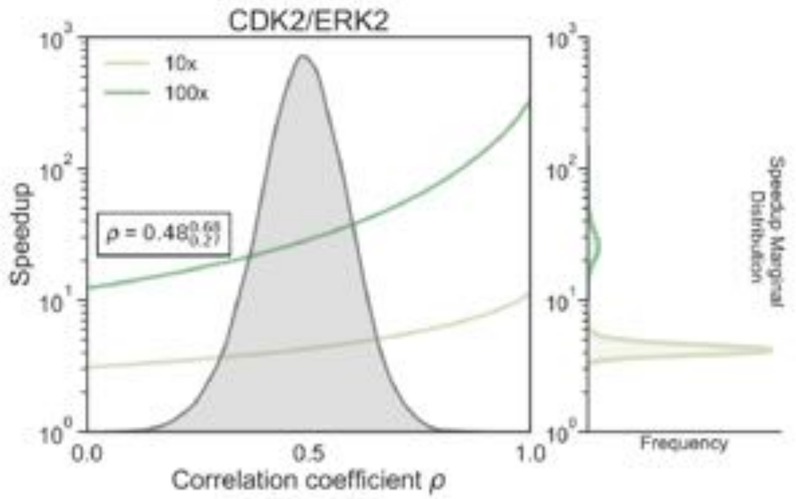
$$\rho \equiv \frac{\mathrm{cov}(\epsilon_1, \epsilon_2)}{\sqrt{\mathrm{var}(\epsilon_1)\mathrm{var}(\epsilon_2)}}$$ of the **error** $$\epsilon_* \equiv \Delta\Delta G_*^{\mathrm{FEP}} - \Delta\Delta G_*^{\mathrm{exp}}$$

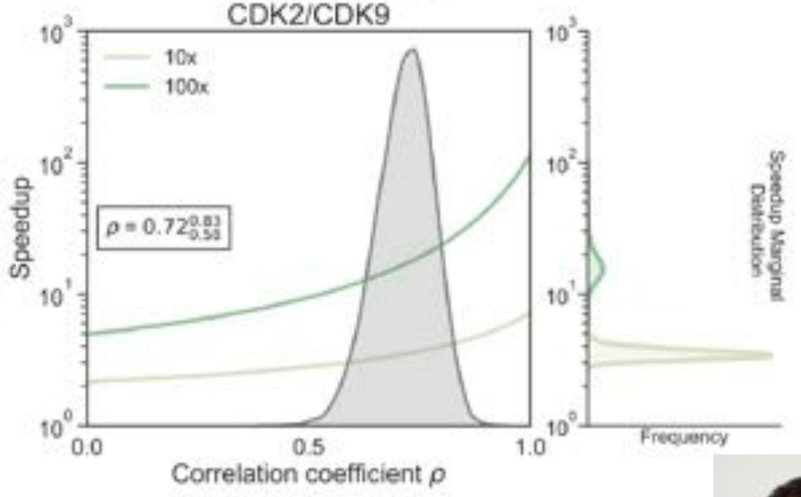
DIFFERENT SELECTIVITY PROBLEMS SHOW DIFFERENT DEGREES OF CANCELLATION











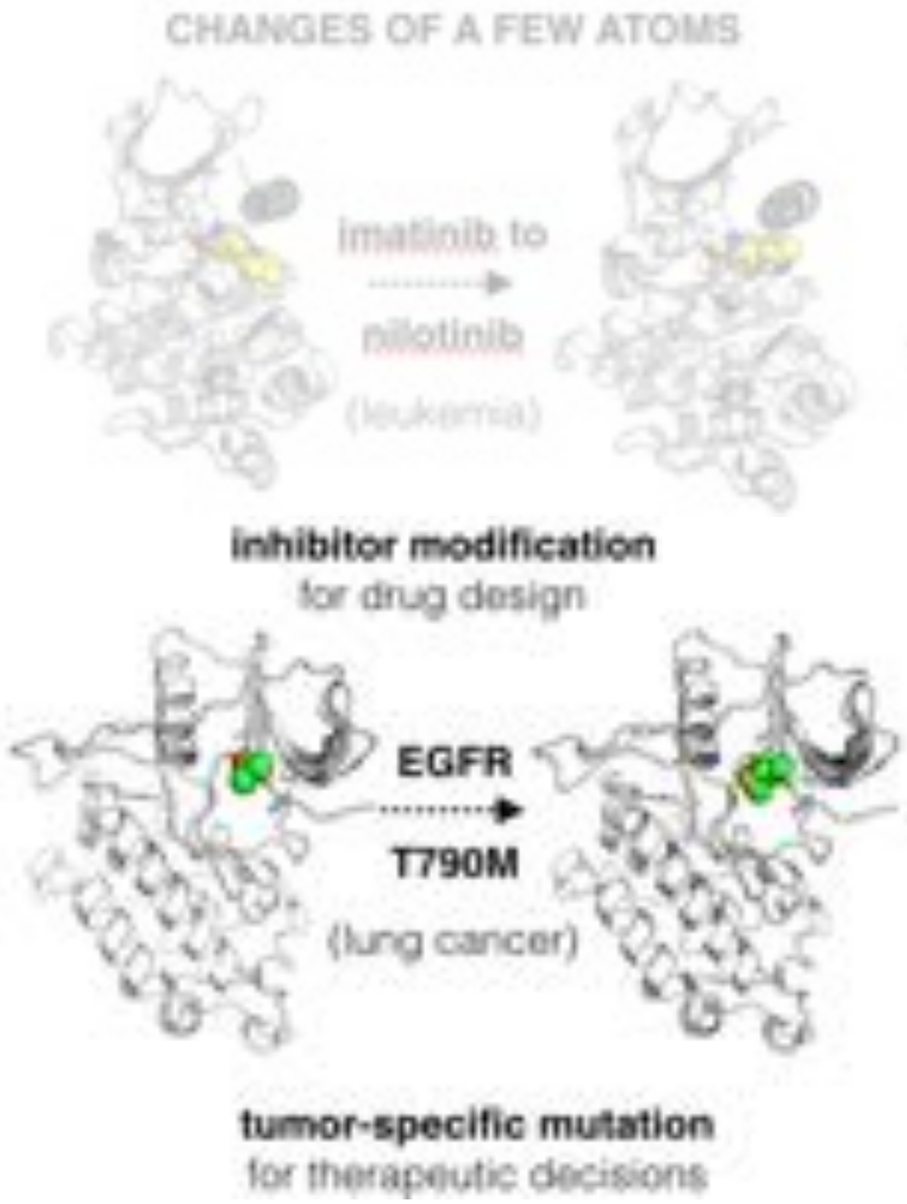
STEVEN ALBANESE

coming soon to bioRxiv

FEP+/OPLS3

LINGLE WANG
SCHRÖDINGER

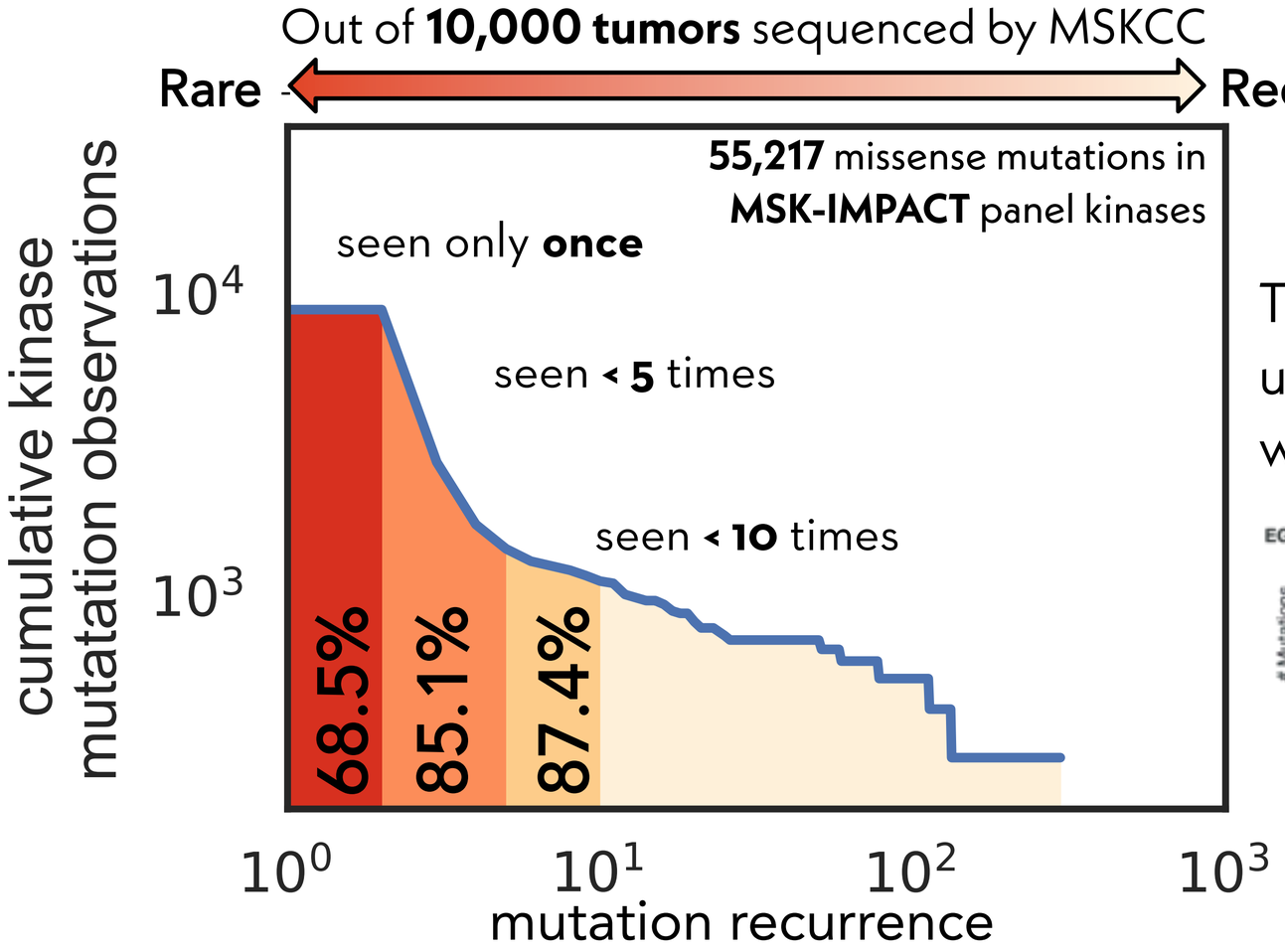
MANY QUESTIONS RELEVANT TO DRUG DISCOVERY INVOLVE COMPUTING THE ΔG OF MODIFYING JUST A FEW ATOMS



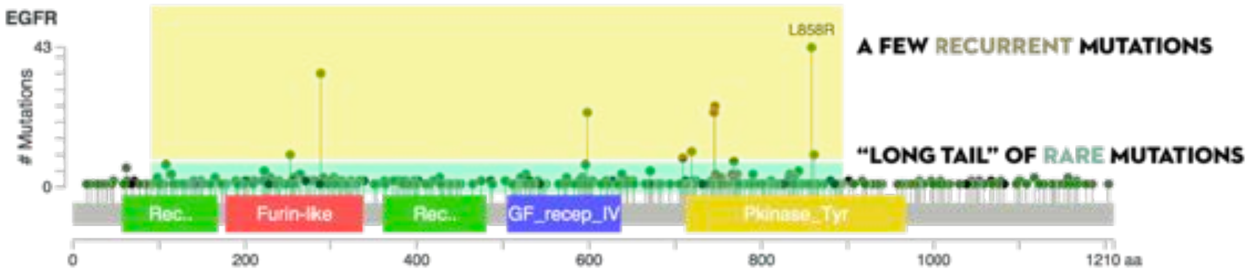
HOW CAN WE DESIGN SPECIFICALLY TARGETED CANCER DRUGS?

HOW CAN WE PREDICT DRUG RESISTANCE AND SUSCEPTIBILITY?

THE LONG TAIL OF CANCER MUTATIONS FRUSTRATES THE PREDICTION OF RESISTANCE



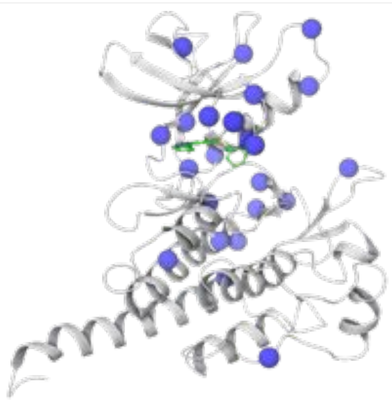
The **vast majority of mutations** are so rare there is unlikely to be clinical or biochemical evidence of whether they confer **drug resistance** or **susceptibility**



STEVEN ALBANESE

Zehir et al. Nature Medicine 23:703, 2017
Hauser, Negron, Albanese, Ray, Steinbrecher, Abel, **Chodera**, Wang. Communications Biology 1:70, 2018

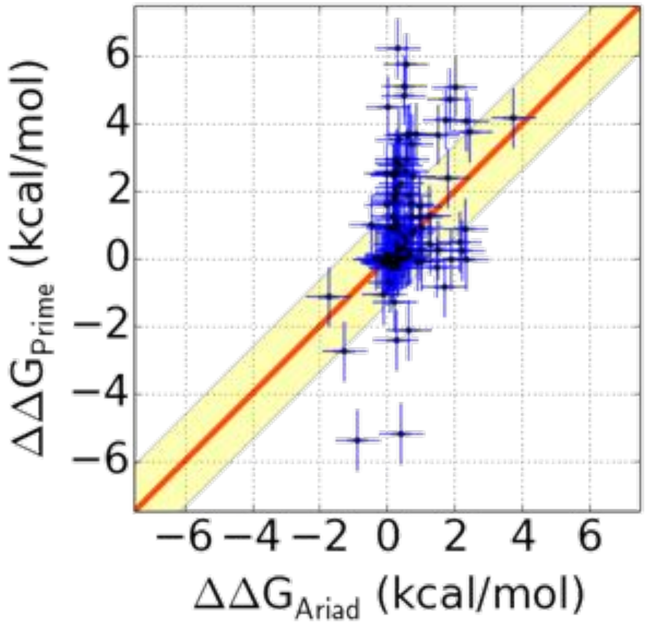
ALCHEMICAL FREE ENERGY CALCULATIONS OUTPERFORM A NULL MODEL



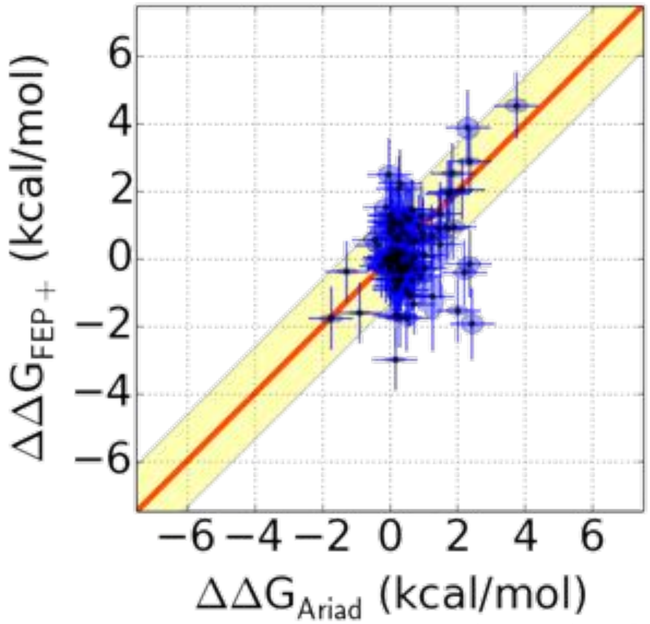
TKI	N_{mut}	R	S
Axitinib	26	0	26
Bosutinib	21	4	17
Dasatinib	21	5	16
Imatinib	21	5	16
Nilotinib	21	4	17
Ponatinib	21	0	21
Subtotal	131	18	113
Erlotinib	7	1	6
Gefitinib	6	0	6
Total	144	19	125

N_{mut} Total number of mutants for which ΔpIC_{50} data was available
Number of **R**esistant, **S**usceptible mutants using 10-fold affinity change threshold

Prime(dock, minimize with MM-GBSA) FEP+ (alchemical free energy)



MUE (kcal/mol)	**1.38** $^{1.58}_{1.21}$	**0.73** $^{0.85}_{0.63}$
RMSE (kcal/mol)	**1.73** $^{1.97}_{1.52}$	**0.92** $^{1.07}_{0.79}$

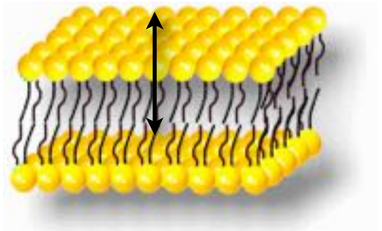


Hauser, Negron, Albanese, Ray, Steinbrecher, Abel, **Chodera**, Wang. Communications Biology 1:70, 2018
https://doi.org/10.1038/s42003-018-0075-x

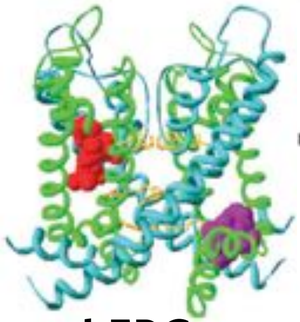
KEVIN HAUSER
SCHRÖDINGER

ALCHEMICAL METHODS COULD ALSO ENABLE THE PREDICTION OF OTHER USEFUL PROPERTIES

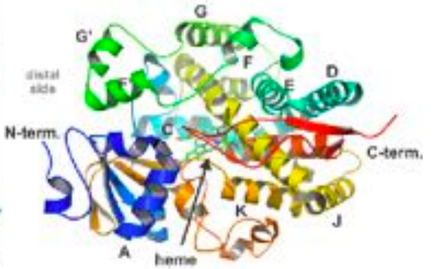
partition coefficients (logP, logD) and permeabilities



structure-enabled ADME/Tox targets



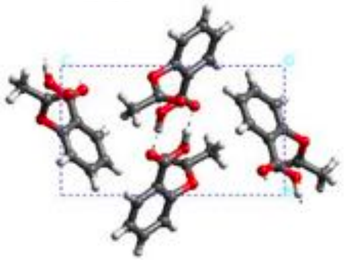
hERG



CYP3A4

porin permeation

crystal polymorphs, solubilities, hygroscopicities, etc.

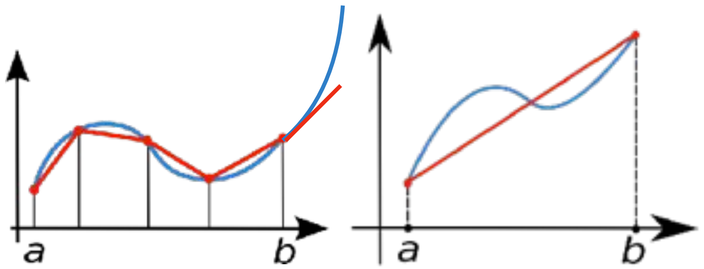


WHAT'S THE DIFFERENCE BETWEEN TI, EXP, BAR, AND MBAR?

THERMODYNAMIC INTEGRATION (TI)

$$\Delta f = \int_{\lambda_1}^{\lambda_2} d\lambda \left\langle \frac{\partial u}{\partial \lambda} \right\rangle_\lambda \approx \frac{\Delta\lambda}{2}\left[\left\langle \frac{\partial u}{\partial \lambda} \right\rangle_{\lambda_1} + \left\langle \frac{\partial u}{\partial \lambda} \right\rangle_{\lambda_2} \right]$$

Jorge et al. *JCTC* **6:**1018, 2010.

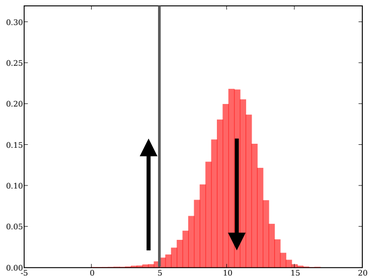


Quadrature error (bias) impossible to quantify; can easily go wrong without noticing.

EXPONENTIAL REWEIGHTING (EXP) / PERTURBATION (FEP)

$$\Delta f = -\ln \left\langle e^{-\Delta u} \right\rangle_{\lambda_1}$$

Zwanzig RW. *JCP* **22:**1420, 1954.
Shirts MR and Pande VS. *JCP* **122:**144107, 2005.

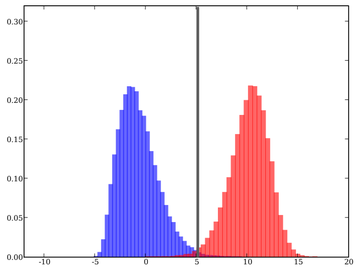


Suffers from large bias and variance

BENNETT ACCEPTANCE RATIO (BAR)

$$\Delta f = \ln \frac{\left\langle \alpha(x)\, e^{-u(x;\lambda_1)} \right\rangle_{\lambda_2}}{\left\langle \alpha(x)\, e^{-u(x;\lambda_2)} \right\rangle_{\lambda_1}}$$

Bennett CH. *J Comput Phys* **22:**245, 1976.
Shirts MR, Bair E, Hooker G, and Pande VS. *PRL* **91:**140601, 2003.

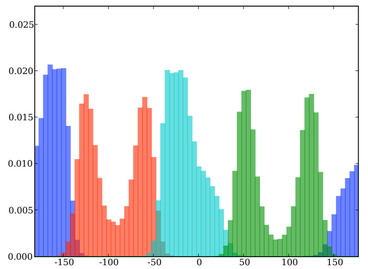


Optimal use of data from two states; minimal variance; bias << variance

MULTISTATE BENNETT ACCEPTANCE RATIO (MBAR)

$$f_i = -\ln \sum_{n=1}^{N} \left[\sum_{k=1}^{K} N_k\, e^{-\Delta u_{ki}(x_n)} \right]^{-1}$$

Shirts MR and Chodera JD. JCP **129:124105**, 2008.



Optimal use of data from many states; minimal variance; bias << variance

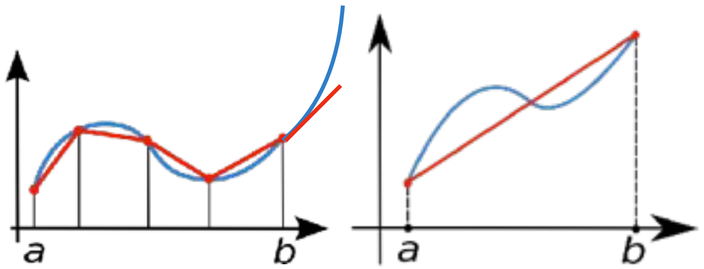
http://www.alchemistry.org/wiki/Free_Energy_Fundamentals

WHAT'S THE DIFFERENCE BETWEEN TI, EXP, BAR, AND MBAR?

THERMODYNAMIC INTEGRATION (TI)

$$\Delta f = \int_{\lambda_1}^{\lambda_2} d\lambda \left\langle \frac{\partial u}{\partial \lambda} \right\rangle_\lambda \approx \frac{\Delta\lambda}{2}\left[\left\langle \frac{\partial u}{\partial \lambda} \right\rangle_{\lambda_1} + \left\langle \frac{\partial u}{\partial \lambda} \right\rangle_{\lambda_2}\right]$$

Jorge et al. *JCTC* **6:**1018, 2010.

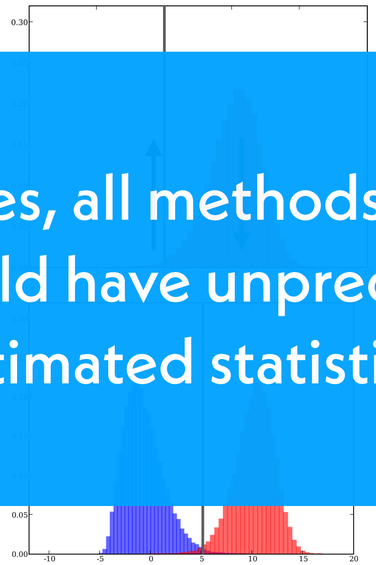


Quadrature error (bias) impossible to quantify; can easily go wrong without noticing.

EXPONENTIAL REWEIGHTING (EXP) / PERTURBATION (FEP)

$$\Delta f = -\ln \left\langle e^{-\Delta u} \right\rangle_{\lambda_1}$$

Zwan
Shirts



Suffers from large bias and variance

For well-optimized alchemical intermediates, all methods give consistent results
For any other case, some estimators could have unpredictably large errors
that won't show up in the estimated statistical error

BEN

$$\Delta f = \ln \frac{\left\langle \alpha(x)\, e^{-u(x;\lambda_1)} \right\rangle_{\lambda_2}}{\left\langle \alpha(x)\, e^{-u(x;\lambda_2)} \right\rangle_{\lambda_1}}$$

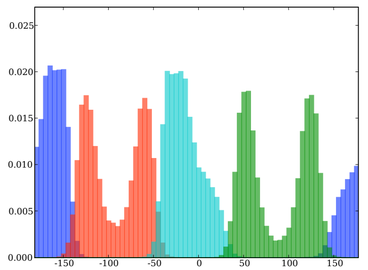
Benne
Shirts MR, Bair E, Hooker G, and Pande VS. *PRL* **91:**140601, 2003.

Optimal use of data from two states; minimal variance; bias << variance

MULTISTATE BENNETT ACCEPTANCE RATIO (MBAR)

$$f_i = -\ln \sum_{n=1}^{N} \left[\sum_{k=1}^{K} N_k\, e^{-\Delta u_{ki}(x_n)} \right]^{-1}$$

Shirts MR and Chodera JD. JCP **129:124105**, 2008.



Optimal use of data from many states; minimal variance; bias << variance

http://www.alchemistry.org/wiki/Free_Energy_Fundamentals

WHAT DO I NEED TO KNOW ABOUT SAMPLING?

Independent simulations

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

simple but dangerous

Hamiltonian replica exchange ★

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

reliable but complex and costly

Single-replica methods

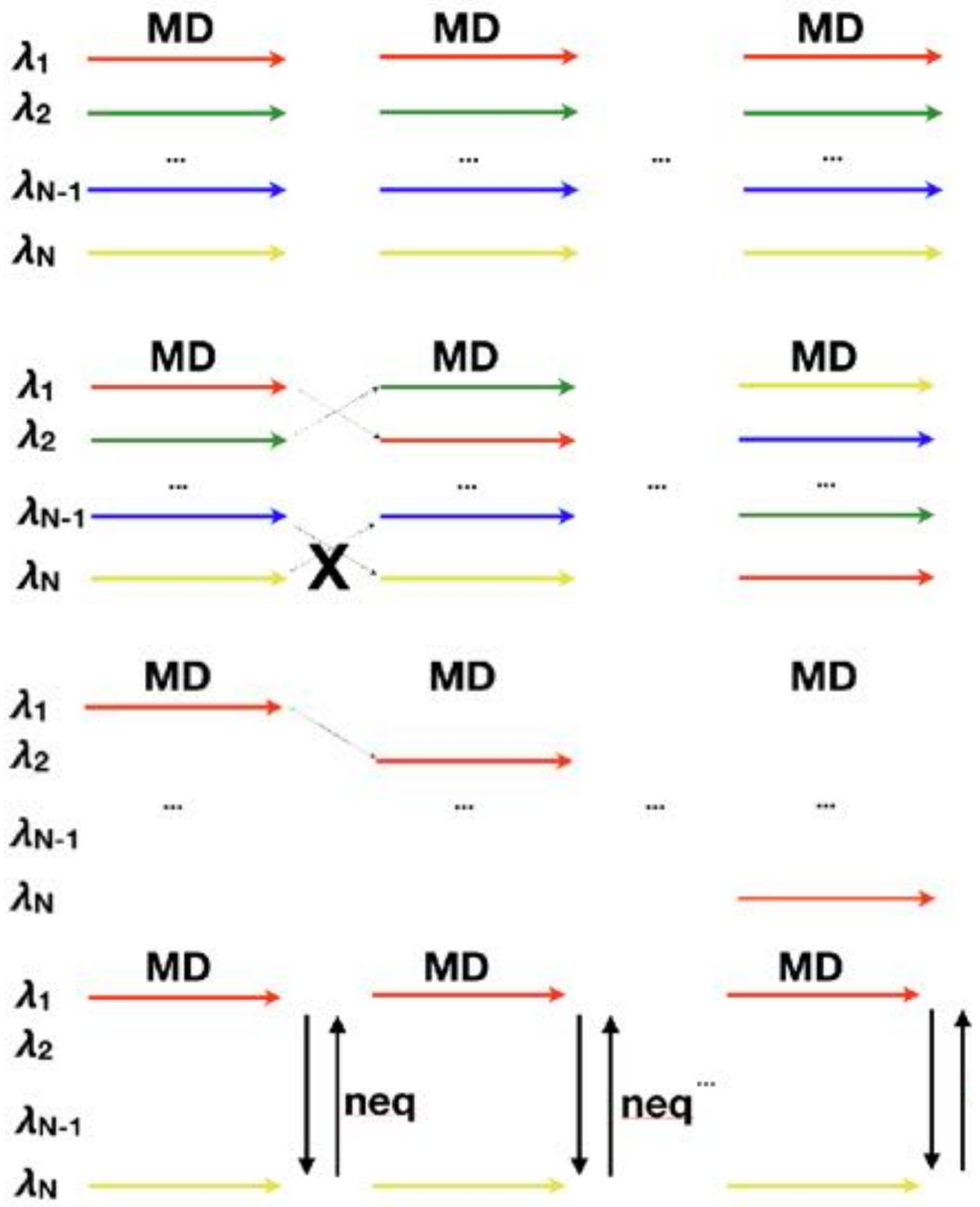
For certainly problems, can converge extremely quickly in a fraction of computer effort; tricky to make reliable

promising but relatively immature

Nonequilibrium methods

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

promising but relatively immature



★ **current best practice**

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

pmx / gromacs
Aldeghi, Heifetz, Bodkin, Knapp, Biggin 2016
https://doi.org/10.1039/C5SC02678D

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

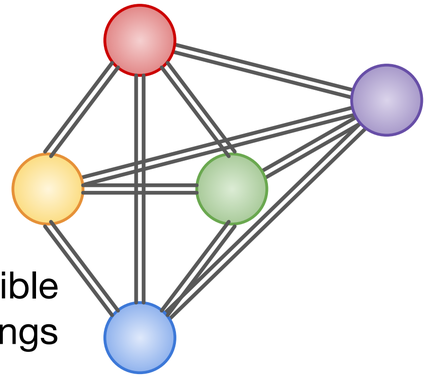
also **OpenMM** (via **openmmtools**)

Hongzhi, Fayer, Wang 2006
https://doi.org/10.1063/1.2424700

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

THERE ARE MULTIPLE WAYS TO CONSTRUCT ALCHEMICAL MAPPINGS FOR FREE ENERGY CALCULATIONS

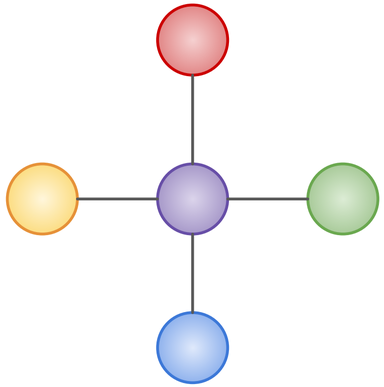
a) complete



multiple possible atom mappings

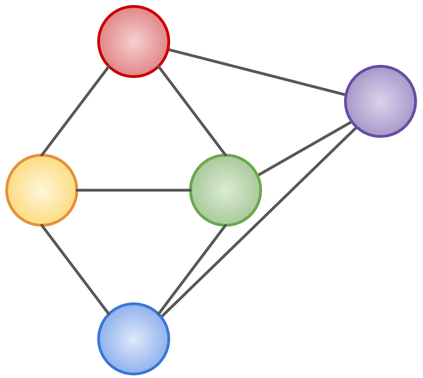
scales poorly with number of ligands

b) minimum spanning tree



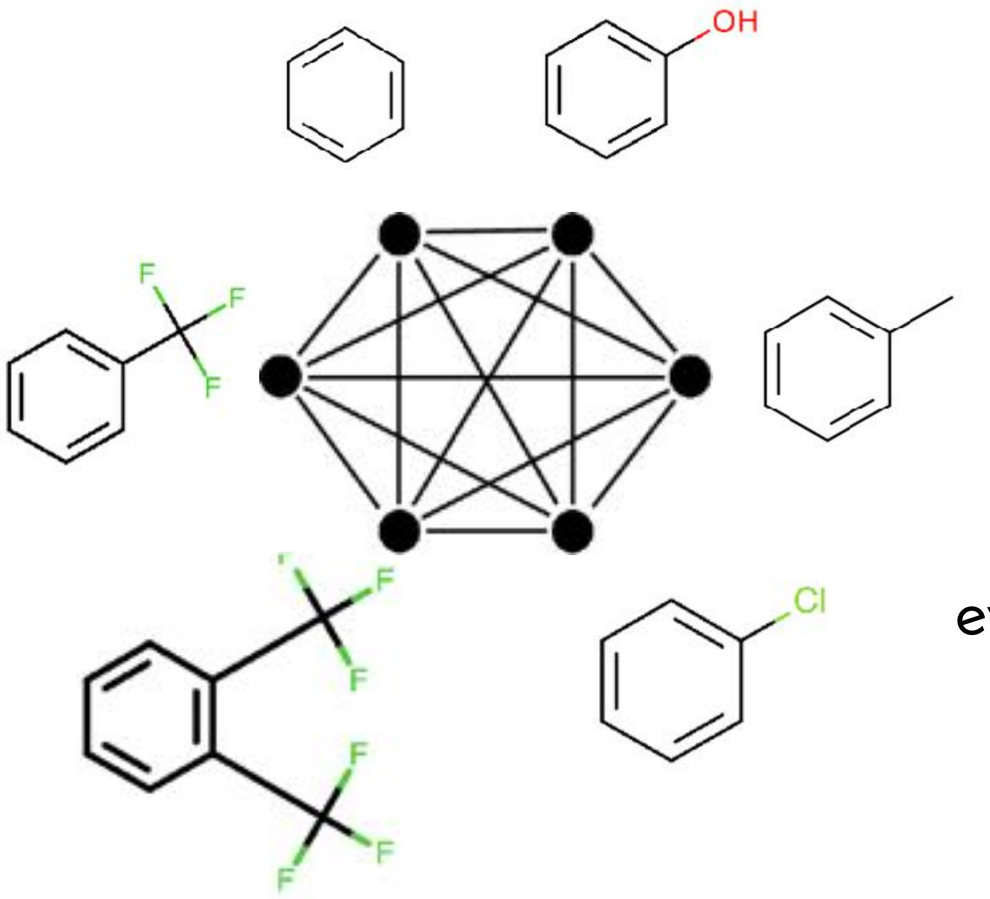
one poor edge can wreck everything

c) cycle closure

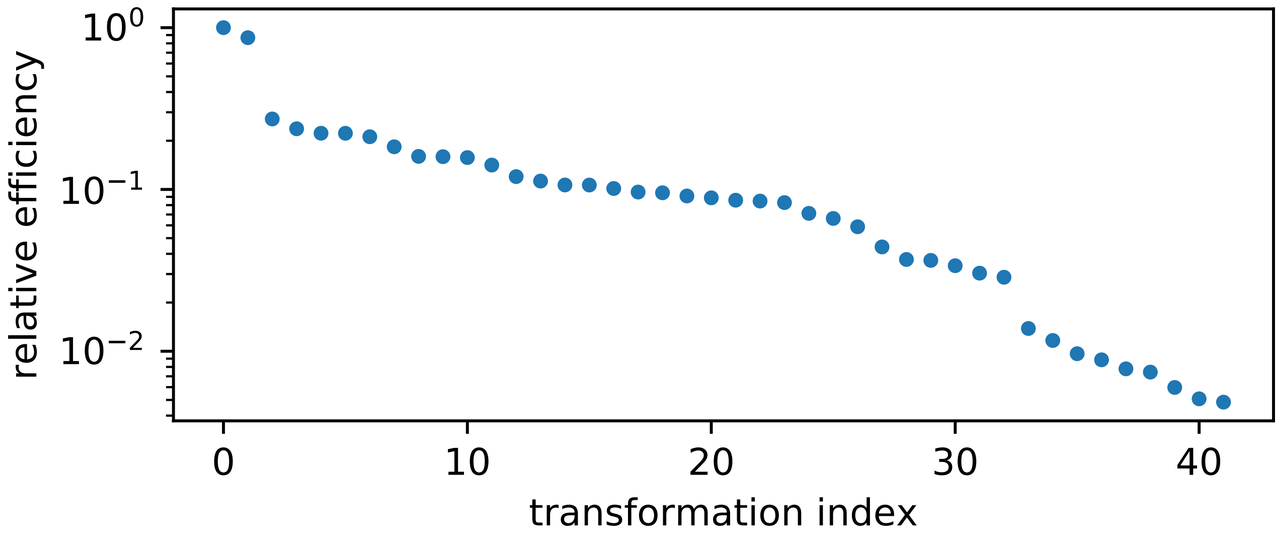


which edges? how much cycle redundancy?

EVEN SIMPLE TRANSFORMATIONS HAVE WIDELY DIFFERING EFFICIENCIES



FreeSolv substituted benzenes in solvent
replica-exchange with dense (50) alchemical states

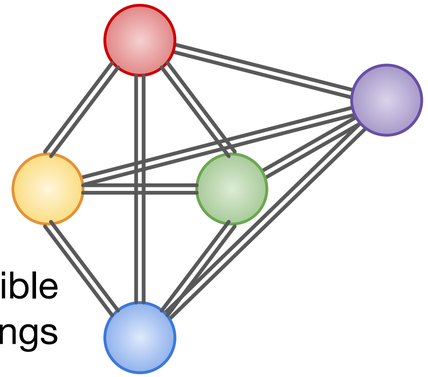


even if we just consider one possible mapping/transformation, some transformations are more than 100x less efficient!

HANNAH BRUCE MACDONALD

THERE ARE MULTIPLE WAYS TO CONSTRUCT ALCHEMICAL MAPPINGS FOR FREE ENERGY CALCULATIONS

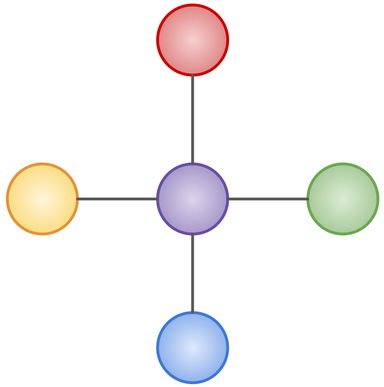
a) complete



multiple possible atom mappings

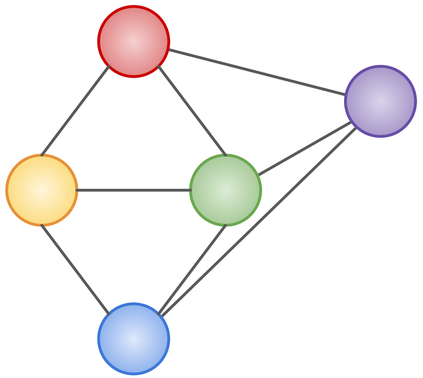
scales poorly with number of ligands

b) minimum spanning tree



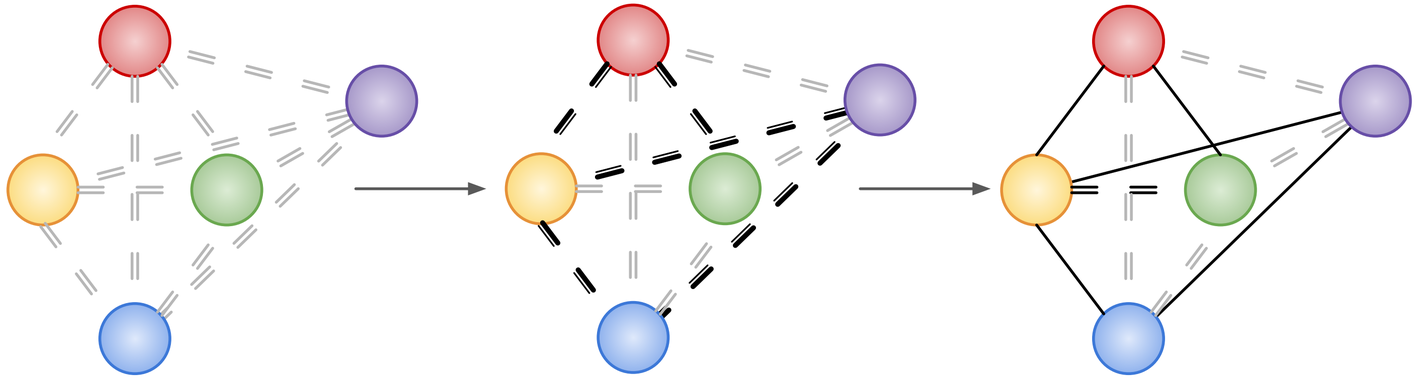
one poor edge can wreck everything

c) cycle closure



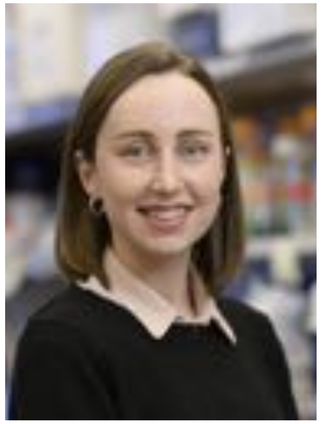
which edges? how much cycle redundancy?

d) adaptive

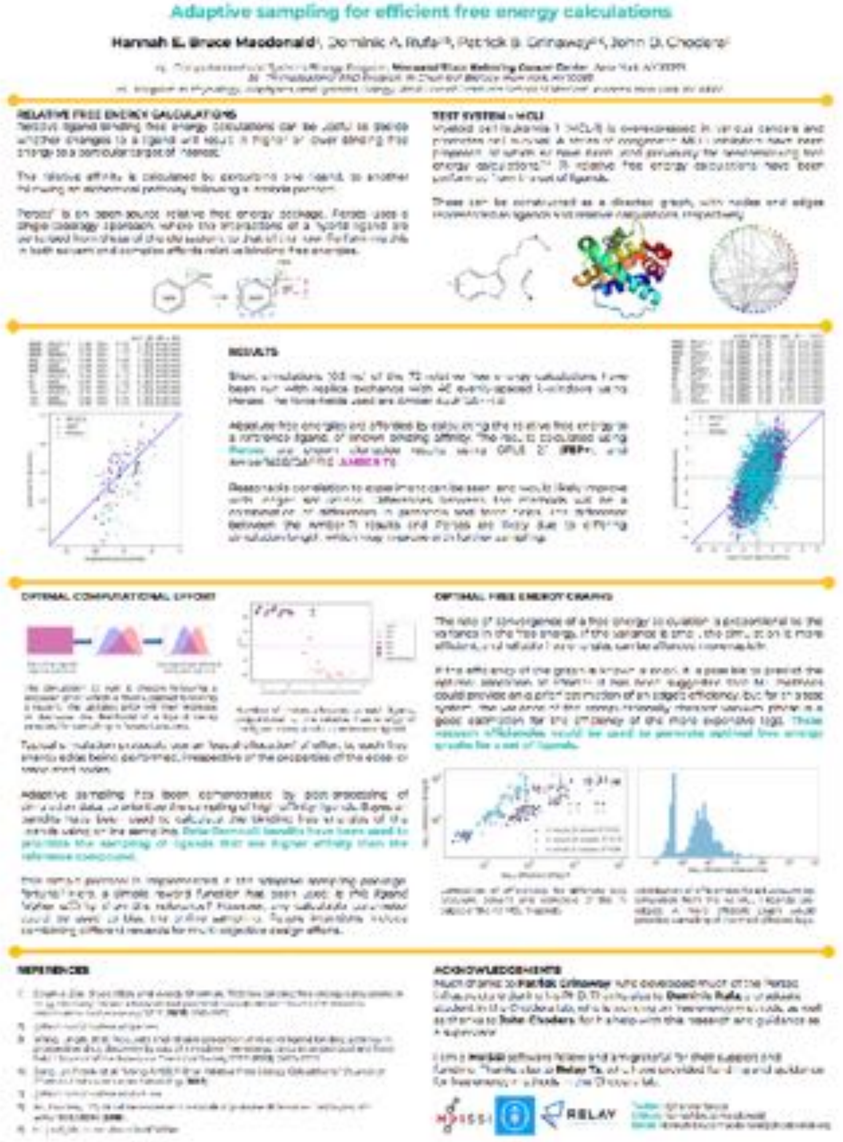


effort is allocated where it is needed to achieve stated goals (which may be different for different projects)

IF YOU'D LIKE TO LEARN MORE ABOUT ADAPTIVE RELATIVEE FREE ENERGY CALCULATIONS...



Hannah Bruce Macdonald
Postdoctoral Research Fellow // MSKCC
MolSSI Software Fellow



Adaptive sampling for efficient free energy calculations

Hannah E. Bruce Macdonald, Dominic A. Rufa, Patrick B. Grinaway, John D. Chodera

RELATIVE FREE ENERGY CALCULATIONS

TEST SYSTEM - MCL1

RESULTS

OPTIMAL COMPUTATIONAL EFFORT

OPTIMAL FREE ENERGY CHUNKS

REFERENCES

ACKNOWLEDGEMENTS

Mon/Tue poster #27

THERE ARE STILL MAJOR LIMITATIONS TO THE DOMAIN OF APPLICABILITY OF FREE ENERGY CALCULATIONS

Multiple high-quality crystal structures of target

Congeneric series of ligands with
all ligands binding in same pose

Only one dominant protonation state
unchanged throughout binding process

No ligand or sidechain tautomerism

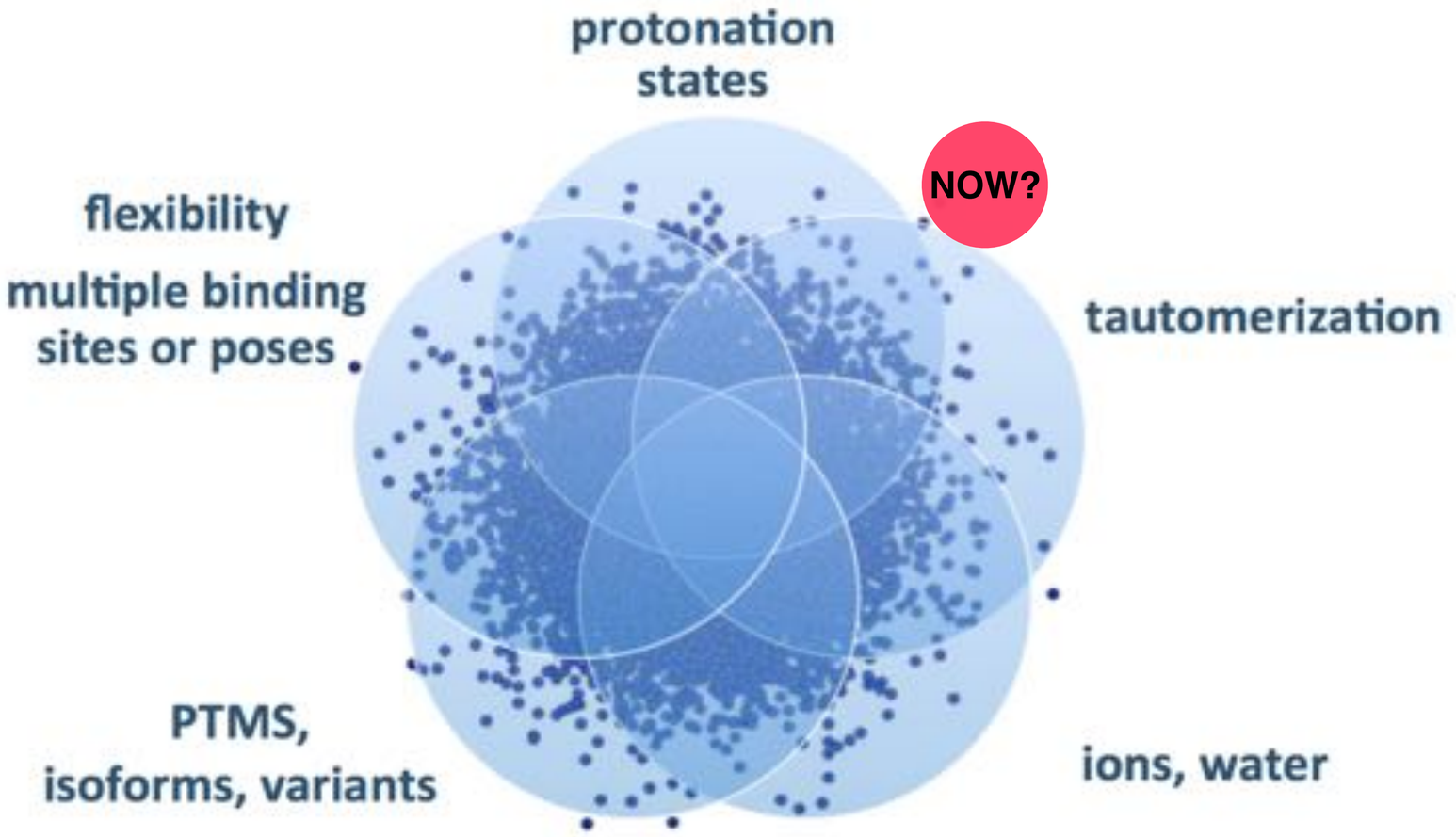
One well-specified, well-resolved isoform/species

No complex cosolvents, binding partners,
slow binding site desolvation events

No exotic chemistries

No metals or prosthetic groups

No membranes?



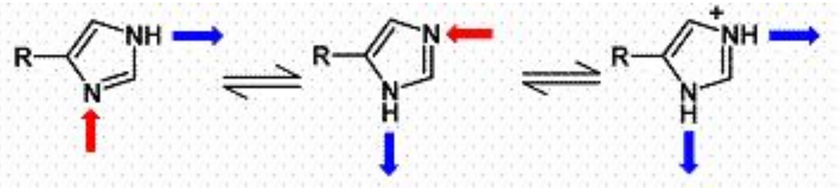
TARGETS
CHALLENGES

FREE ENERGY CALCULATIONS FAIL FOR THREE MAIN REASONS

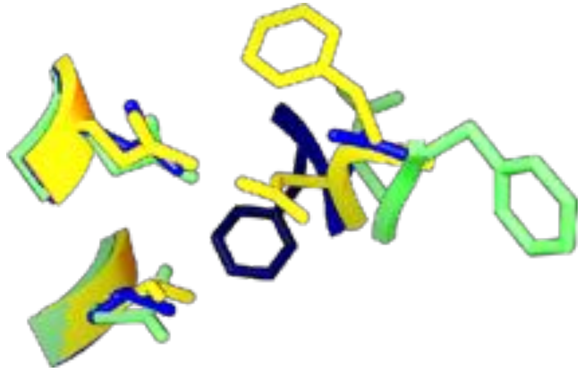
1. The forcefield does a poor job of modeling the physics of our system

$$V(\mathbf{q}) = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

2. We're missing some essential chemical in our simulations
(e.g. protonation states, tautomers, covalent association)



3. We haven't sampled all of the relevant conformations

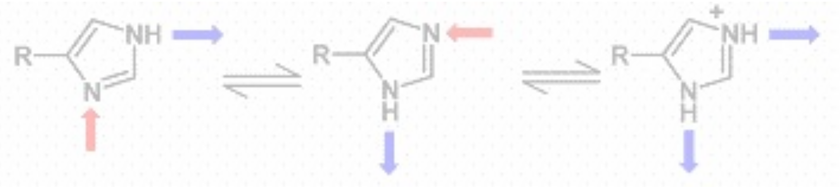


FREE ENERGY CALCULATIONS FAIL FOR THREE MAIN REASONS

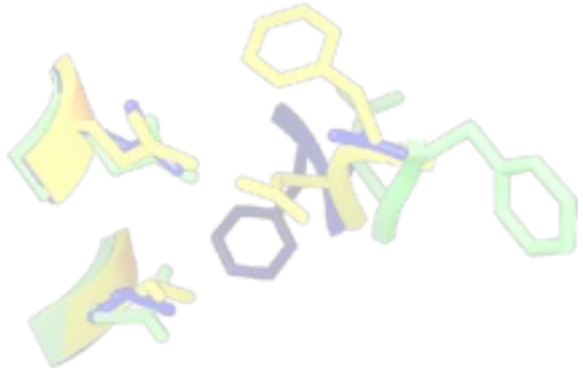
1. The forcefield does a poor job of modeling the physics of our system

$$V(\mathbf{q}) = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

2. We're missing some essential chemical in our simulations (e.g. protonation states, tautomers, covalent association)



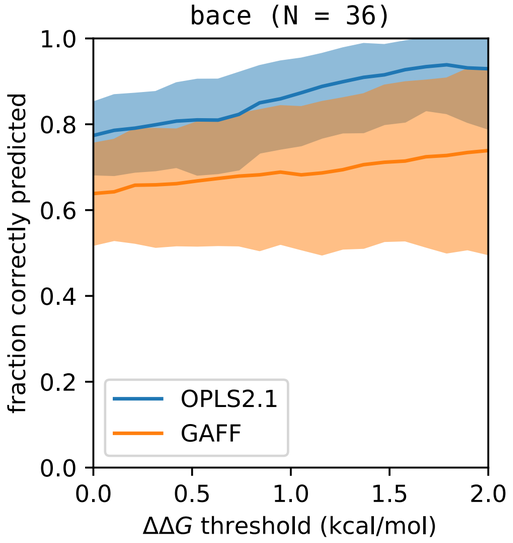
3. We haven't sampled all of the relevant conformations

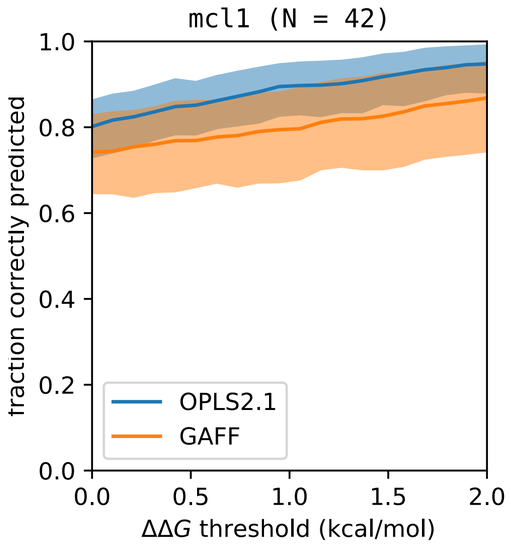


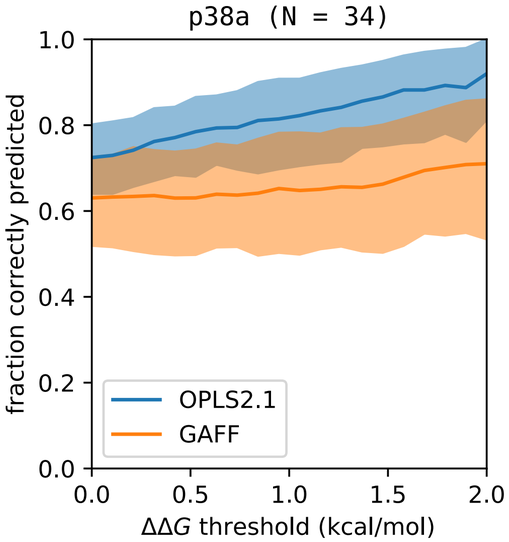
PUBLIC FORCE FIELDS NEED TO CATCH UP

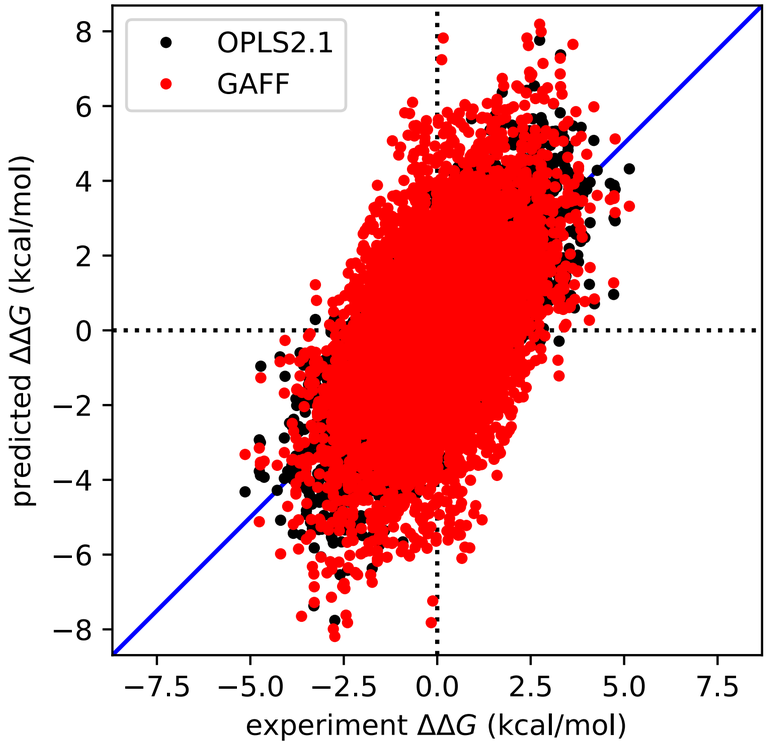
FRACTION OF TIME
SIGN OF TRANSFORMATION IS CORRECTLY PREDICTED
AMBER14SB/GAFF1.8 VS OPLS2.1 (SCHRÖDINGER JACS PAPER)

```
                    all within-target pairs ΔΔG (N = 5620)
RMSE: OPLS2.1    1.37 [95%:  1.34,  1.39] kcal/mol
RMSE: GAFF       1.97 [95%:  1.93,  2.01] kcal/mol
MUE : OPLS2.1    1.09 [95%:  1.06,  1.11] kcal/mol
MUE : GAFF       1.55 [95%:  1.51,  1.58] kcal/mol
R2  : OPLS2.1    0.10 [95%:  0.06,  0.15] kcal/mol
R2  : GAFF      -0.87 [95%: -0.98, -0.76] kcal/mol
rho : OPLS2.1    0.73 [95%:  0.72,  0.74] kcal/mol
rho : GAFF       0.53 [95%:  0.51,  0.55] kcal/mol
```









Song, Lee, Zhu, York, Merz 2019
https://doi.org/10.1021/acs.jcim.9b00105
Wang, Wu, Deng, Kim, ... Abel 2015
https://doi.org/10.1021/ja512751q

https://github.com/jchodera/jacs-dataset-analysis

HOW ARE FORCEFIELDS MADE?

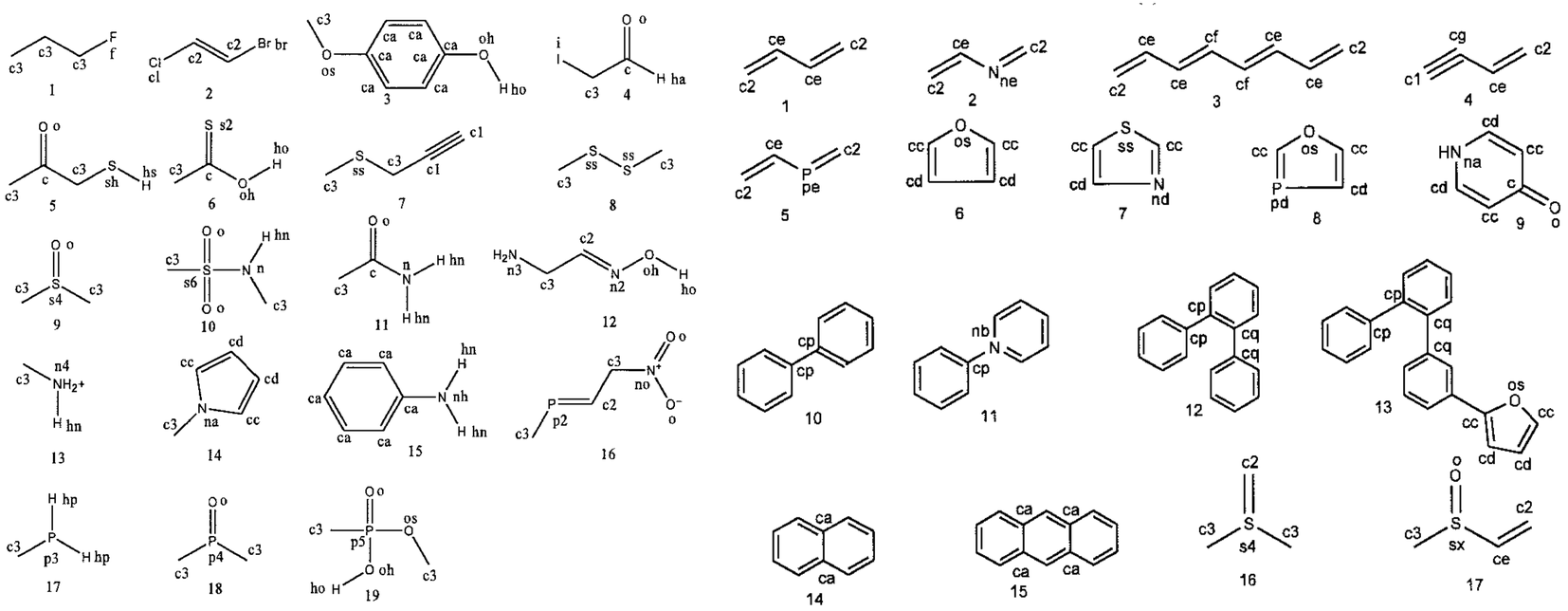
experimental data
quantum chemistry
keen chemical intuition

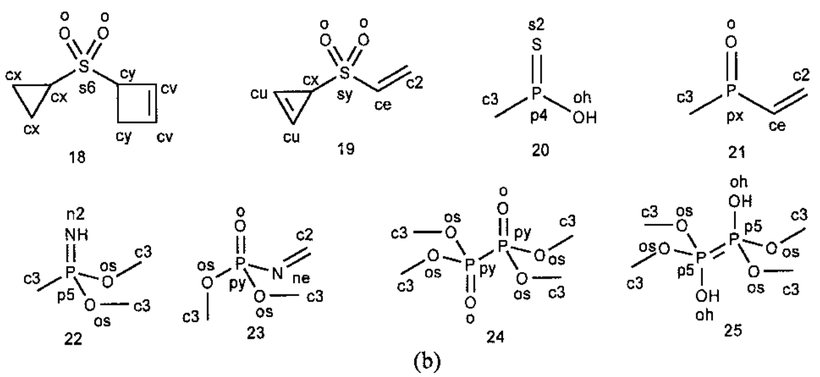
heroic effort by graduate students and postdocs

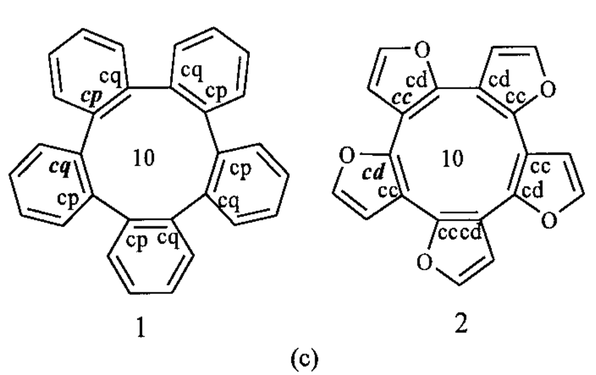
a parameter set we desperately hope someone actually uses

AS DRUG DISCOVERY EXPLORES NEW PARTS OF CHEMICAL SPACE, HOW CAN FORCEFIELDS KEEP UP?

The Generalized Amber Forcefield (GAFF) was parameterized with this chemical universe:







- GAFF 1 was finished in 1999
- Extension to new chemical space is nontrivial
- Parameter fitting code was never released
- Atom types cause numerous complications

Wang J, Wolf RM, Caldwell JW, Kollman PA, and Case DA. J Comput Chem 25:1157, 2004.

THE OPEN FORCE FIELD INITIATIVE
HOW IS IT OPEN?

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

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

Open source infrastructure: for improving force fields with in-house data

Open science: everything we do is free, permissively licensed, and online

http://openforcefield.org

THE OPEN FORCE FIELD CONSORTIUM:
A PARTNERSHIP FOR ADVANCING BIOMOLECULAR MODELING

INDUSTRY PARTNERS

BASF
Bayer
BMS
Boehringer Ingelheim
GSK
Merck KGaA
Pfizer
QuLab
Roche
Vertex

SOFTWARE SCIENTISTS

MOLECULAR SOFTWARE SCIENCES INSTITUTE

ACADEMIC RESEARCHERS

CHRISTOPHER BAYLY
OPENEYE SCIENTIFIC

JOHN CHODERA
SLOAN KETTERING INSTITUTE

MICHAEL GILSON
UNIVERSITY OF CALIFORNIA, SAN DIEGO

DAVID MOBLEY
UNIVERSITY OF CALIFORNIA, IRVINE

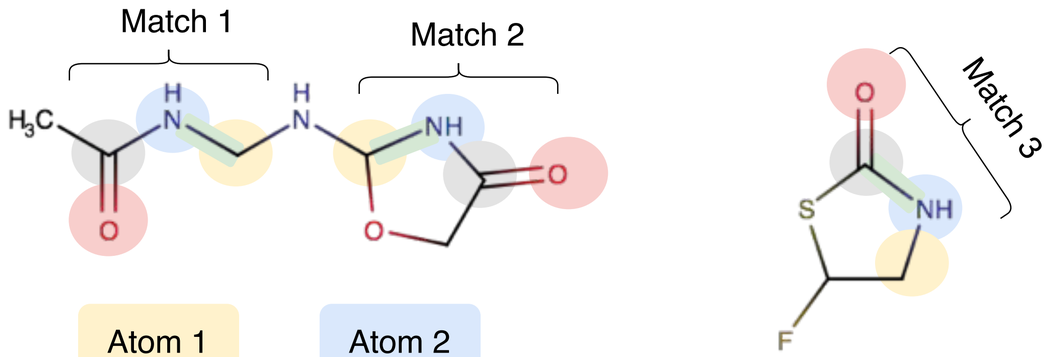
MICHAEL SHIRTS
UNIVERSITY OF COLORADO, BOULDER

LEE-PING WANG
UNIVERSITY OF CALIFORNIA, DAVIS

http://openforcefield.org/consortium

24+ contributing researchers

THE SMIRKS NATIVE OPEN FORCE FIELD (SMIRNOFF) FORMAT AVOIDS THE COMPLEXITY OF ATOM TYPING



MATCH BONDS DIRECTLY:

Atom 1 Atom 2

[#6X4:1]-[#7X3:2]-[#6X3]=[#8X1+0]

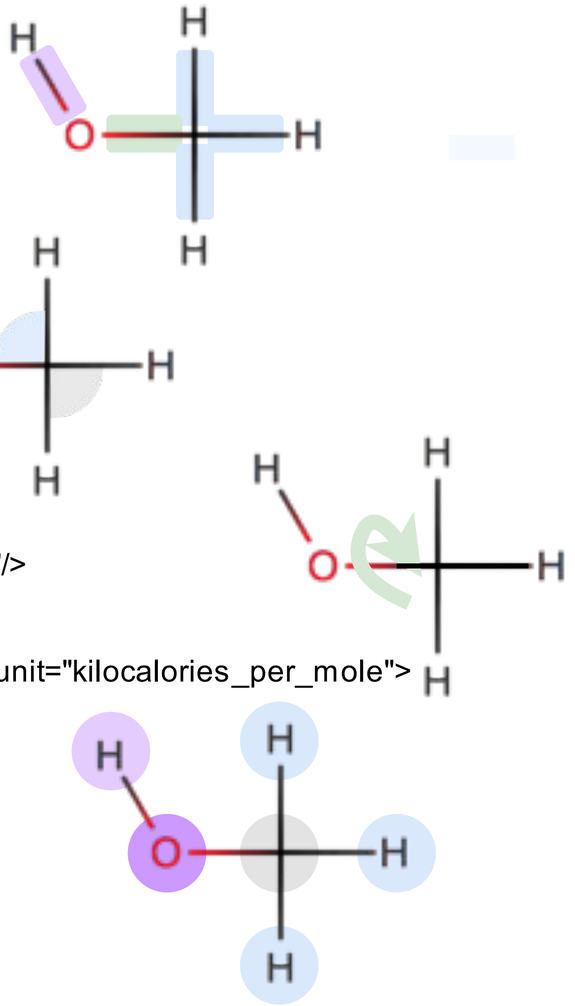
Labeled bond | Alpha carbon | Beta oxygen

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

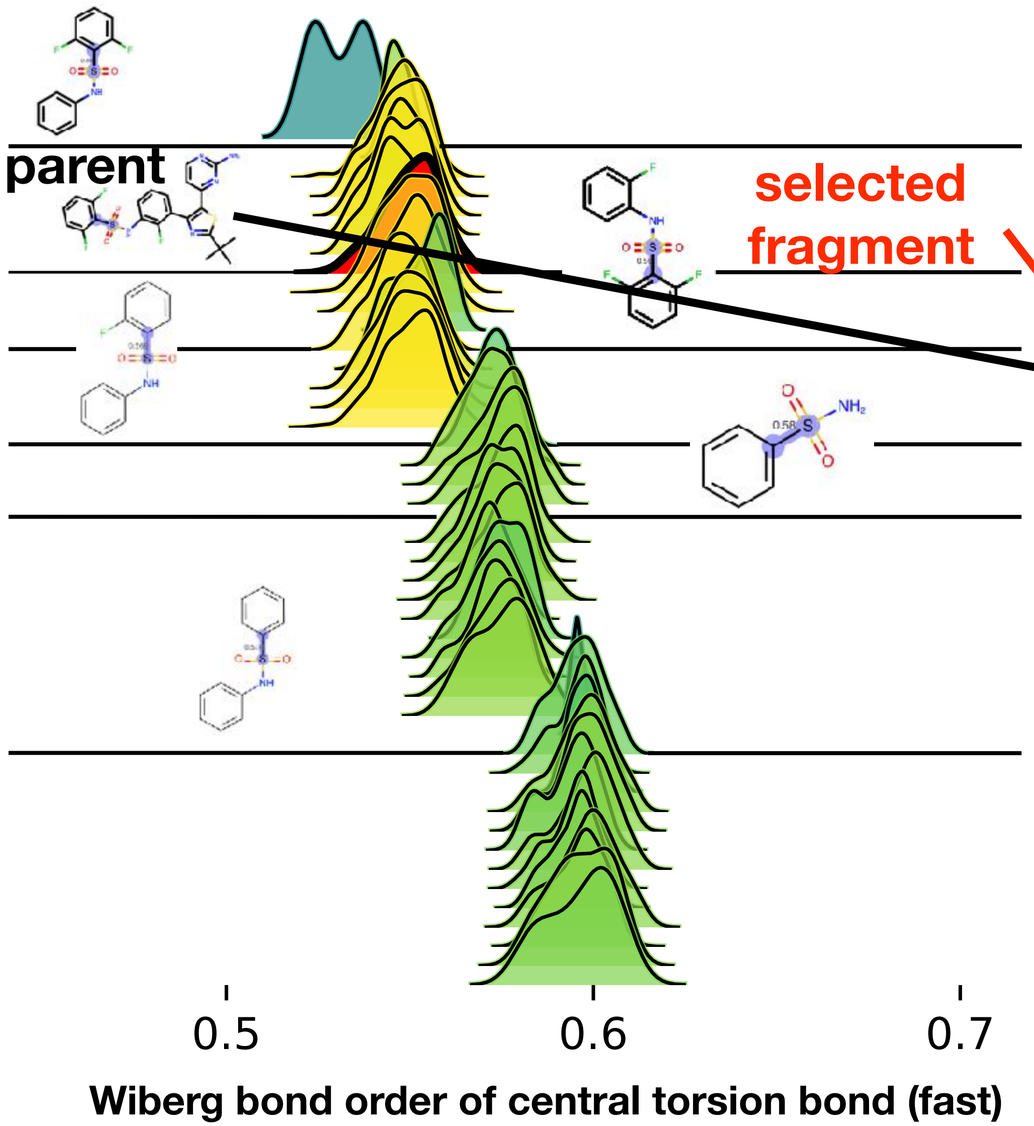
https://github.com/open-forcefield-group/openforcefield/blob/master/The-SMIRNOFF-force-field-format.md

SMIRNOFF VASTLY SIMPLIFIES THE CONSTRUCTION OF BIOMOLECULAR FORCE FIELDS

```
<?xml version="1.0"?>
<SMIRNOFF>

<HarmonicBondForce length_unit="angstroms" k_unit="kilocalories_per_mole/angstrom**2">
  <Bond smirks="[#6X4:1]-[#1:2]" length="1.090" k="680.0"/>
  <Bond smirks="[#6X4:1]-[#8&amp;X2&amp;H1:2]" length="1.410" k="640.0"/>
  <Bond smirks="[#8X2:1]-[#1:2]" length="0.960" k="1106.0"/>
</HarmonicBondForce>

<HarmonicAngleForce angle_unit="degrees" k_unit="kilocalories_per_mole/radian**2">
  <Angle smirks="[a,A:1]-[#6X4:2]-[a,A:3]" angle="109.50" k="100.0"/>
  <Angle smirks="[#1:1]-[#6X4:2]-[#1:3]" angle="109.50" k="70.0"/>
  <Angle smirks="[#6X4:1]-[#8X2:2]-[#1:3]" angle="108.50" k="110.0"/>
</HarmonicAngleForce>

<PeriodicTorsionForce phase_unit="degrees" k_unit="kilocalories_per_mole">
  <Proper smirks="[a,A:1]-[#6X4:2]-[#8X2:3]-[#1:4]" idivf1="3" periodicity1="3" phase1="0.0" k1="0.50"/>
</PeriodicTorsionForce>

<NonbondedForce coulomb14scale="0.833333" lj14scale="0.5" sigma_unit="angstroms" epsilon_unit="kilocalories_per_mole">
  <Atom smirks="[#1:1]" rmin_half="1.4870" epsilon="0.0157"/>
  <Atom smirks="[$([#1]-[#6]-[#7,#8,#9,#16,#17,#35]):1]" rmin_half="1.3870" epsilon="0.0157"/>
  <Atom smirks="[#1$(*-[#8]):1]" rmin_half="0.0000" epsilon="0.0000"/>
  <Atom smirks="[#6:1]" rmin_half="1.9080" epsilon="0.1094"/>
  <Atom smirks="[#8:1]" rmin_half="1.6837" epsilon="0.1700"/>
  <Atom smirks="[#8X2+0$(*-[#1]):1]" rmin_half="1.7210" epsilon="0.2104"/>
</NonbondedForce>

</SMIRNOFF>
```

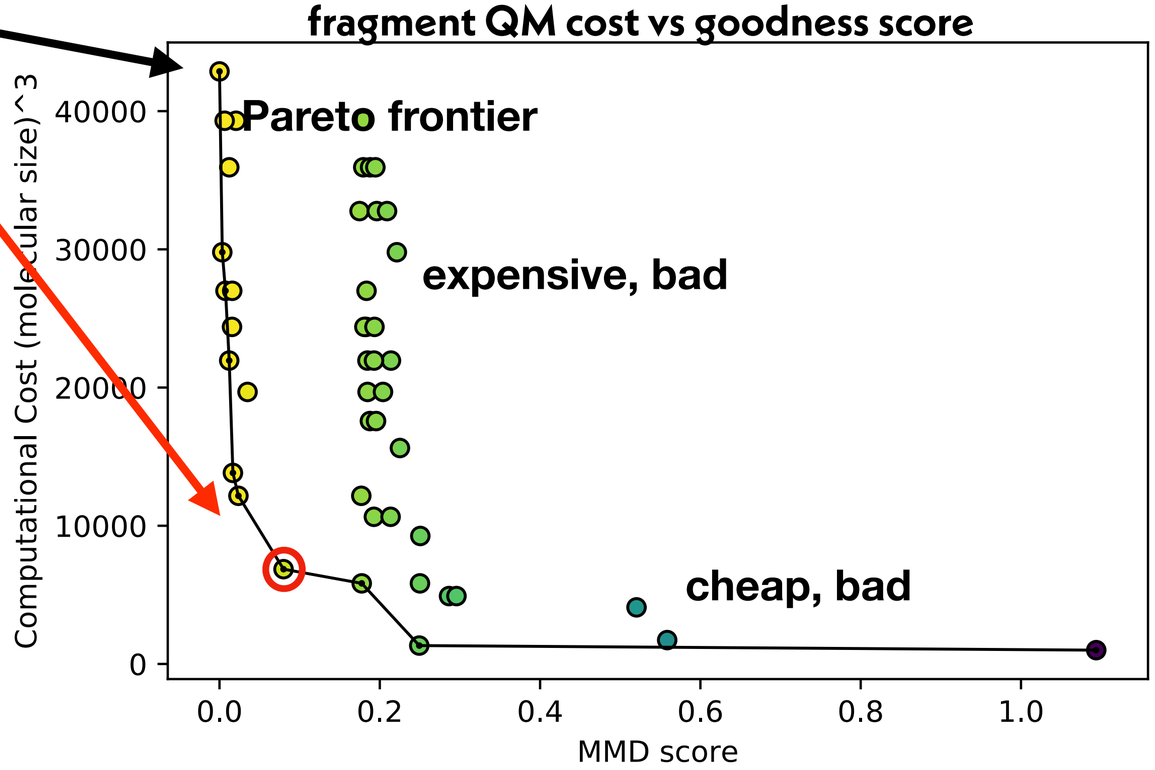


AUTOMATED FRAGMENTATION REDUCES QM TORSION DRIVE COSTS WHILE PRESERVING CHEMICAL ENVIRONMENT

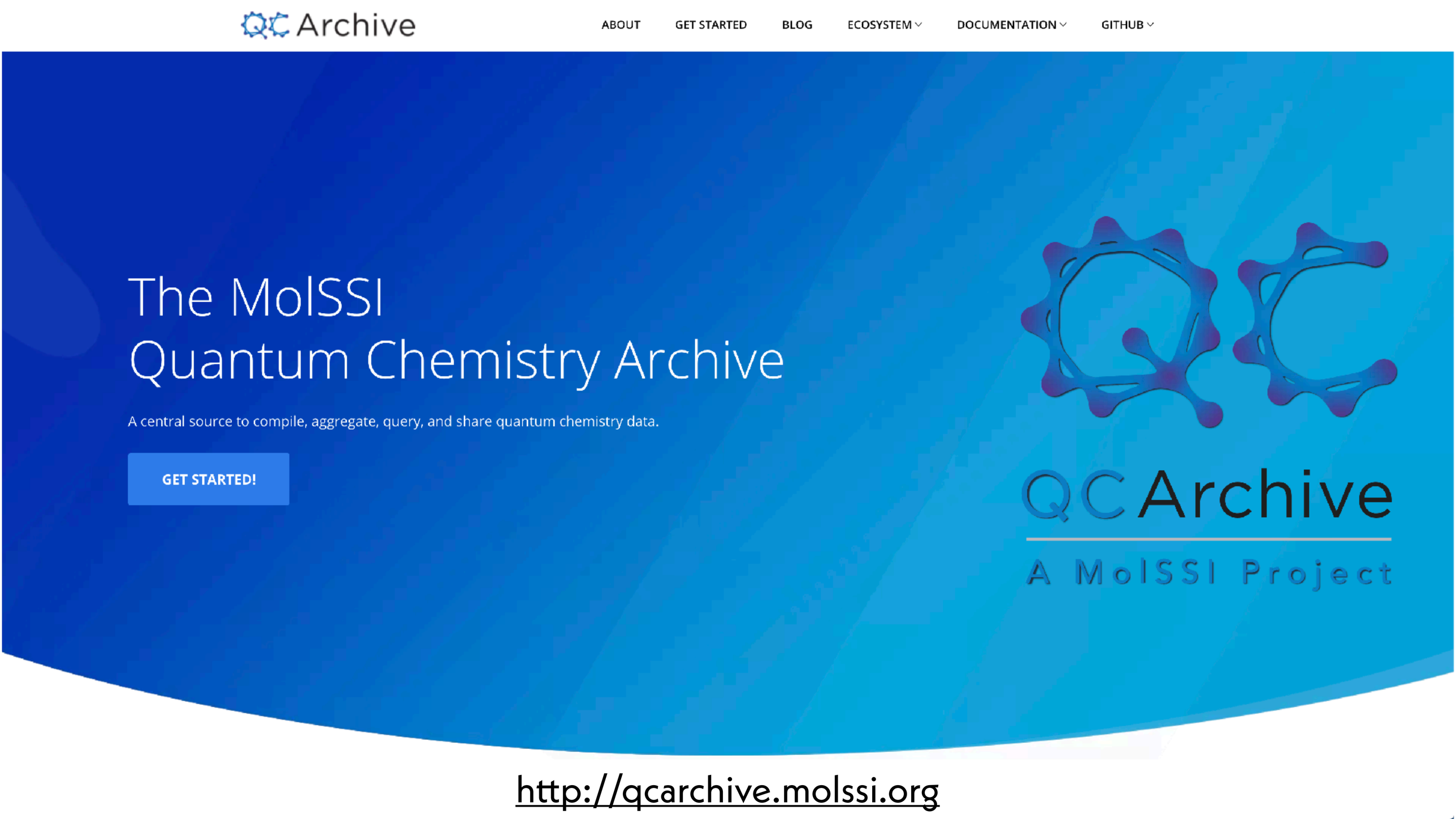


CHAYA STERN

Graduating PhD student
(and cheminformatics wizard)



https://zenodo.org/record/3238643



http://qcarchive.molssi.org

CHECK OUT THE OPEN FORCE FIELD POSTERS

Jeffrey Wagner
Open Force Field Consortium
Software Scientist
the Open Force Field toolkit

Mon/Tue #66

Simon Boothroyd
XtalPi/Open Force Field Consortium
Distinguished Postdoctoral Fellow
the Open Force Field toolkit

Mon/Tue #26

Christina Schindler
Merck KGaA (Consortium Partner)
the Open Force Field Consortium

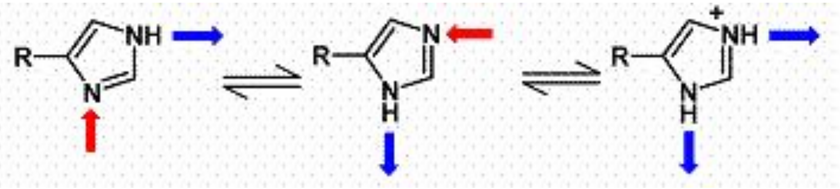
Wed/Thu

FREE ENERGY CALCULATIONS FAIL FOR THREE MAIN REASONS

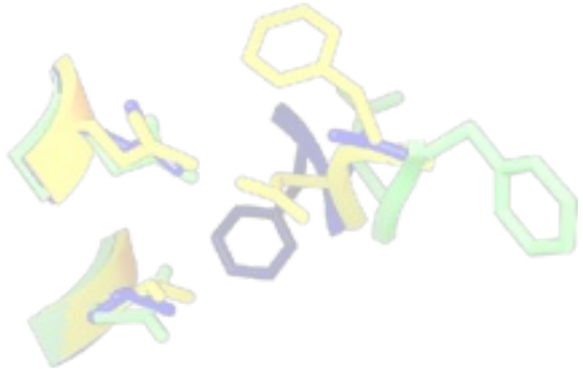
1. The forcefield does a poor job of modeling the physics of our system

$$V(\mathbf{q}) = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

2. We're missing some essential chemical in our simulations (e.g. protonation states, tautomers, covalent association)

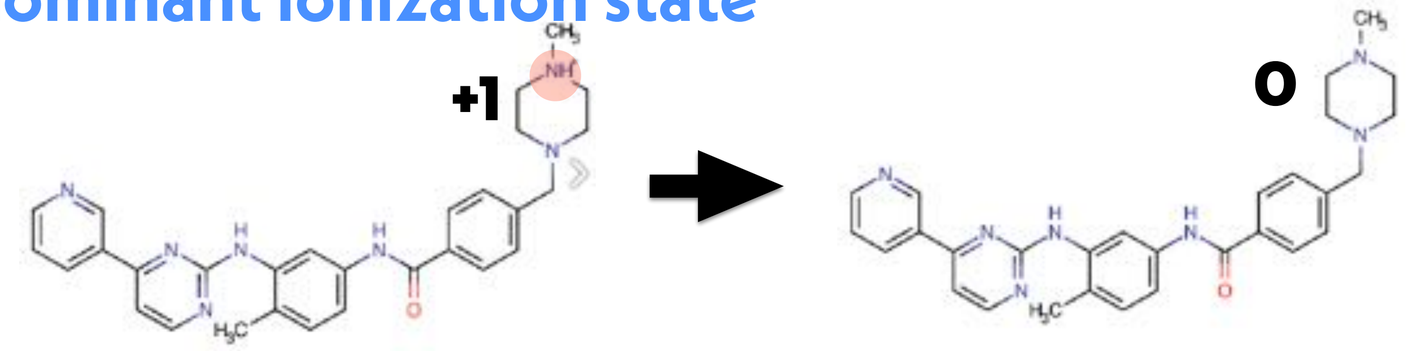


3. We haven't sampled all of the relevant conformations

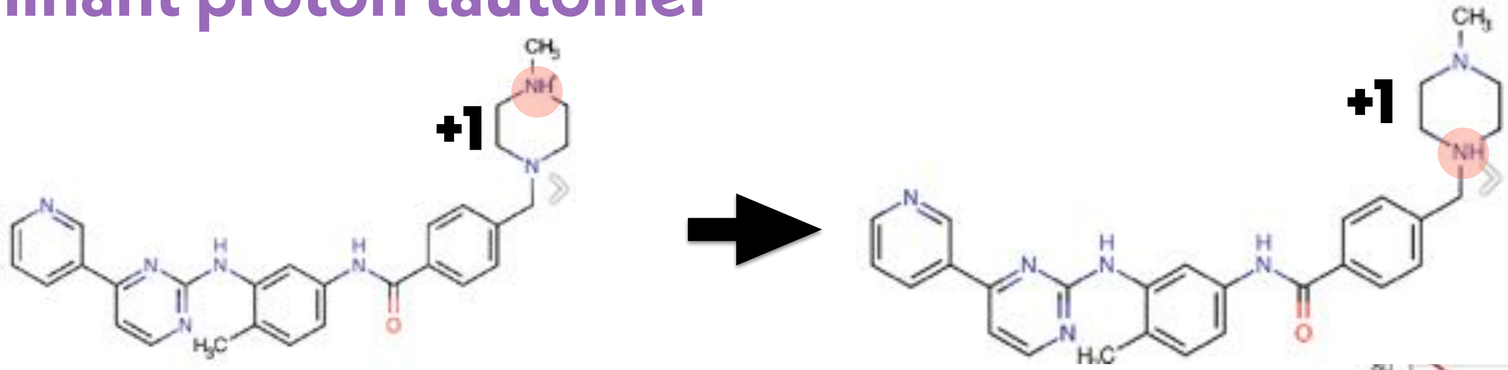


PROTONATION STATE EFFECTS CAN CAUSE LARGE MODELING ERRORS

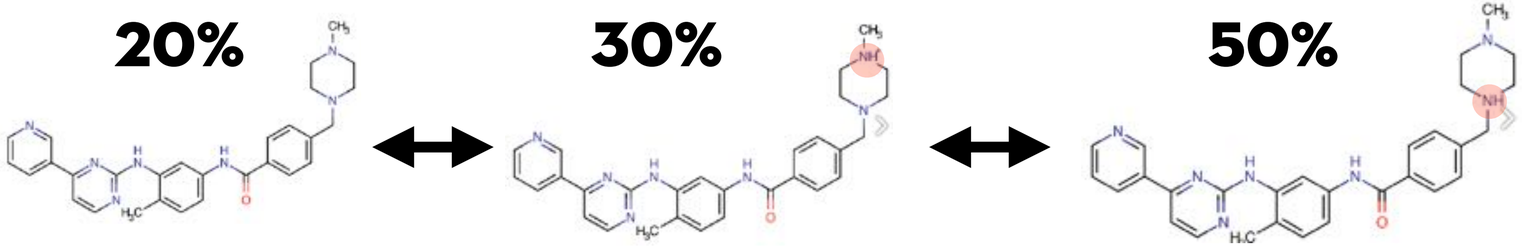
Change in dominant ionization state

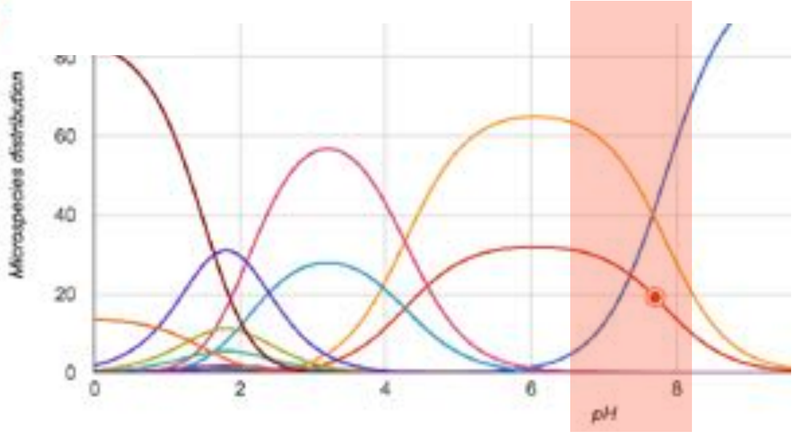


Shift in dominant proton tautomer

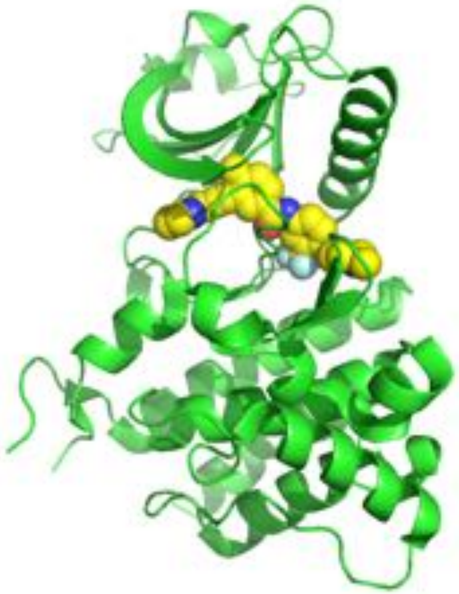


Population of mixture of protonation states



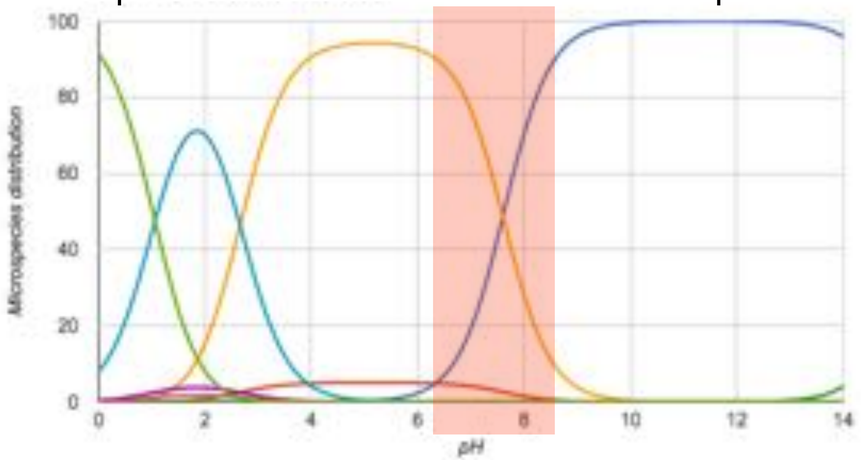


pH-SENSITIVITY OF KINASE:INHIBITOR BINDING SUGGESTS A ROLE FOR PROTONATION STATE EFFECTS

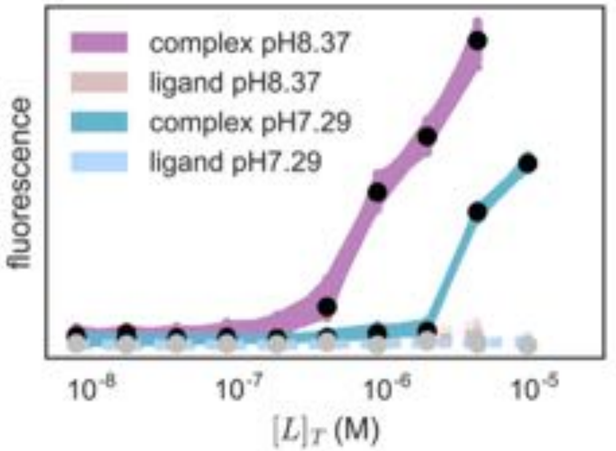


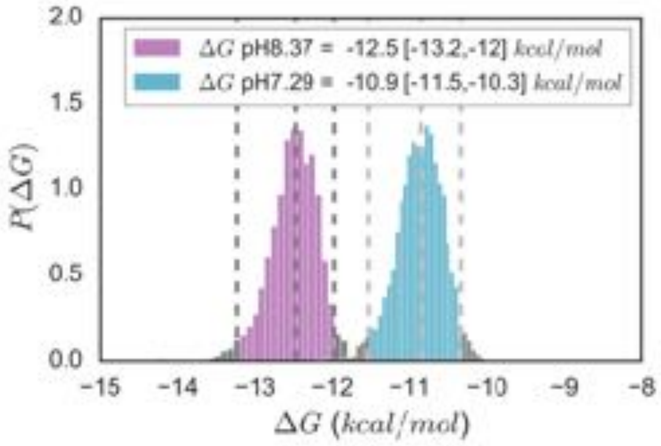
ponatinib:DDR1
(pdbid:3ZOS)

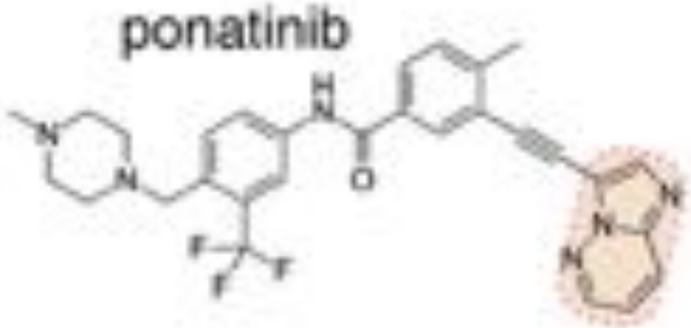
ponatinib populates a mixture
of protonation states in solution at pH 7.4

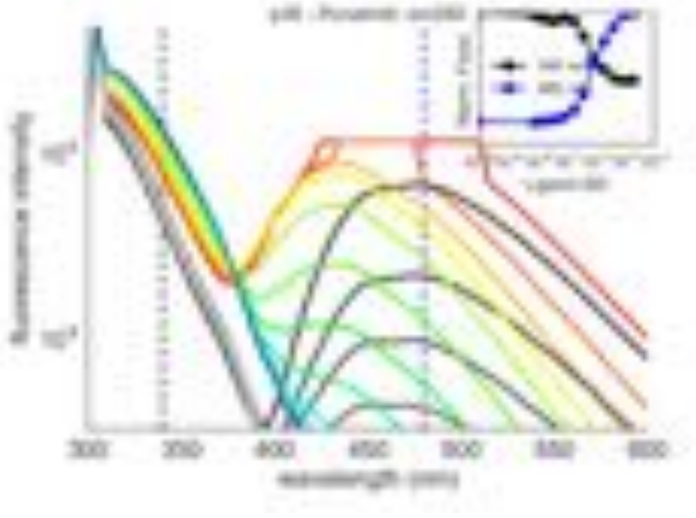


ΔpH of 1.1 ➡ ΔΔG of 1.6 kcal/mol



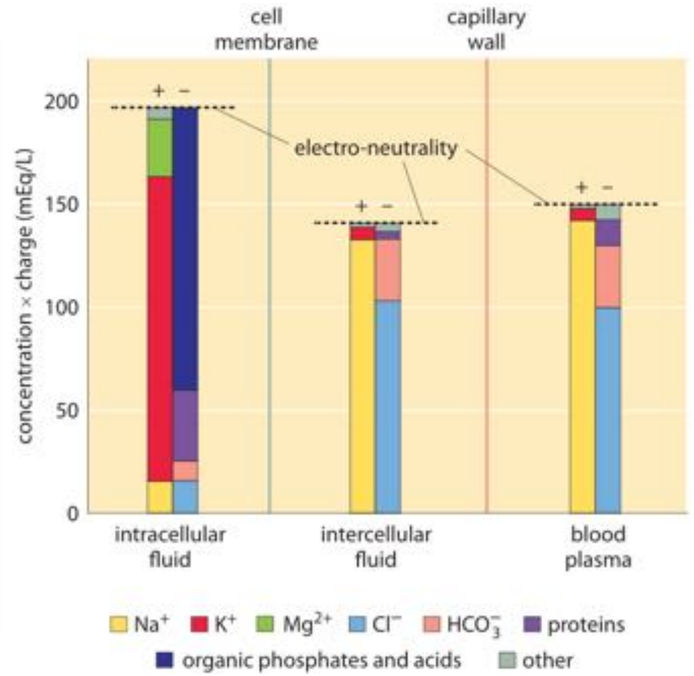




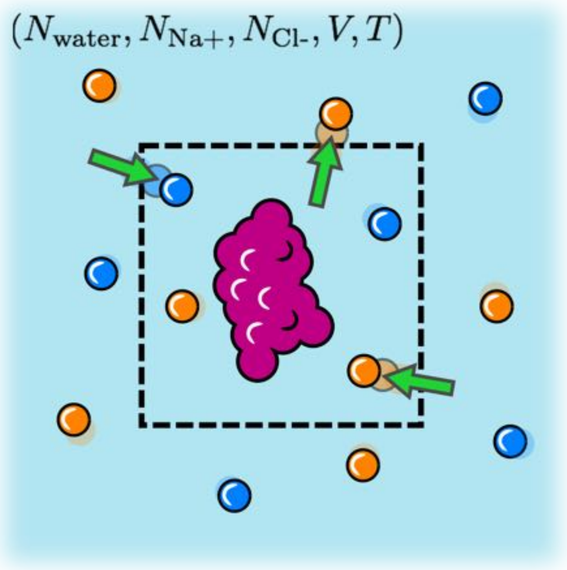


SONYA HANSON

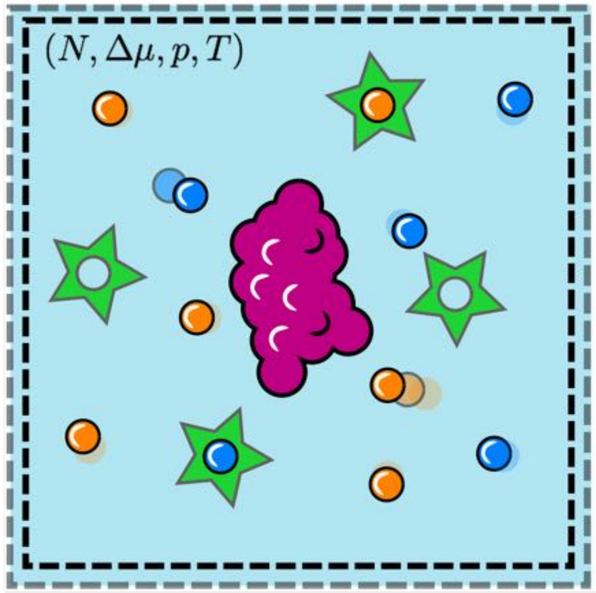
REALISTIC SALT CONCENTRATIONS CAN BE MODELED WITH GRAND CANONICAL MONTE CARLO



real system



semigrand ensemble



semigrand ensemble

$$\pi(x, \theta; \Delta\mu, N, p, T) = \frac{1}{\Xi(\Delta\mu, N, p, T)} e^{-\beta[U(x,\theta)+pV(x)+n(\theta)\Delta\mu]}$$

number of ions

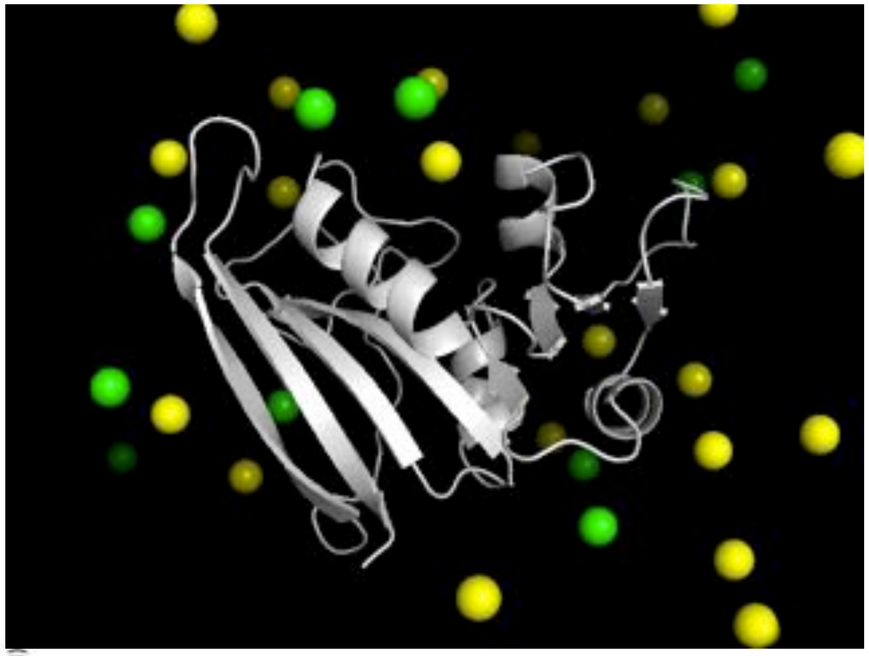
difference in chemical potential between salt and water

$$\Xi(\Delta\mu, N, p, T) = \sum_{\theta, \text{ subject to } \sum_i^N \theta_i = -z} \int \mathrm{d}x\, \pi(x, \theta, \Delta\mu, N, p, T),$$

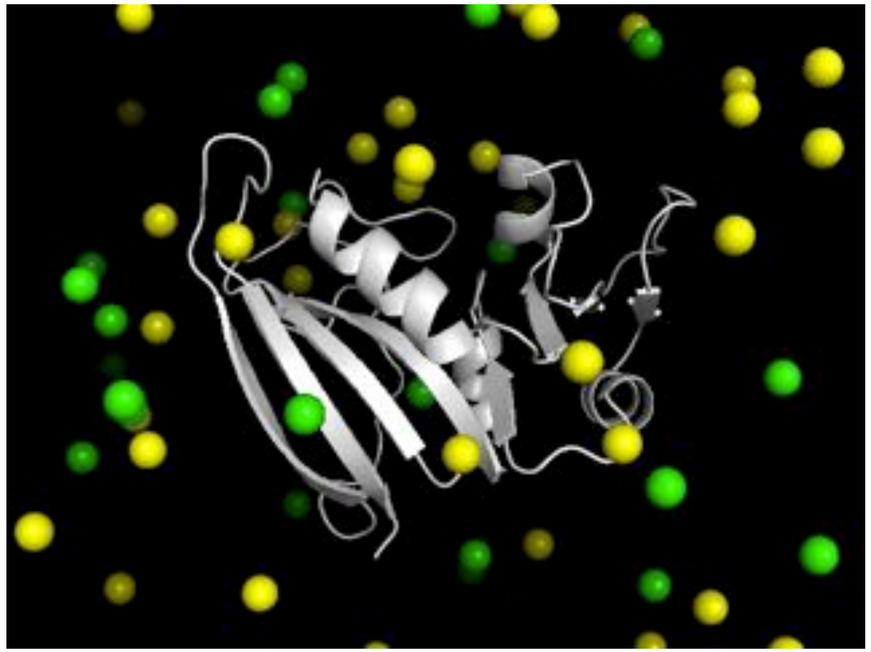
GREGORY ROSS

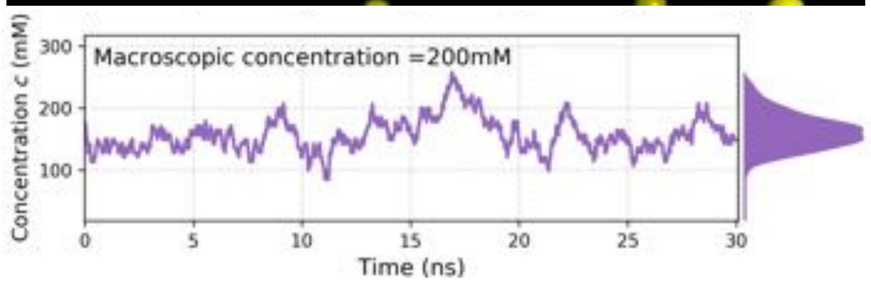
COUNTERION ENVIRONMENTS AROUND BIOMOLECULES ARE HIGHLY DYNAMIC

100 mM NaCl



200 mM NaCl



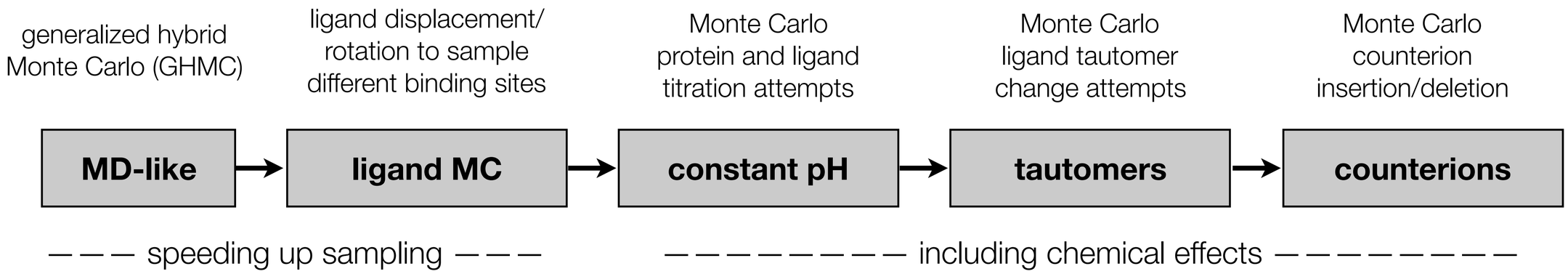


DHFR in TIP3P with PME

AMBER99SB-ILDN with Cheatham-Joung ion parameters

GREGORY ROSS

MARKOV CHAIN MONTE CARLO (MCMC) PROVIDES A FLEXIBLE FRAMEWORK FOR MIX-AND-MATCH ENHANCEMENTS



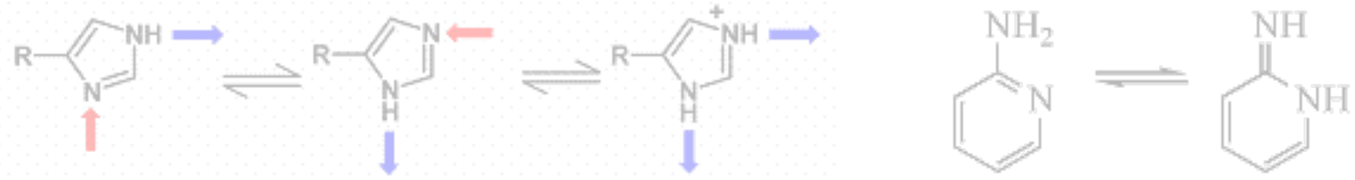
THIS IS AN OPPORTUNITY TO MIX-AND-MATCH COMMUNITY-SOURCED SAMPLING ALGORITHMS

FREE ENERGY CALCULATIONS FAIL FOR THREE MAIN REASONS

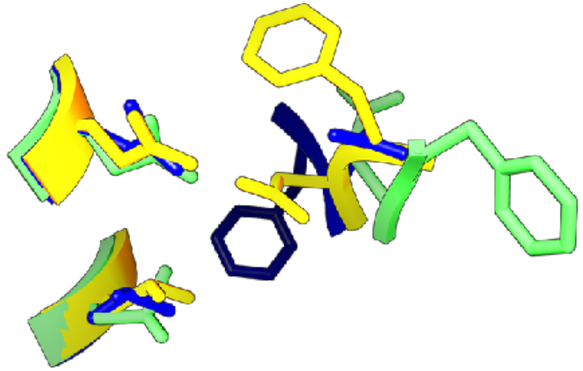
1. The forcefield does a poor job of modeling the physics of our system

$$V(\mathbf{q}) = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

2. We're missing some essential chemical in our simulations
(e.g. protonation states, tautomers, covalent association)

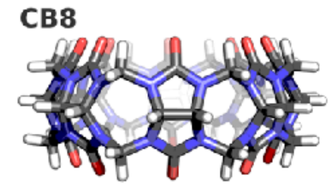


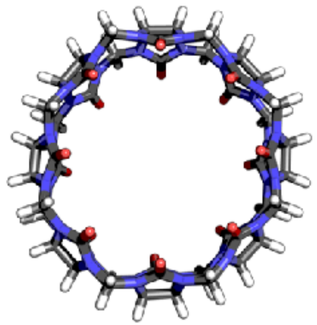
3. We haven't sampled all of the relevant conformations

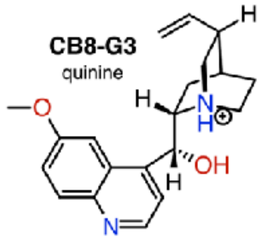


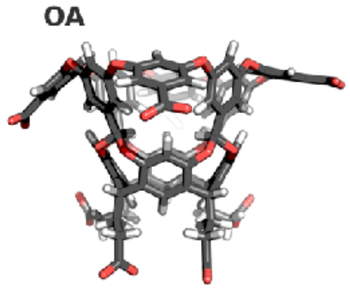
HOW QUICKLY DO DIFFERENT FREE ENERGY PACKAGES SAMPLE TO ACHIEVE CONVERGENCE

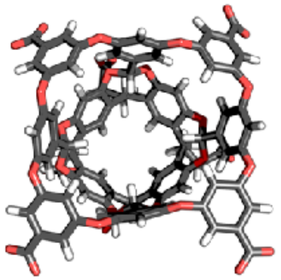
SAMPL6 SAMPLing Challenge

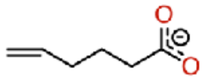




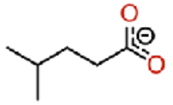




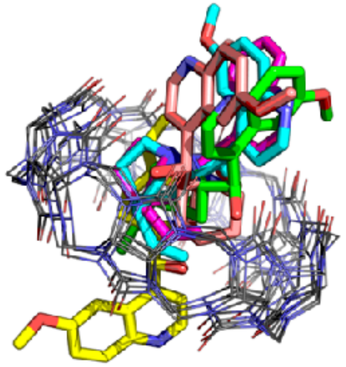


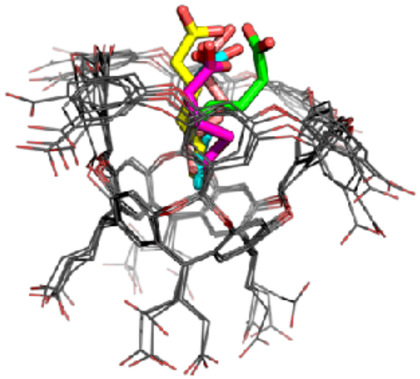


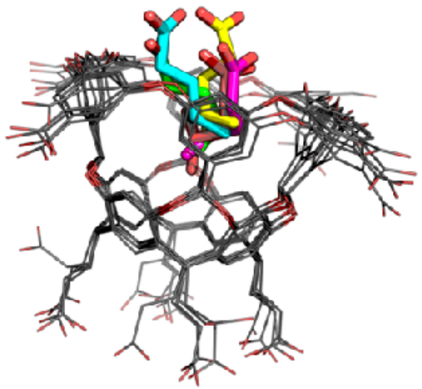
OA-G3
5-hexenoic acid

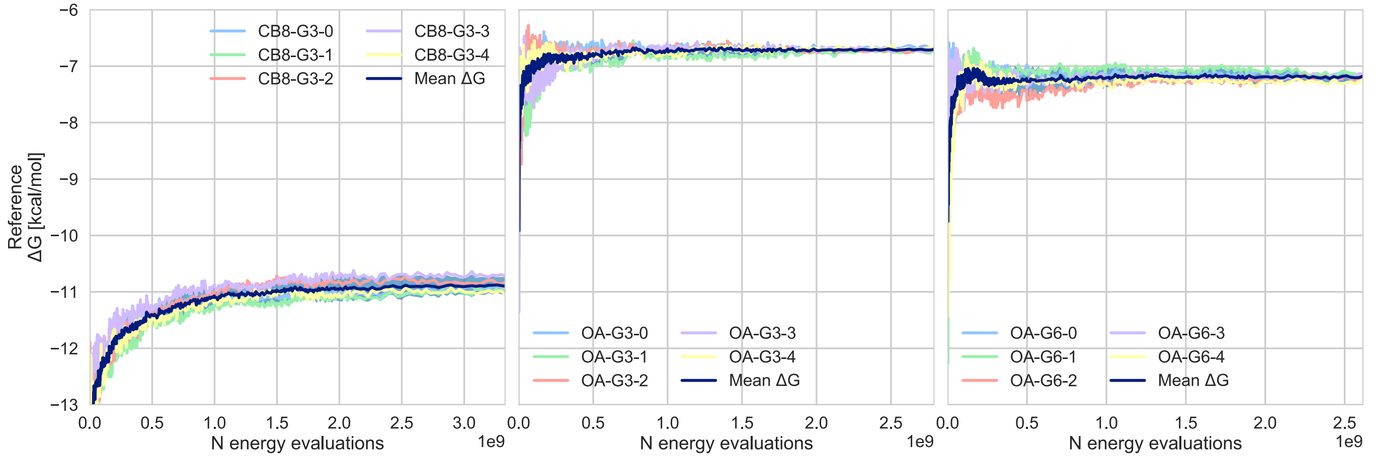


OA-G6
4-methylpentanoic acid







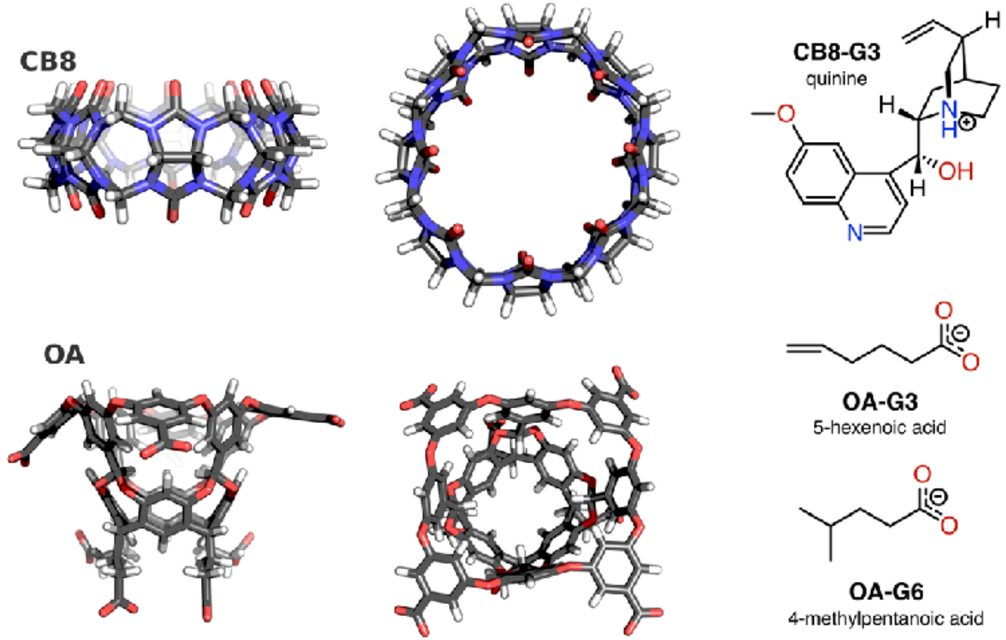


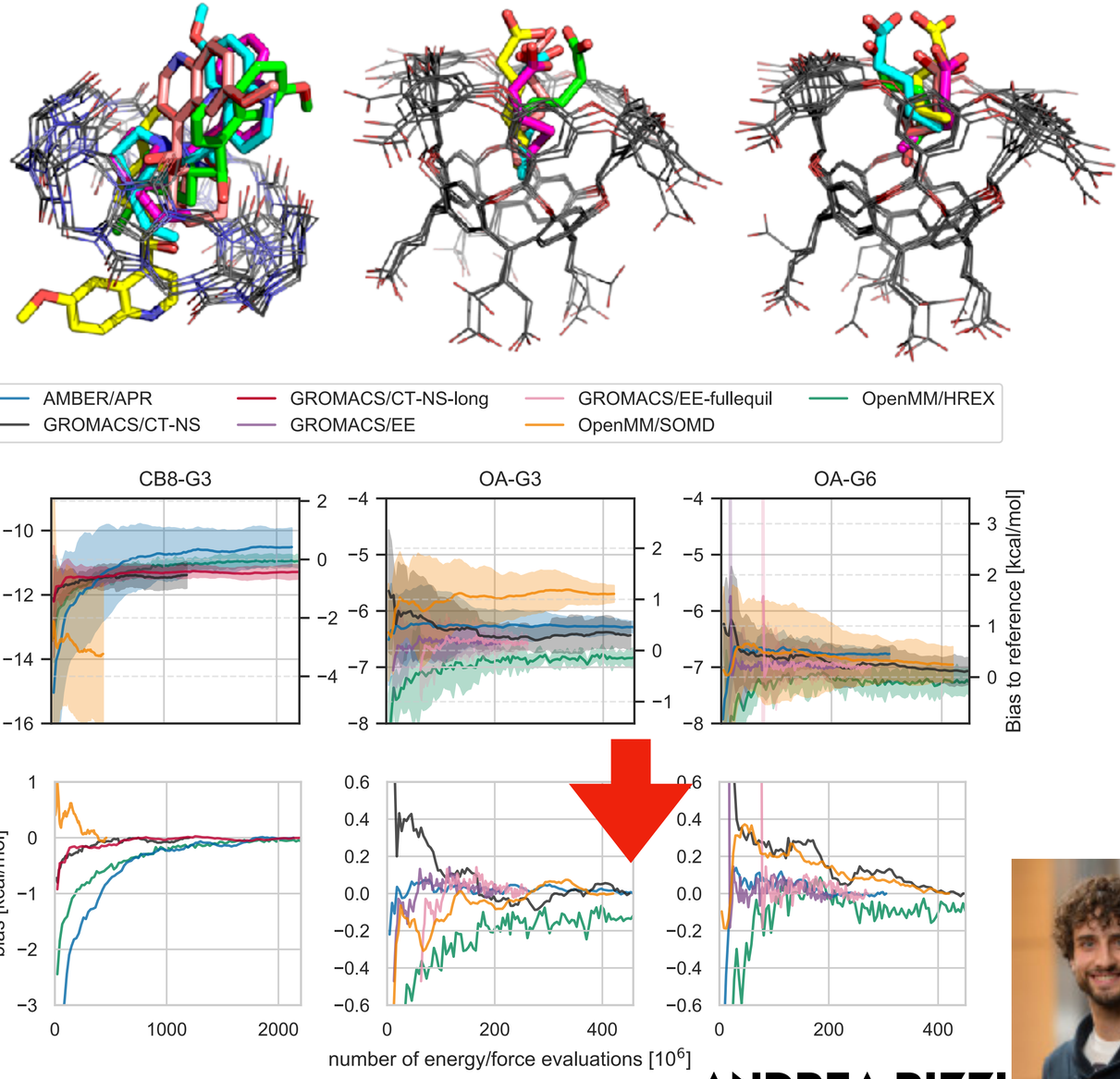
For a single method, rate at which five replicates with different initial configurations converge should measure efficiency

ANDREA RIZZI

HOW QUICKLY DO DIFFERENT FREE ENERGY PACKAGES SAMPLE TO ACHIEVE CONVERGENCE

SAMPL6 SAMPLing Challenge





But none of the methods can agree on what the ΔG is for the same force field, partial charges, and simulation box!

ANDREA RIZZI

MOLSSI SPONSORED MOLECULAR SIMULATION INTEROPERABILITY WORKSHOP

Can we adopt a standard set of force field terms to support in all major simulation packages?

Can we adopt a single way to encode force fields or parameterized molecular systems?

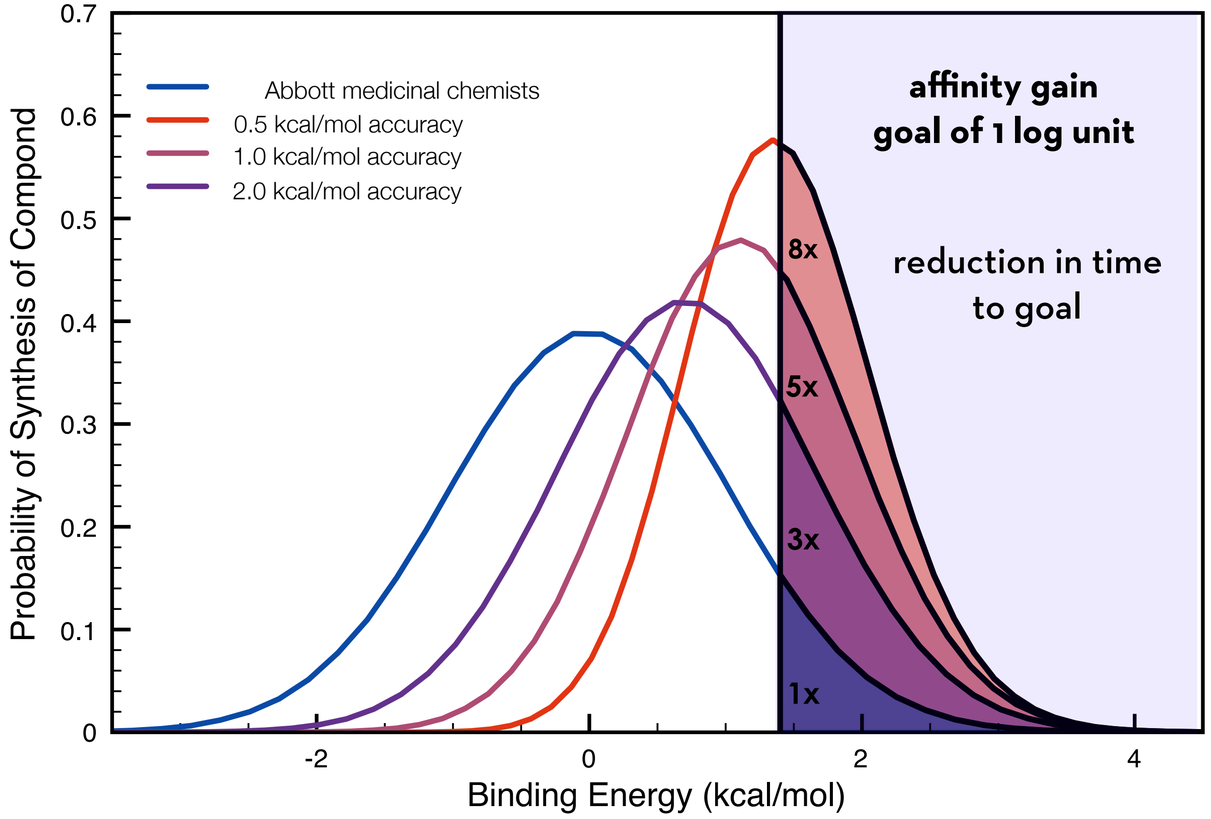
Can we adopt a unifying input standard for initiating molecular simulations?

3-5 NOV 2019 - WILLIAMSBURG, BROOKLYN

HOW CAN WE MEASURE IMPACT?

HOW CAN WE ASSESS WHERE WE ARE?

The simplest statistical model of lead optimization

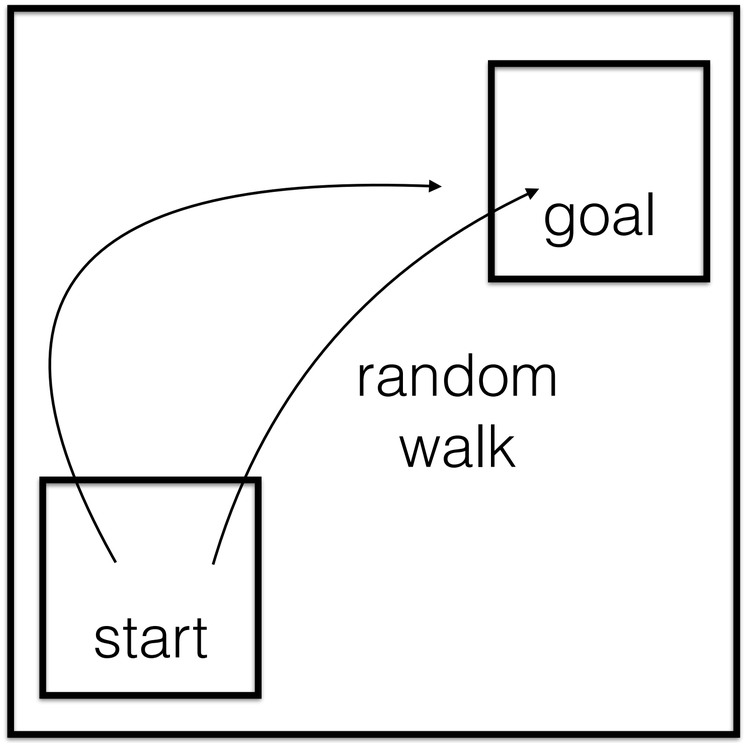


With industry or NIH NCATS data, more sophisticated statistical models could provide more realistic cost/benefit analysis of predictive modeling tools

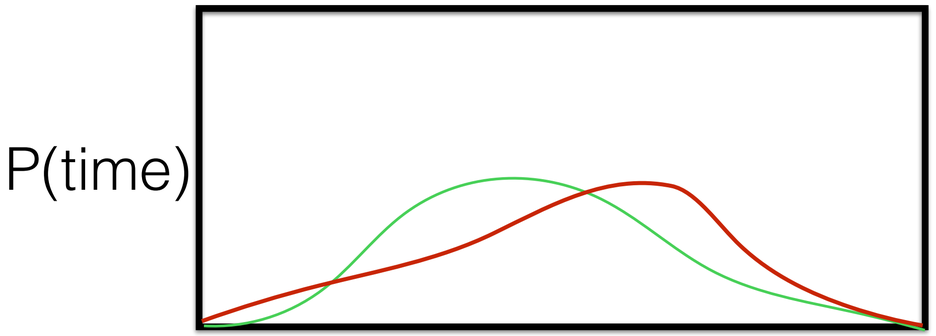
M. R. Shirts, D. L. Mobley and Scott P. Brown. "Free energy calculations in structure-based drug design", in Drug Design: Structure- and Ligand-Based Approaches, pgs. 61-86, 2010.

HOW CAN WE ASSESS WHERE WE ARE?

Can we build a statistical model of impact?



How do we measure
impact of a predictive tool
with X kcal/mol accuracy?



Where can we get appropriate statistics about predictive accuracy?

D3R Drug Design Data Resource

Collect unpublished protein-ligand data, typically from pharma

Use the data to hold blinded prediction challenges

Hold workshops and to discuss and learn

Drive progress in computer-aided drug design

D3R's NIH grant runs out this year as its funding mechanism terminates. What next?

http://drugdesigndata.org

PIs: Michael Gilson and Rommie Amaro

WE NEED TO BE BETTER, AS A COMMUNITY, ABOUT PARTICIPATING IN BLIND CHALLENGES

All free energy participants for Cathepsin challenge of GC4

Show 100 entries

Download data

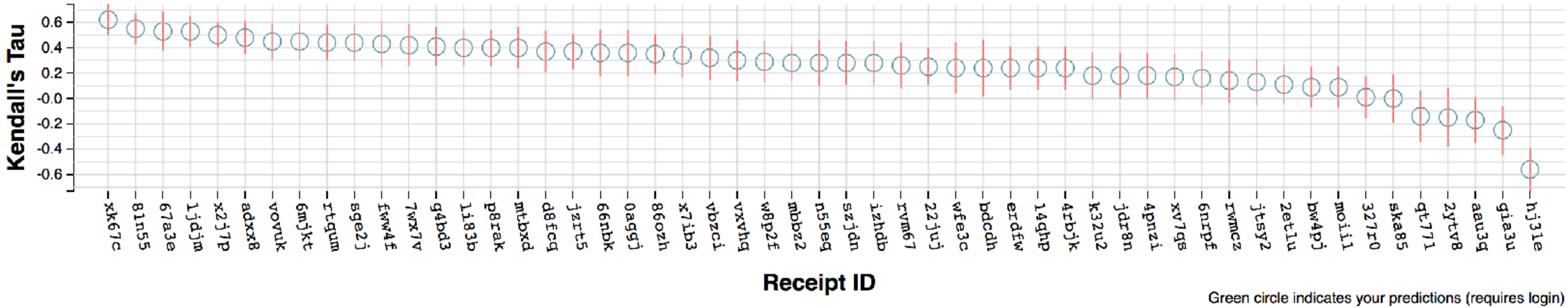
Receipt ID	Submitter Name	PI/Group Name	Number of Ligands	Kendall's τ	Kendall's τ Error	Spearman's ρ	Spearman's ρ Error	Pearson's r	Pearson's r Error	$RMSE_c$	$RMSE_c$ Error	Method Name
3gjm2	Junjie Zou	Carlos simmerling/Daniel Raleigh	39	0.62	0.09	0.8	0.11	0.82	0.1	0.49	0.08	am1-bcc/ti
tkkqh	Chuan Tian	Carlos simmerling	39	0.62	0.09	0.8	0.11	0.82	0.1	0.49	0.08	am1-bcc/ti
53cvi	Junjie Zou	Carlos simmerling/Daniel Raleigh	39	0.61	0.09	0.78	0.11	0.8	0.1	0.5	0.08	am1-bcc/ti
szgth	Chuan Tian	Carlos simmerling	39	0.61	0.09	0.78	0.11	0.8	0.1	0.5	0.08	am1-bcc/ti
d8n3z	Antonia Mey	Julien michel	39	0.09	0.12	0.07	0.17	0.09	0.16	1.16	0.13	alchemical free energies using biosimspace.

https://drugdesigndata.org/php/d3r/gc4/combined/free-energy/index.php?component=1147&group=default

IT'S HARD TO LEARN FROM FAILURE
WHEN FAILURES ARE TOO COMPLEX TO UNDERSTAND

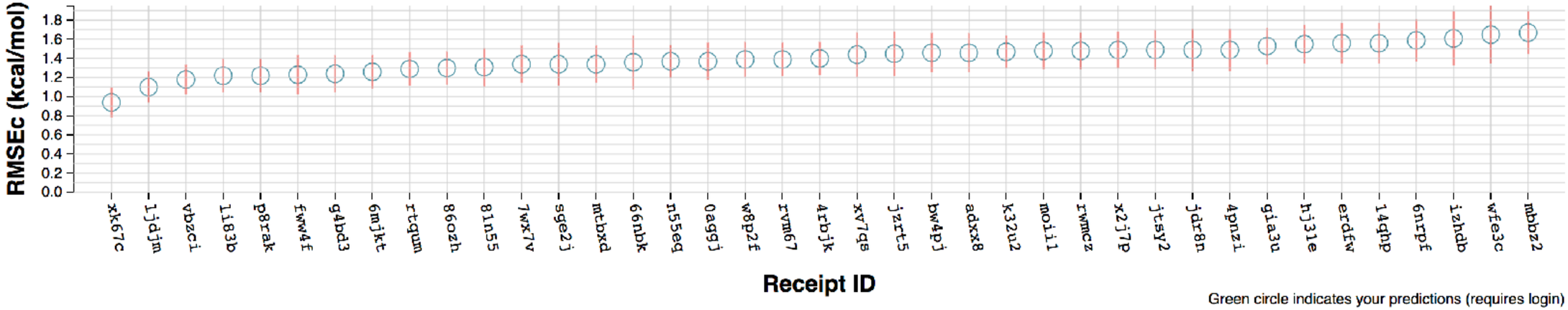
Grand Challenge 2

Free Energy Set 2 (Stage 2) - Kendall's Tau



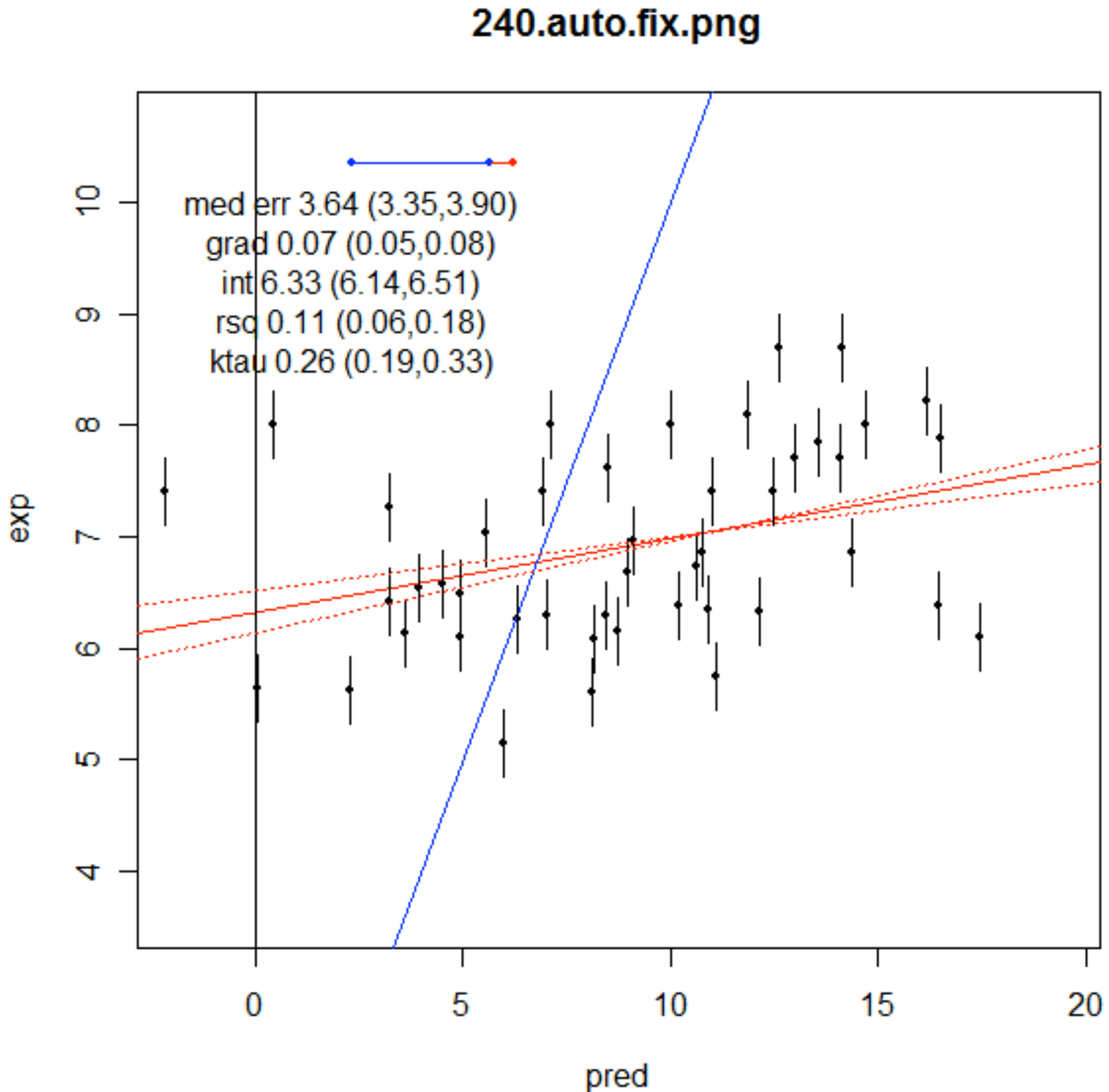
Green circle indicates your predictions (requires login)

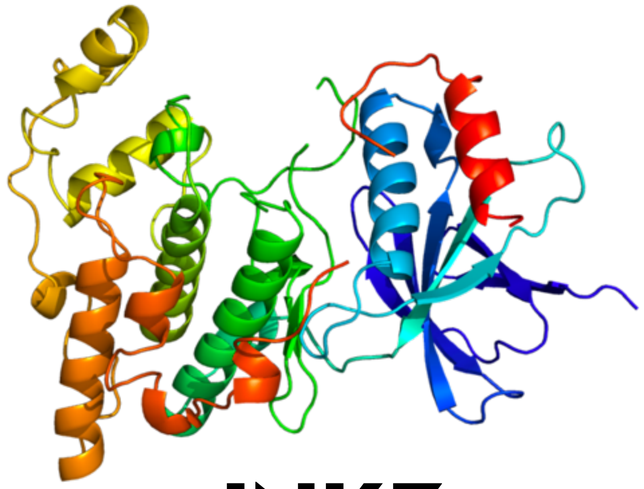
Free Energy Set 2 (Stage 2) - RMSEc



Green circle indicates your predictions (requires login)

IT'S HARD TO LEARN FROM FAILURE
WHEN FAILURES ARE TOO COMPLEX TO UNDERSTAND





JNK3 (2007 BLIND CHALLENGE)

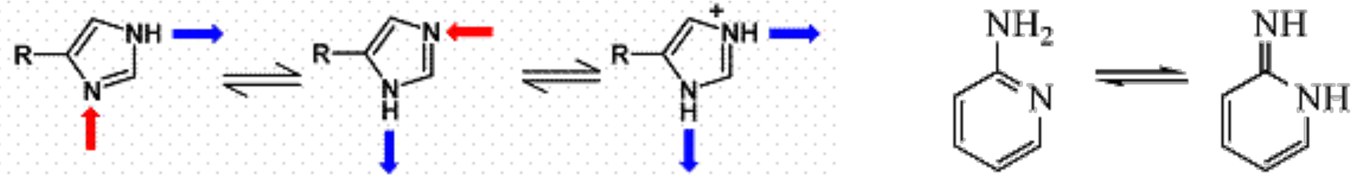
THE USUAL SUSPECTS ALL CONSPIRE

1. The forcefield does a poor job of modeling the physics of our system

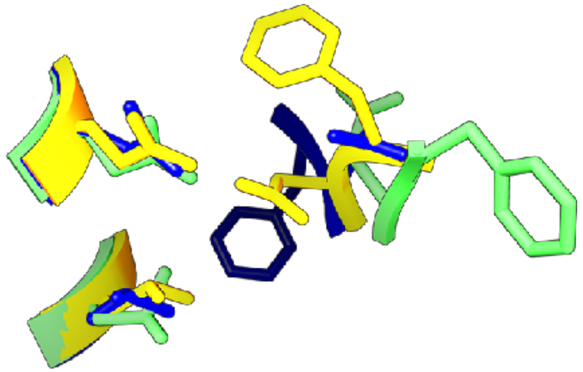
$$V(\mathbf{q}) = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

2. We're missing some essential chemical effects in our simulations

(e.g. protonation states, tautomers, cosolvents, counterions, covalent association)



3. We haven't sampled all of the relevant conformations of protein or ligand



CAN WE "DIVIDE AND CONQUER" BY ISOLATING INDIVIDUAL CHALLENGES?

SAMPL BLIND COMMUNITY CHALLENGES

Model systems of **intermediate complexity** to focus community on challenges in blind tests

Model protein-ligand systems

Isolate individual physical challenges (e.g. binding of charged ligands)

Physical properties

Tests of forcefield accuracy in hydrated or protein-like environments

Isolate chemical effects (protonation states, ligand conformations) without slow protein timescales

Host-guest systems

Binding of small drug-like molecules with protein-like affinities, without slow protein timescales

SAMPL0
2007

JNK3 kinase inhibitors
hydration free energies

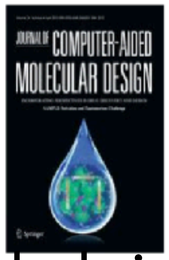
Anthony Nicholls

SAMPL1
2008

CDK2 kinase inhibitors
hydration free energies

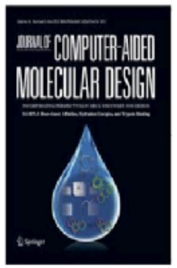
SAMPL2
2009

hydration free energies
tautomer ratios



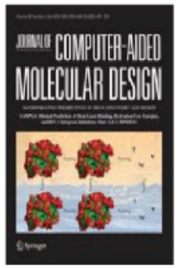
SAMPL3
2011

trypsin inhibitors
hydration free energies



SAMPL4
2013

HIV-1 integrase inhibitors
hydration free energies
octoacid host-guest
CB7 host-guest

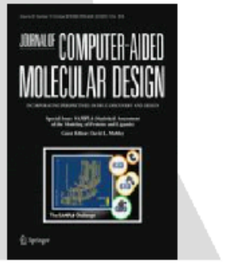


SAMPL5
2016

distribution coefficients
CBClip host-guest
CB7 host-guest

SAMPL6
Part I - 2018

acid dissociation constants
host-guest affinity
sampling



SAMPL6
Part II - 2019

partition coefficients

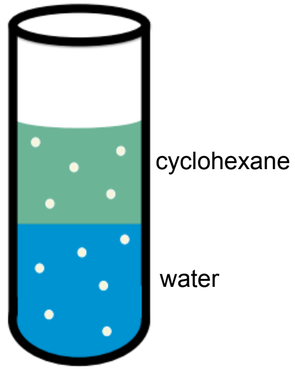
https://samplchallenges.github.io/

handed over to academic stewards . . . . . .

BLIND CHALLENGES CAN DRIVE PROGRESS BY FOCUSING COMMUNITY EFFORT

SAMPL0 2007	SAMPL1 2008	SAMPL2 2009	SAMPL3 2011	SAMPL4 2013	SAMPL5 2016
hydration free energies	hydration free energies	hydration free energies	hydration free energies	hydration free energies	WE RAN OUT OF DATA!

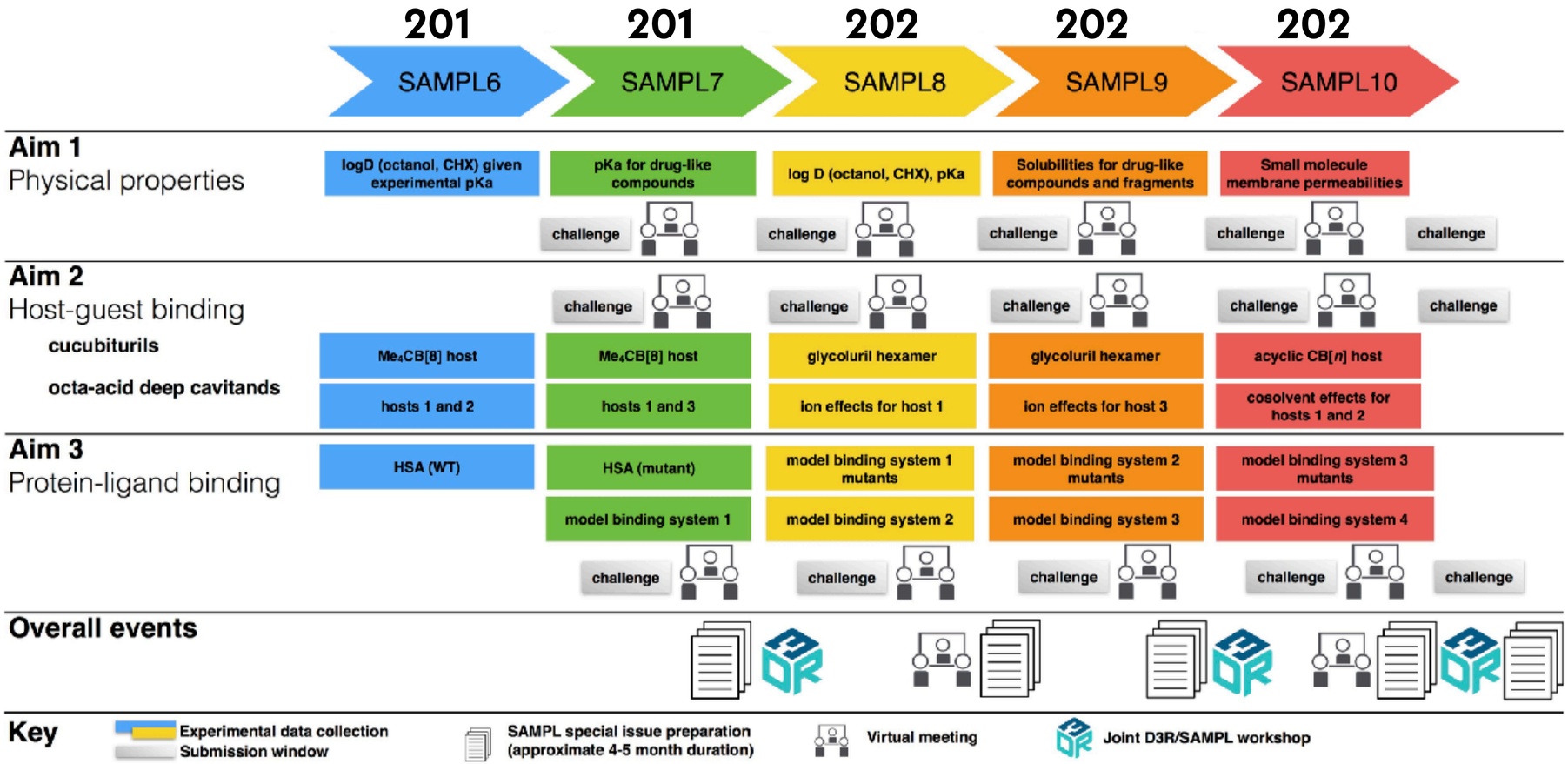
Lots of disagreement in predictions → **Can tell when experiments are wrong** X



https://samplchallenges.github.io/

SAMPL WILL FIELD MANY MORE CHALLENGES

https://samplchallenges.github.io/



Mehtap Işık
MSKCC senior PhD student

Mon/Tue #9

Working with NIH NCATS and DiamondMX / XChem and Daniel Keedy (CCNY) to collect new protein:ligand datasets, but we're always looking for industry collaborations!

D3R / SAMPL MEETING: 22-23 AUG 2019

Home / D3R 2019 Workshop

D3R 2019 Workshop

https://drugdesigndata.org

Overview

Agenda

Workshop Venue

Accommodations

Online Registration

Contact

Overview

Registration is now open!

We are pleased to invite you to attend the D3R 2019 Workshop that will be held on August 20 - 23, 2019 at the Hotel La Jolla

Registration Fees:
Faculty & Regular Attendees: $200
Students or Post docs: $100

Dates: Aug 20-23, 2019
Location: Hotel La Jolla
Hotel Block: Hotel La Jolla with group code "UCD3R"
Online Registration: https://www.eventbrite.com/e/registration-62017434

Save the date: Joint GCC / EuroSAMPL meeting 1-5 Nov 2020

HOW CAN WE, AS A COMMUNITY SOLVE CHALLENGES IN OUR FIELD TO IMPROVE SUCCESS RATES?

CURRENT LICENSING MODELS FOR FREE ENERGY CALCULATION CODES

	COMMERCIAL	ACADEMIC/MIXED	PERMISSIVE OPEN SOURCE
Cost to license	**$$$**	**industry $$$** **academics ∅-$**	∅
Modify?	**NO**	**YES** (source provided)	**YES!**
Use in your own tools?	**LOL NO**	**HARD** **(users need to get their own licenses)**	**YES!**
Examples	**FEP+**	**AMBER TI GPU** **CHARMM**	**gromacs** **NAMD** **OpenMM**

WE CAN LOOK TO THE MACHINE LEARNING ECOSYSTEM FOR INSPIRATION

What did TensorFlow accomplish?

* Created new opportunities
* Accelerated rate of progress

OPEN SOURCE SOFTWARE ECOSYSTEMS HAVE THE POTENTIAL TO ACCELERATE PROGRESS

Omnia

empowering molecular simulation

AN OPEN SOURCE, HIGH PERFORMANCE, HIGH USABILITY ECOSYSTEM FOR PREDICTIVE BIOMOLECULAR SIMULATION

OpenMM

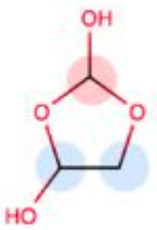
PDB FIXER

MSM BUILDER

PyEMMA

MDTRAJ

ENSEMBLER



http://omnia.md

OPEN SOURCE SOFTWARE FOR PHYSICAL MODELING CAN HAVE IMPACT

OPENMM

We contribute to the development of the GPU-accelerated OpenMM molecular modeling package, one of the fastest platforms for molecular simulation.

Install with conda | Anaconda Cloud 7371480 | downloads 188k total

that's a lot of impact!

OPENMMTOOLS

A useful library layer of high-quality integrators, a composable MCMC framework, advanced sampling schemes, alchemical factories, and test systems for OpenMM.

Read the documentation at openmmtools.readthedocs.io
Go to the GitHub page

Install with conda | Anaconda Cloud 0.18.1 | downloads 75k total

YANK

An open, extensible platform for GPU-accelerated alchemical binding free energy calculations using the OpenMM toolkit.

- Read the documentation at getyank.org
- Go to the YANK GitHub page
- Read the YANK paper on identifying unknown binding sites
- Get it through conda

Install with conda | Anaconda Cloud 0.24.0 | downloads 8k total | DOI 10.5281/zenodo.2577832

http://choderalab.org/code

THOUGHTFUL LICENSING MODELS CAN ENSURE RESEARCH HAS MAXIMUM SCIENTIFIC IMPACT

Goal: Ensure our work has **maximum impact** by allowing use, modification, and redistribution. Aim to explicitly rescind any restrictions that would prevent this.

reproducible research product

paper
data
experiment (code)
documentation

Stodden, Victoria, Enabling Reproducible Research: Open Licensing for Scientific Innovation (March 3, 2009). International Journal of Communications Law and Policy, Forthcoming. Available at SSRN: https://ssrn.com/abstract=1362040

https://web.stanford.edu/~vcs/papers/Licensing08292008.pdf

THOUGHTFUL LICENSING MODELS CAN ENSURE RESEARCH HAS MAXIMUM SCIENTIFIC IMPACT

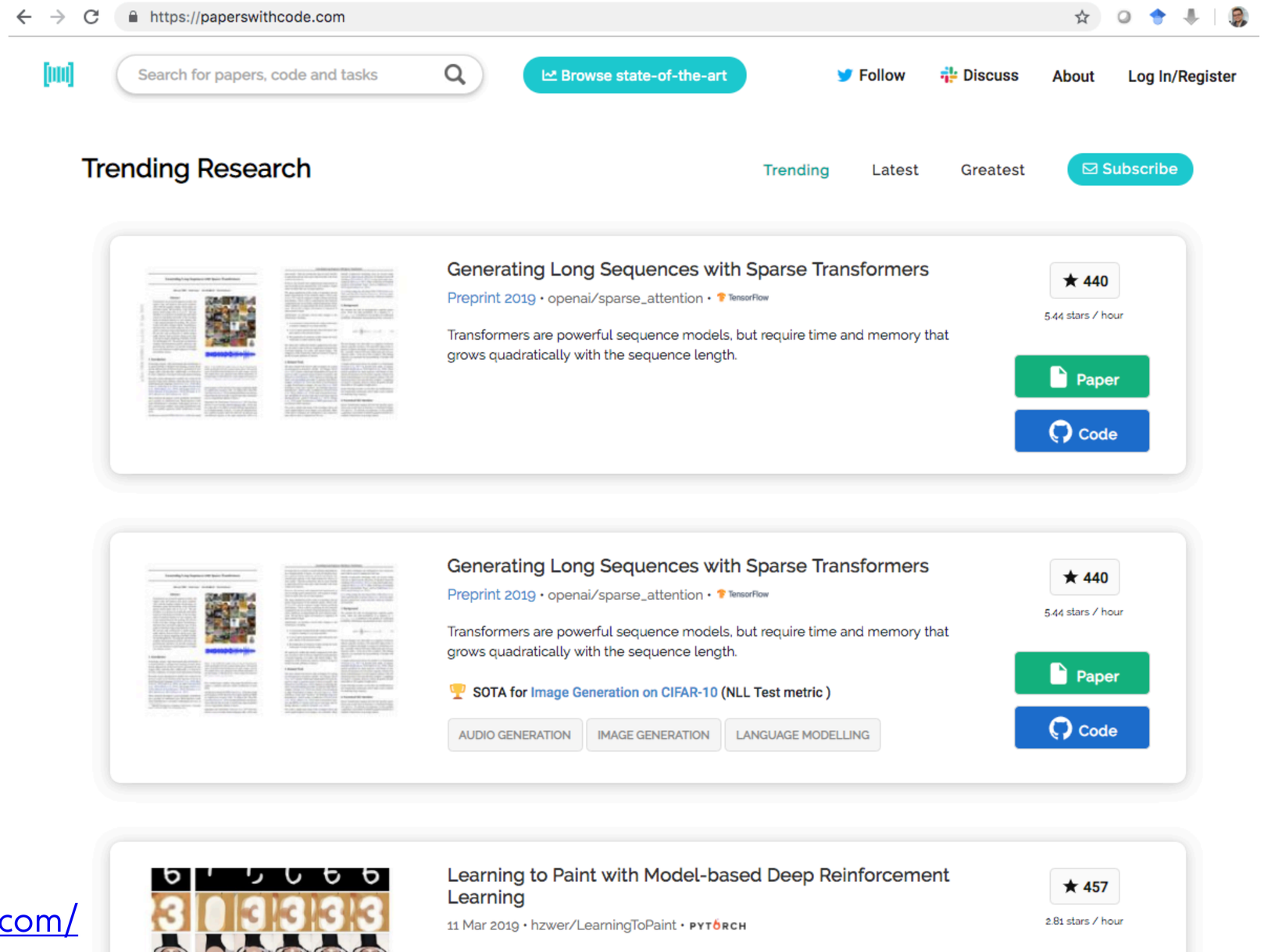
Goal: Ensure our work has **maximum impact** by allowing use, modification, and redistribution. Aim to explicitly rescind any restrictions that would prevent this.

reproducible research product	**licenses that encourage others to build on work**
paper	**CC-BY 4.0** https://creativecommons.org/licenses/by/4.0/
data	**CC-BY 4.0**
experiment (code)	**MIT, BSD 3-clause** https://opensource.org/licenses/MIT
documentation	**CC-BY 4.0**

Stodden, Victoria, Enabling Reproducible Research: Open Licensing for Scientific Innovation (March 3, 2009). International Journal of Communications Law and Policy, Forthcoming. Available at SSRN: https://ssrn.com/abstract=1362040

https://web.stanford.edu/~vcs/papers/Licensing08292008.pdf

MACHINE LEARNING CELEBRATES REPRODUCIBLE RESEARCH



https://paperswithcode.com/

MODERN ML FRAMEWORKS PROVIDE A LEVEL OF ABSTRACTION THAT ENABLES HIGH PRODUCTIVITY

```
import tensorflow as tf
mnist = tf.keras.datasets.mnist

(x_train, y_train),(x_test, y_test) = mnist.load_data()
x_train, x_test = x_train / 255.0, x_test / 255.0

model = tf.keras.models.Sequential([
  tf.keras.layers.Flatten(input_shape=(28, 28)),
  tf.keras.layers.Dense(128, activation='relu'),
  tf.keras.layers.Dropout(0.2),
  tf.keras.layers.Dense(10, activation='softmax')
])

model.compile(optimizer='adam',
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])

model.fit(x_train, y_train, epochs=5)
model.evaluate(x_test, y_test)
```

Run code now Try in Google's interactive notebook

load your tools

grab a dataset

define a new kind of model

declare your objectives in training it

fit it

use it

https://www.tensorflow.org/overview

Why can't we make it this easy to do new things in computer-aided drug discovery?

WE'RE TRYING!

This is a complete free energy code built with **openmmtools**

```
>>> # Alchemically modify the system
>>> ligand_atoms = mdtraj_topology.select('resname B2')
>>> alchemical_region = alchemy.AlchemicalRegion(alchemical_atoms=ligand_atoms)
>>> factory = alchemy.AbsoluteAlchemicalFactory()
>>> alchemical_system = factory.create_alchemical_system(system, alchemical_region)

>>> # Initialize compound thermodynamic states at different temperatures and alchemical states.
>>> protocol = {'temperature': [300, 310, 330, 370, 450] * unit.kelvin,
...             'lambda_electrostatics': [1.0, 0.5, 0.0, 0.0, 0.0],
...             'lambda_sterics': [1.0, 1.0, 1.0, 0.5, 0.0]}
>>> alchemical_state = alchemy.AlchemicalState.from_system(alchemical_system)
>>> compound_states = states.create_thermodynamic_state_protocol(
...     alchemical_system, protocol=protocol, composable_states=[alchemical_state])

>>> mcmc_move_scheme = mcmc.SequenceMove(move_list=[
...     mcmc.MCDisplacementMove(atom_subset=ligand_atoms),
...     mcmc.MCRotationMove(atom_subset=ligand_atoms),
...     mcmc.LangevinSplittingDynamicsMove(timestep=2.0*unit.femtoseconds, n_steps=n_steps,
...                                        reassign_velocities=True, n_restart_attempts=6)
... ])

>>> # Run the combined Hamiltonian replica exchange + parallel tempering simulation.
>>> hrex_tempering = ReplicaExchange(compound_states, sampler_states, mcmc_move_scheme)
```

ANDREA RIZZI

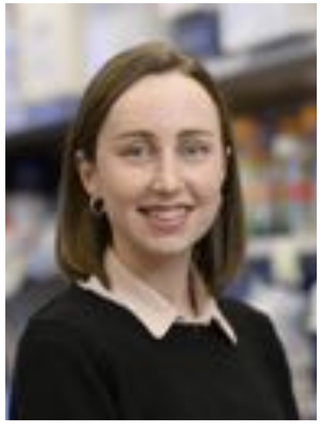
http://github.com/choderalab/openmmtools

#alchemy2020

May 5-7, 2020 (tentative), Boston/Cambridge

PRIMARY ORGANIZERS ("Free Energy TNG")

Kira Armacost
Merck



Hannah Bruce Macdonald
MSKCC

http://tiny.cc/alchemy2020

#alchemy2021

May 2021, Berlin

ORGANIZERS

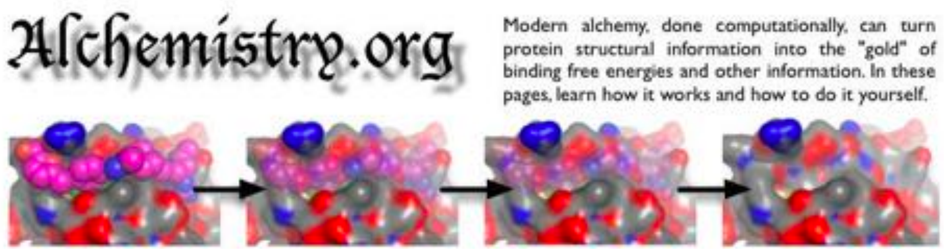
Bert de Groot
MPI Göttingen

?

RESOURCES FOR LEARNING MORE ABOUT ALCHEMICAL FREE ENERGY METHODS

Living Journal of Computational Molecular Sciences (LiveCoMS)
https://www.livecomsjournal.org/
Forthcoming review on alchemical free energy best practices
led by Antonia Mey (soon!)

alchemistry.org
Community resource for alchemical free energy theory and methods



The alchemistry Slack
https://tiny.cc/join-alchemistry-slack

Recent reviews
Aldeghi, Bluck, Biggin. Comput Drug Discovery Design 199, 2018.
Cournia, Allen, Sherman. JCIM 57:2017
Abel, Wang, Mobley. Curr Topics Med Chem 17:2577, 2017
Mobley, Gilson. Annu Rev Biophys 46:531, 2017
Klimovich, Shirts, Mobley. JCAMD 29:397, 2015

CHODERA LAB

National Institutes of Health

SCHRÖDINGER.

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