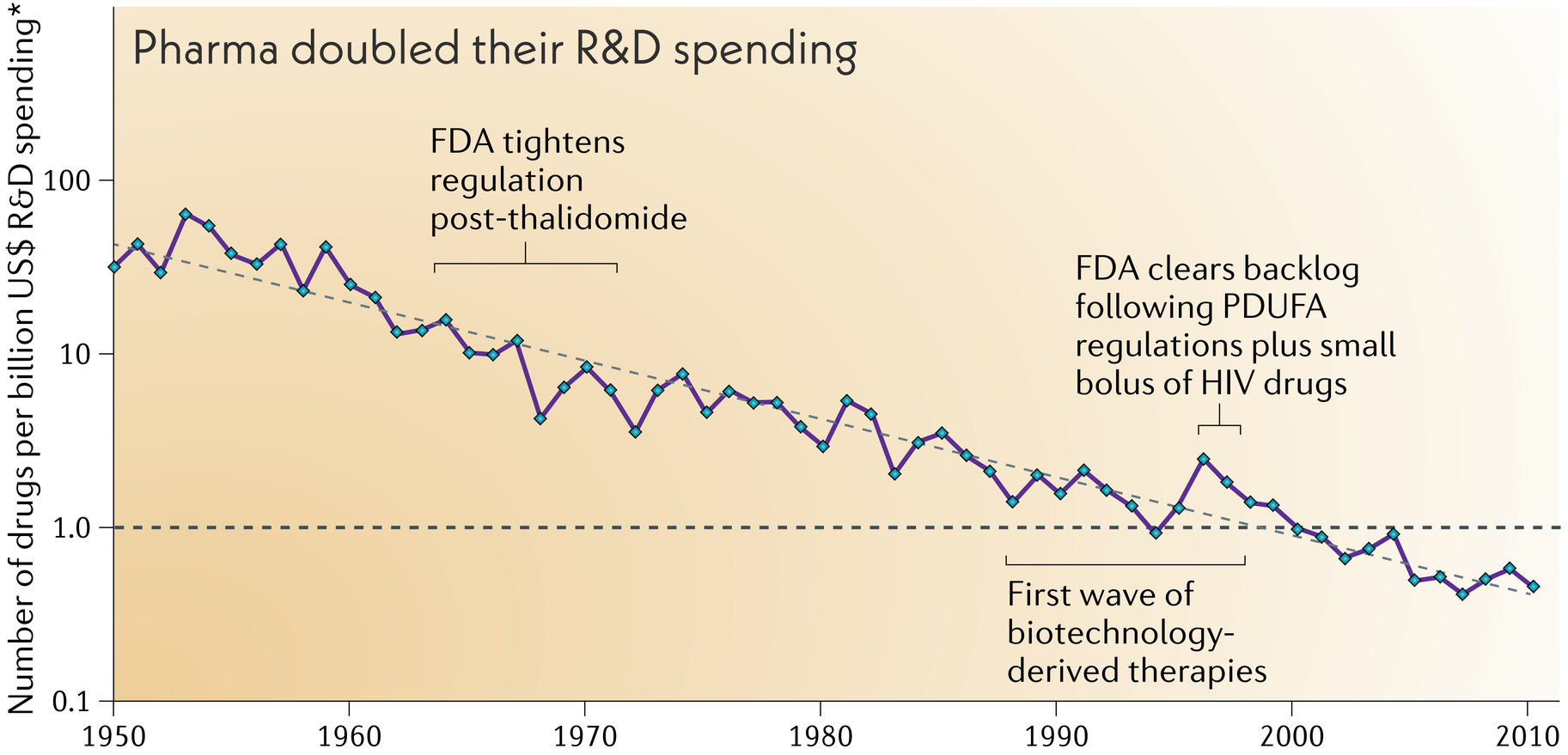
REDESIGNING DRUG DESIGN

John D. Chodera
MSKCC Computational and Systems Biology Program
Sildes will be posted to http://www.choderalab.org/news

DISCLOSURES:
Scientific Advisory Board, OpenEye Scientific, Redesign Science*, Interline Therapeutics*
All funding sources: http://choderalab.org/funding
* Denotes equity interests

DRUG DISCOVERY IS HIGHLY INEFFICIENT

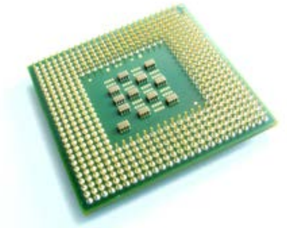
a Overall trend in R&D efficiency (inflation-adjusted)

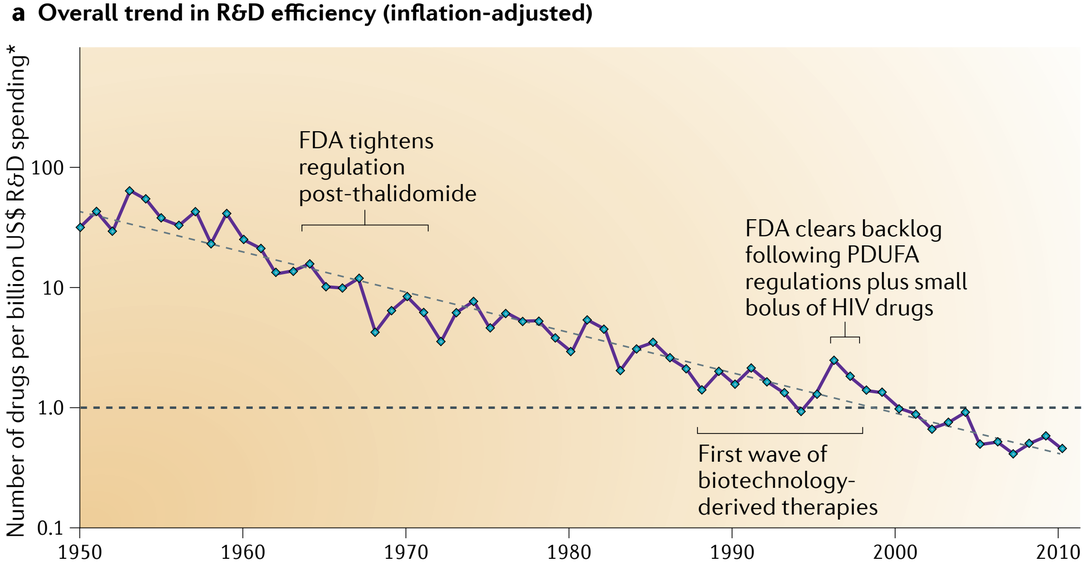


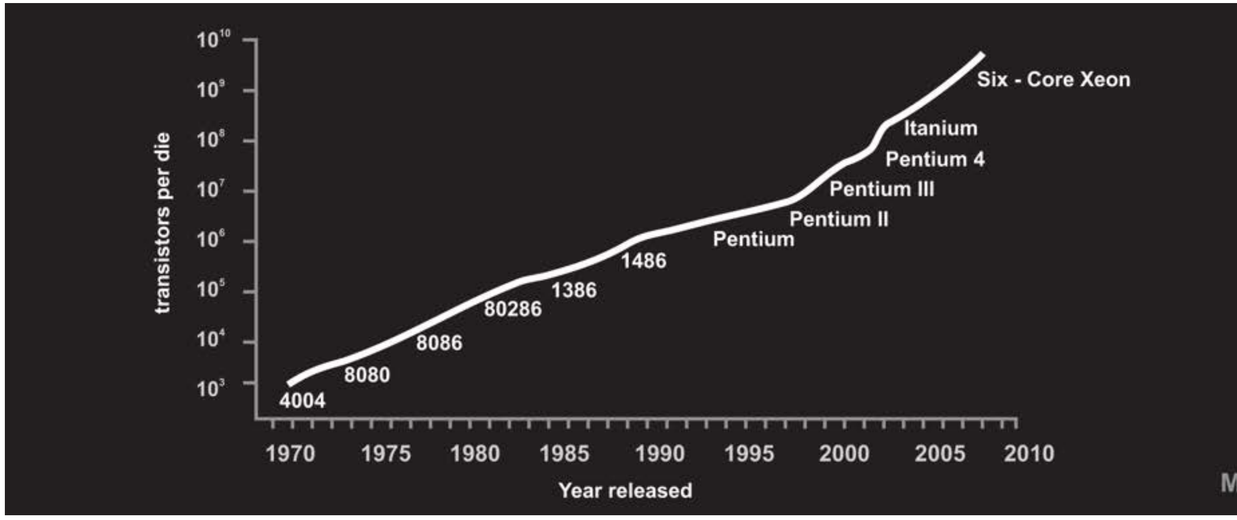
←NOW: >$2.6B/drug

Scannell et al. Nature Rev Drug Disc 11:191, 2012
DiMasi et al. J Health Econ 47:20, 2016

DRUG DISCOVERY IS HIGHLY INEFFICIENT





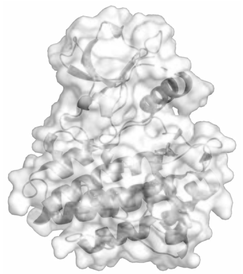


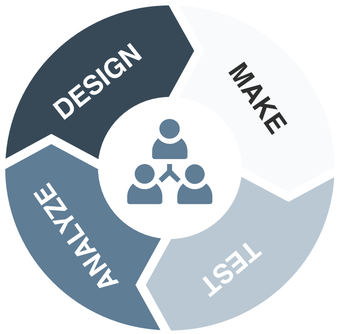
EROOM'S LAW

MOORE'S LAW

DRUG DISCOVERY AND DEVELOPMENT IS COSTLY, TIME-CONSUMING, AND INEFFICIENT

	Target Product Profile (TPP)	Target ID	Discovery	Preclinical	Phase 1 safety	Phase 2 efficacy	Phase 3 effectiveness	NDA new drug approval
value*			$8M	$24M	$58M	$321M	$20M	$1.66B
stage cost*			$13.5M	$3M	$3M	$8M		$3M
% survival			48%	38%	25%	12%	7%	6%
duration*			4.5 y	2.3 y	2.1 y	3.1 y	2.5 y	9 mo





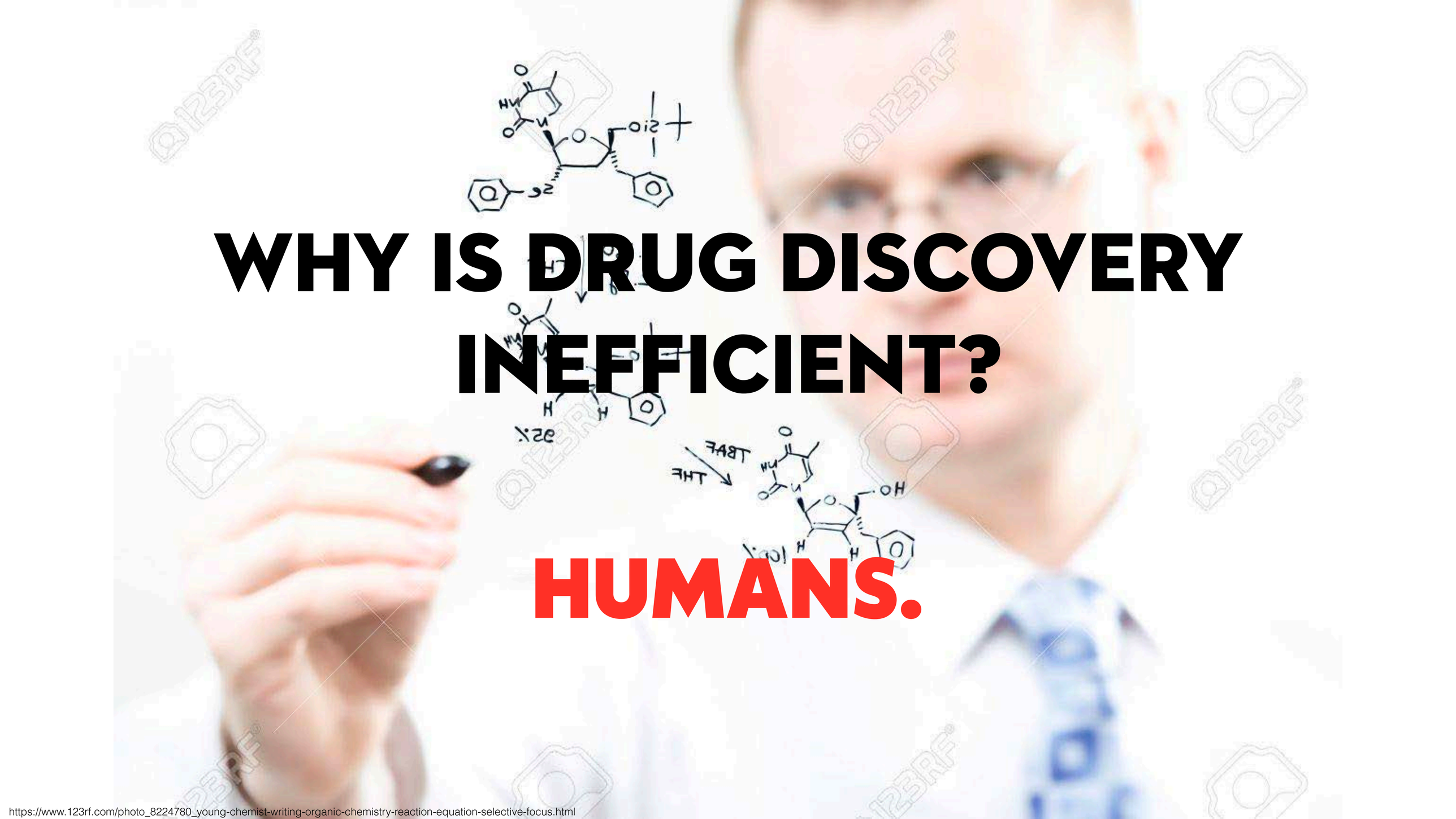
Human-driven design iterations

* denotes mean

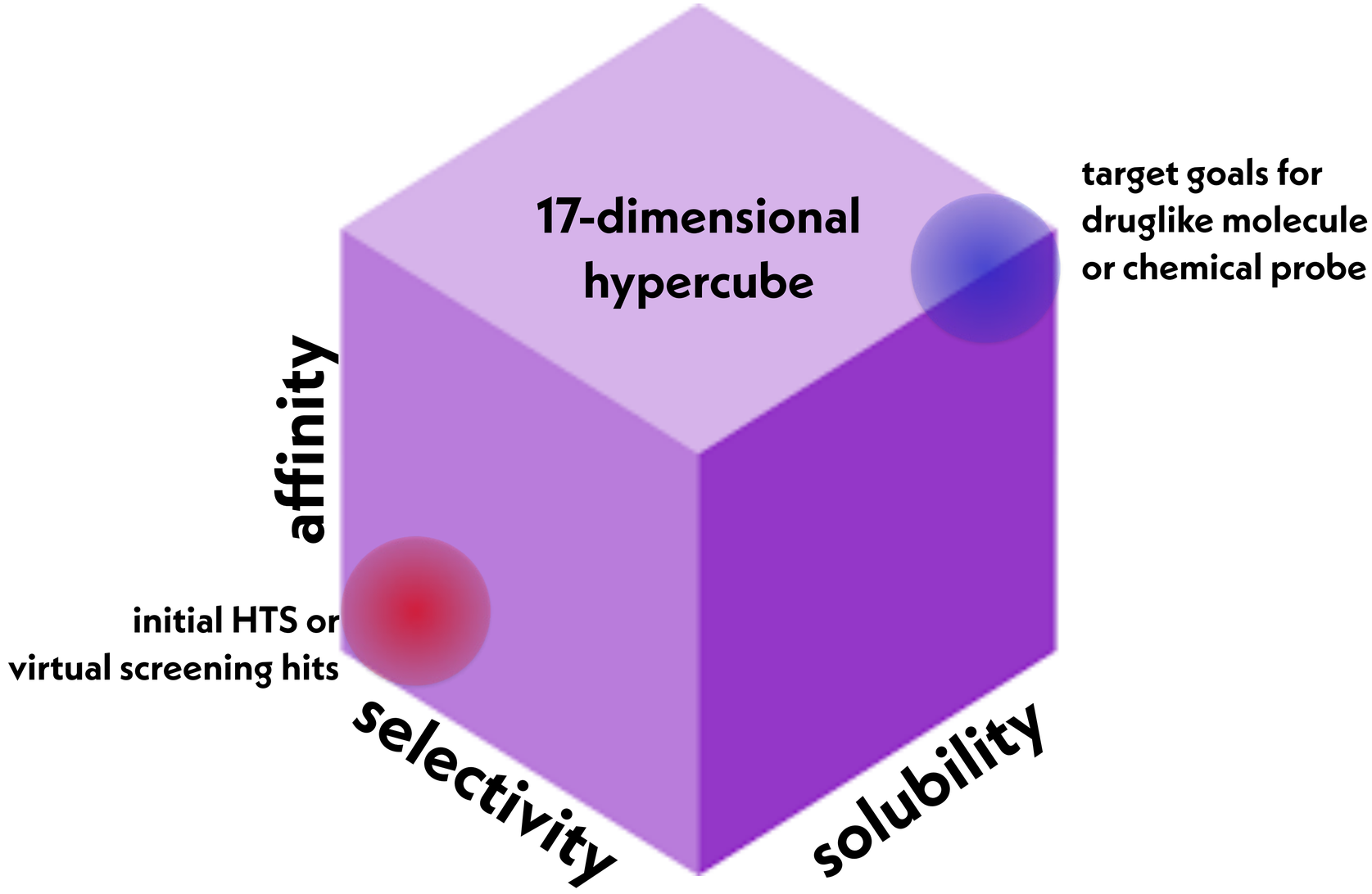
sources: [1] [2] [3] [4] [5]

Global annual prescription drug **market will reach $1.6T** by 2026 [5]

WHY IS DRUG DISCOVERY INEFFICIENT?



WE'RE FACING COMPLEX MULTI-OBJECTIVE DESIGN PROBLEMS



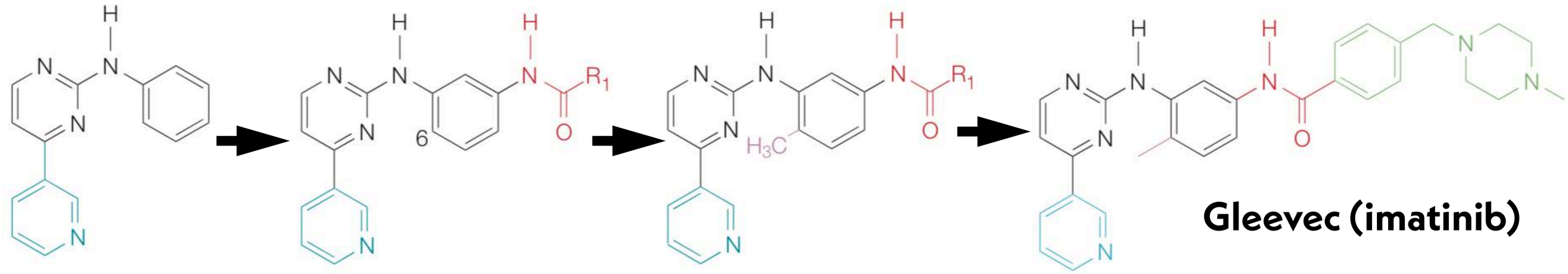
HUMANS FOCUS ON SOLVING ONE PROBLEM AT A TIME, ONE STEP AT A TIME

enhances cellular activity

selectivity for tyrosine kinase

eliminates PKC affinity

increases solubility and oral availability



Gleevec (imatinib)

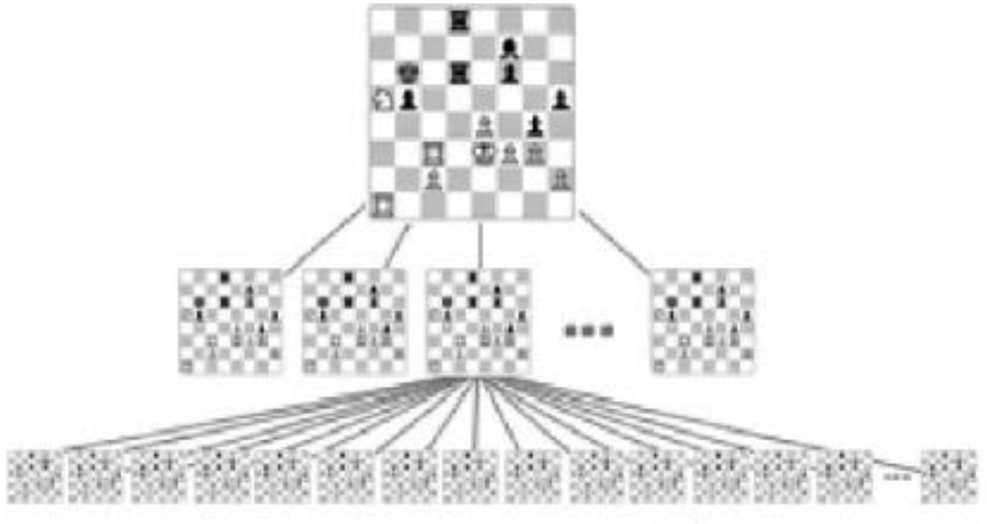
Humans are just not very good at multi-objective optimization.

Herrling PL. Prog. Drug Res. 62, 2005

Capdeville R, Buchdunger E, Zimmerman J, and Matter A. Nat. Rev. Drug Disc. 1:493, 2002

WHY ARE COMPUTERS GOOD AT CHESS?

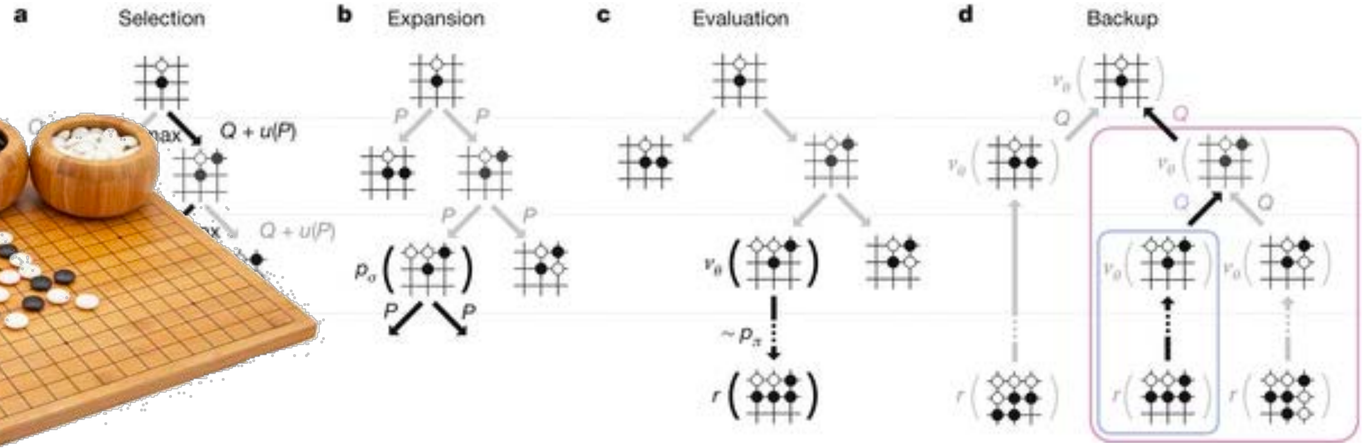
Chess Grand Master Garry Kasparov loses to Deep Blue in 1997
http://scienceblogs.com/startswithabang/2013/02/17/weekend-diversion-chess-is-almost-solved/



Computers can easily look several steps ahead to evaluate positions in light of expected outcomes.

OK, BUT CHEMICAL SPACE IS ENORMOUS. SURELY COMPUTERS CAN'T COPE WITH THIS

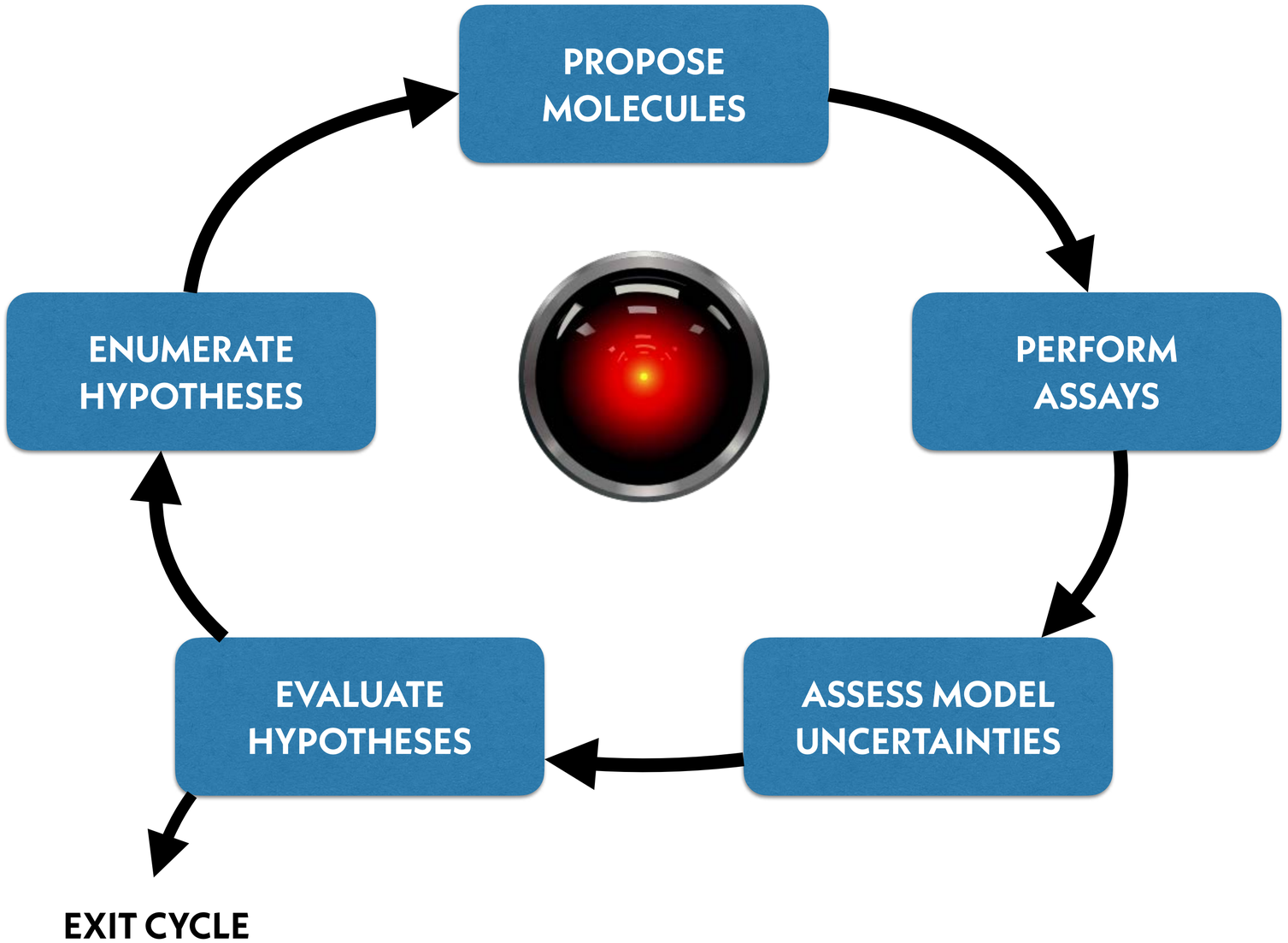
SURELY COMPUTERS COULD NEVER MASTER A GAME AS COMPLEX AS GO



Computers can **still** easily look several steps ahead to evaluate positions in light of expected outcomes.

(AlphaGo) Silver et al. Nature 529:484, 2015

MODEL-DRIVEN DRUG DISCOVERY REMOVES HUMANS FROM THE EQUATION



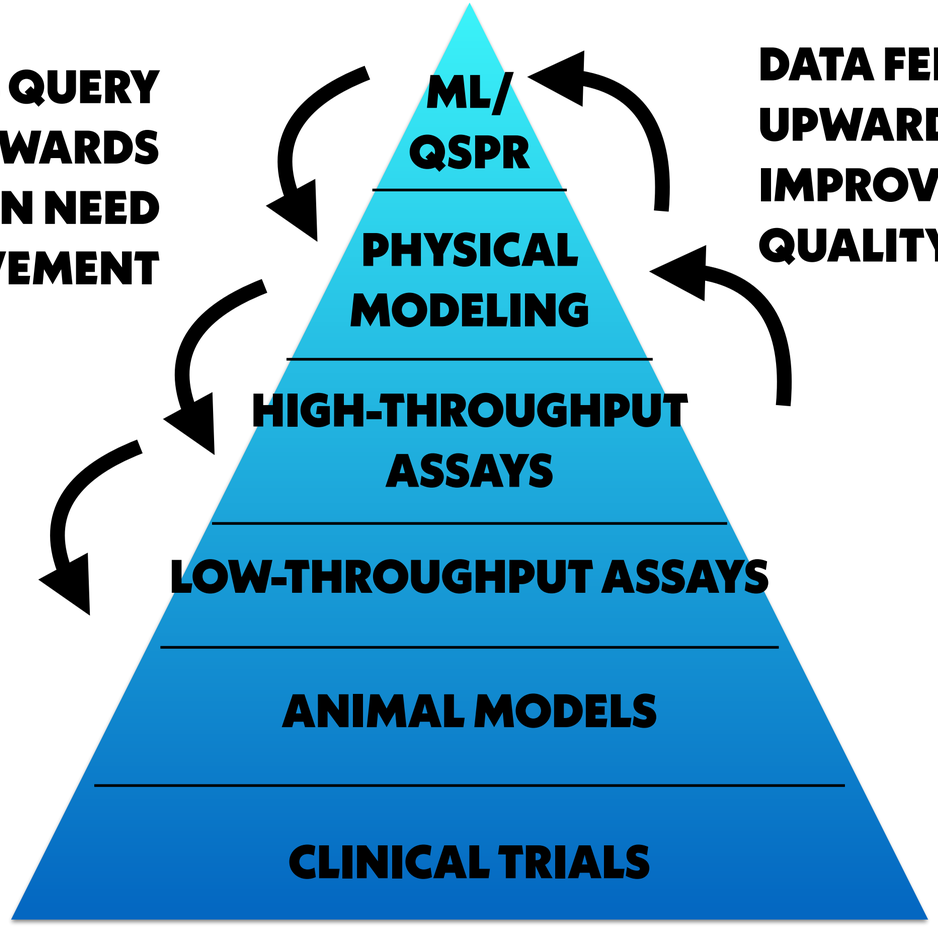
- reasoning based on probabilities of outcomes
- economic modeling of progress through cycle enables utility-driven decisions: prioritize $, time, or total probability of success
- liberates humans to study failures and develop improved models
- make rational exit decisions
- enable scientists to simultaneously field multiple projects

DRUG DISCOVERY CAN BE VIEWED AS AN AMORTIZED INFERENCE PROBLEM

MANY MOLECULES EVALUATED

FEW MOLECULES EVALUATED

MODELS QUERY DOWNWARDS WHEN IN NEED OF IMPROVEMENT



DATA FEEDS UPWARDS TO IMPROVE MODEL QUALITY

$

FAST/CHEAP

SLOW/EXPENSIVE

$$$

OK, BUT WHO WOULD BE CRAZY ENOUGH TO FUND THIS?

DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

Defense Advanced Research Projects Agency › News And Events

Accelerated Molecular Discovery

OCT. 18, 2018

Webinar

- autonomous multi-parameter optimization
- automated synthetic chemistry
- automated measurements
- automated model synthesis from experimental data

BIOLOGY IS INDUSTRIALIZED,
WHILE MEDICINAL CHEMISTRY IS STUCK IN A PRE-INDUSTRIAL STAGE

Biology

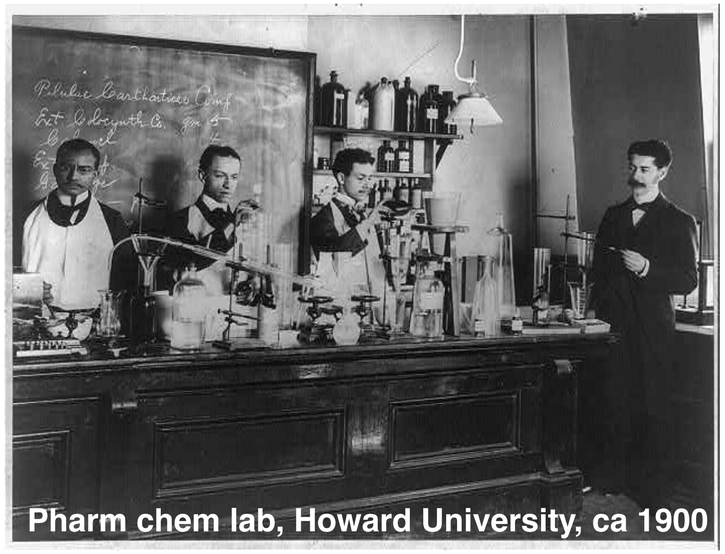
Cold Spring Harbor Laboratory, ca 1930

Recursion Pharmaceuticals

1900s

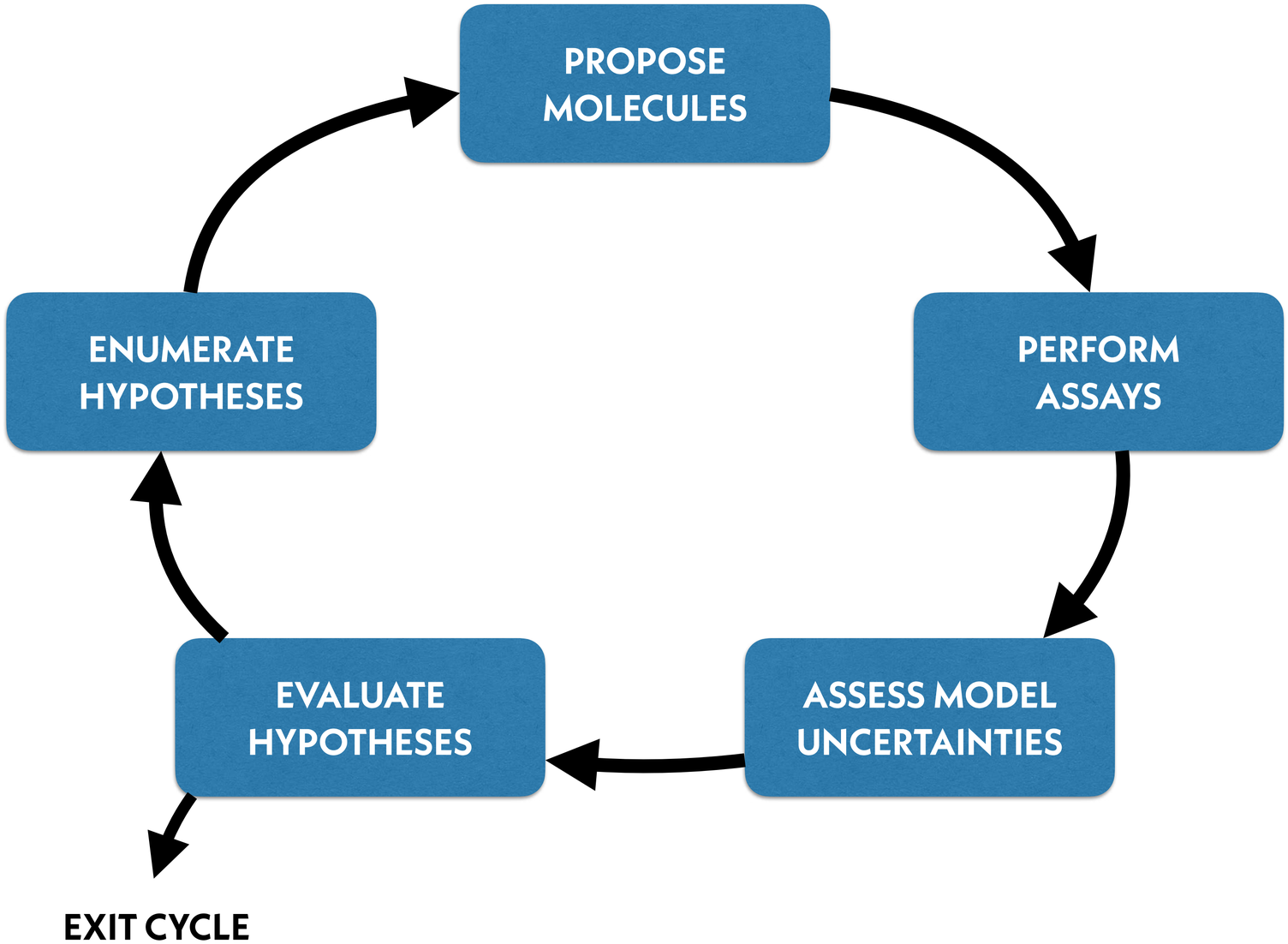
TODAY

Medicinal Chemistry



Pharm chem lab, Howard University, ca 1900

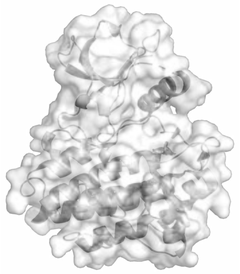
HOW WOULD AUTONOMOUS DISCOVERY WORK?

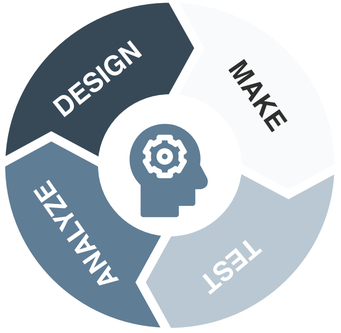


AN AUTONOMOUS DISCOVERY PLATFORM WOULD REASON THROUGH ALL POSSIBLE FUTURES TO SELECT OPTIMAL ACTIONS

Target Product Profile (TPP)

Target ID





Autonomous decision-making engine

Preclinical → Phase 1 (safety) → Phase 2 (efficacy) → Phase 3 (effectiveness)

$0M
$62M
11.1 years

Preclinical → Phase 1 (safety) → Phase 2 (efficacy)

Preclinical → Phase 1 (safety) → Phase 2 (efficacy) → Phase 3 (effectiveness) → NDA (new drug approval)

$1.6B
$43M
7.4 y

$0M
$3.2M
2.8 years

Preclinical → Phase 1 (safety)

Preclinical → Phase 1 (safety) → Phase 2 (efficacy) → Phase 3 (effectiveness) → NDA (new drug approval)

$258M
$112M
9.3 y

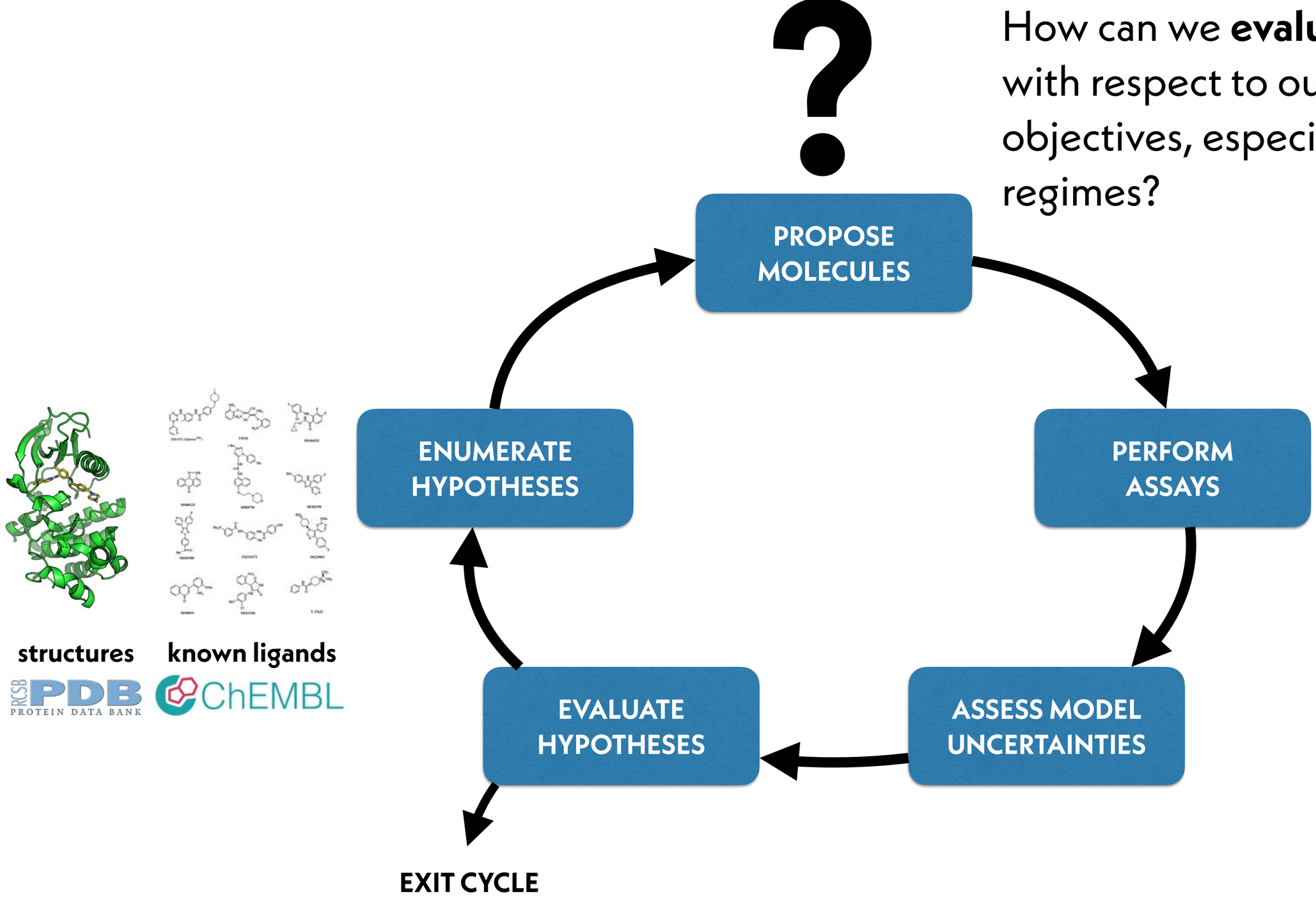
Preclinical → Phase 1 (safety)

SIMULATE ALL POSSIBLE FUTURES - - - - - - - - - - ->

DECISION POINT WITH OPTIONS

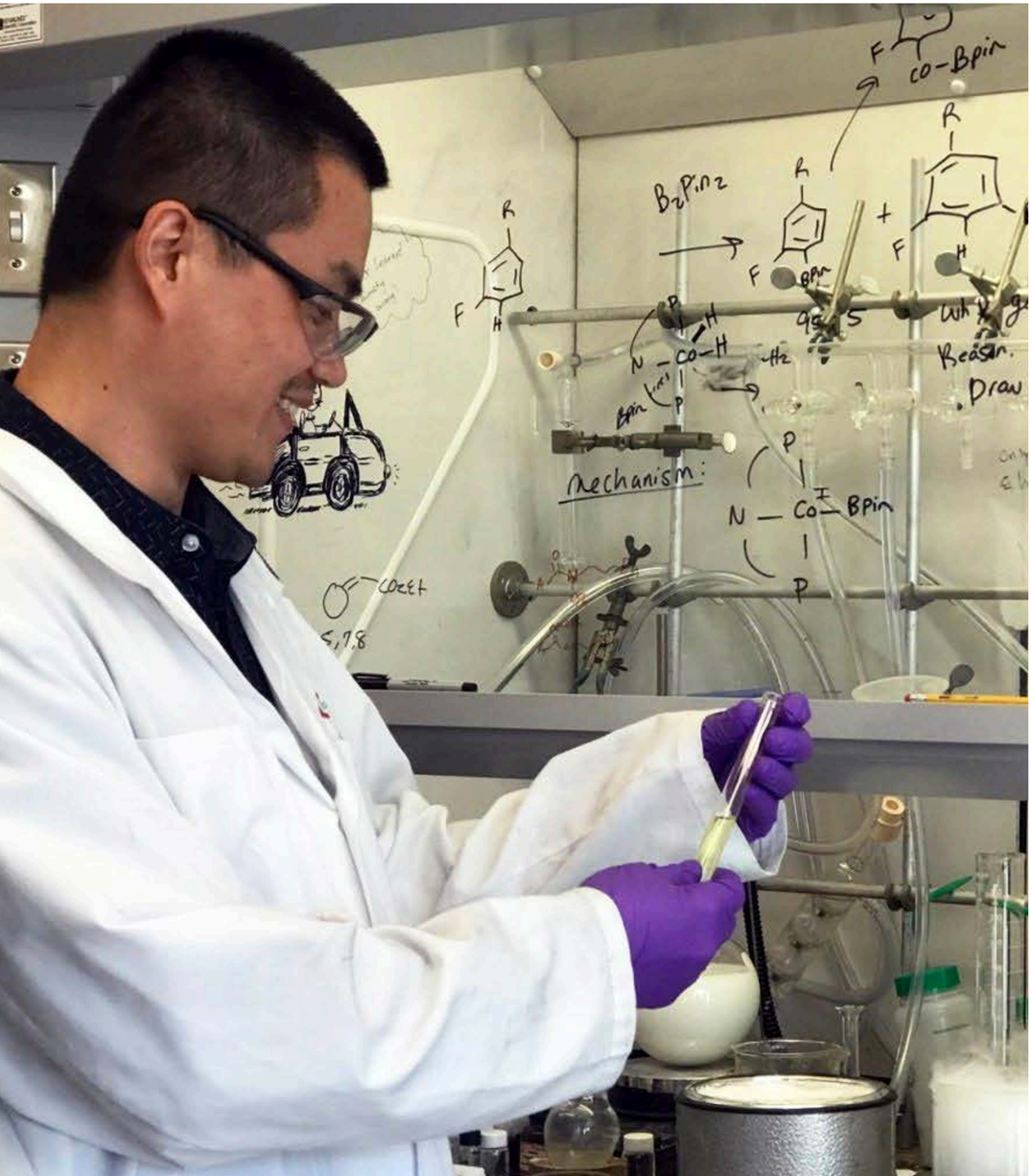
value
total cost
duration

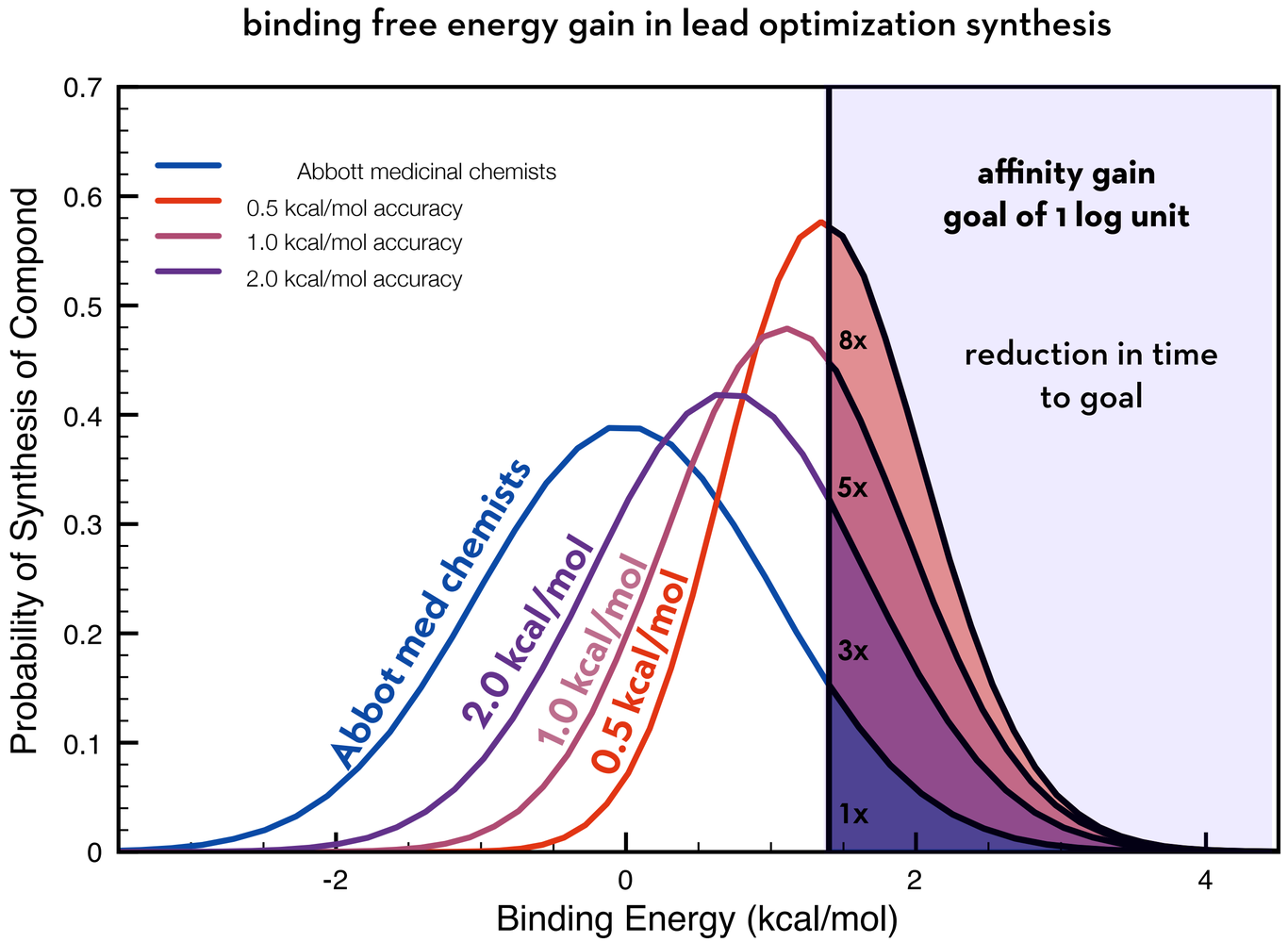
How can we **evaluate** molecules with respect to our design objectives, especially in data-poor regimes?



Several excellent graphics taken from http://www.optibrium.com/

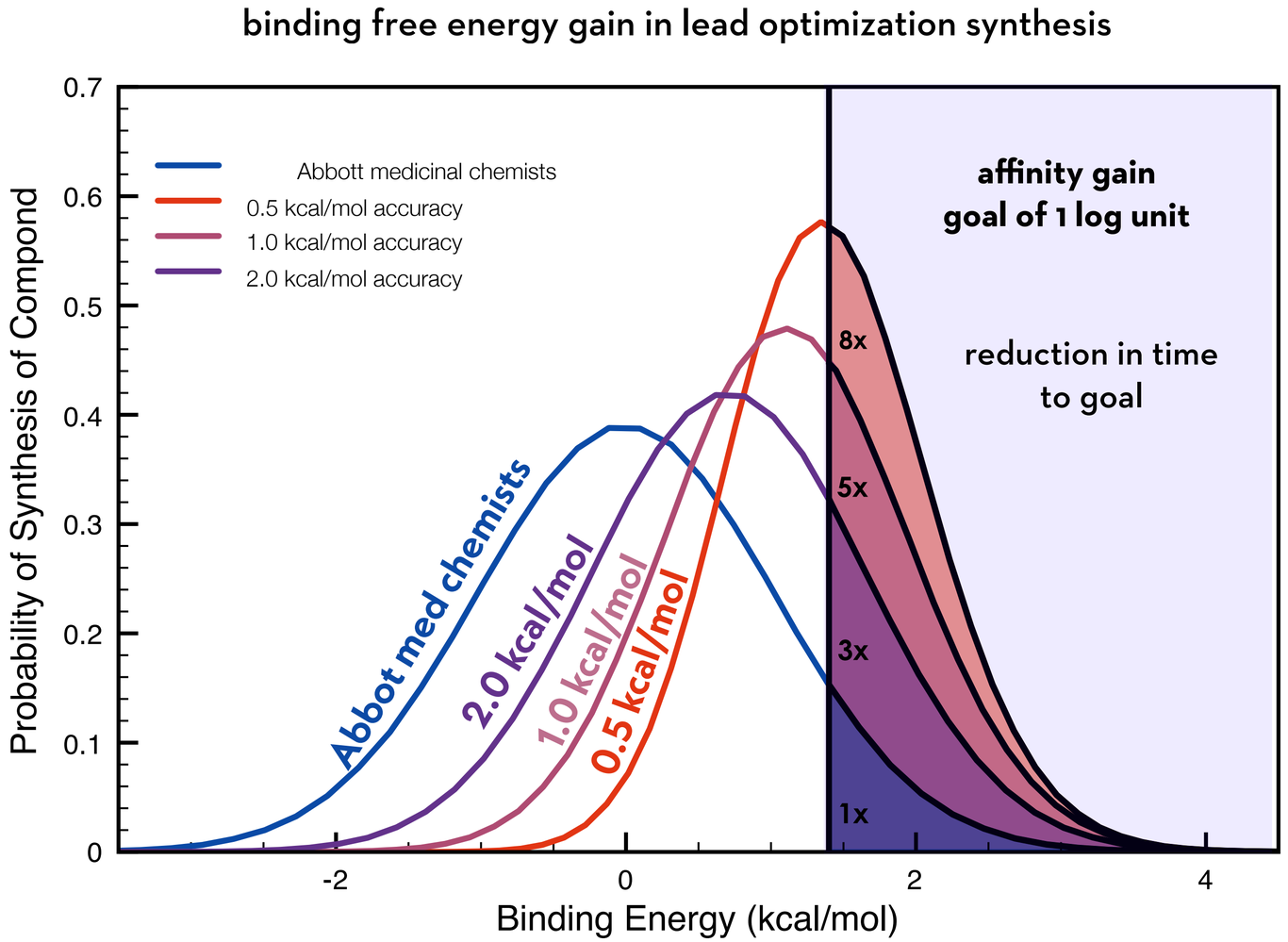
HOW ACCURATE DO PREDICTIONS NEED TO BE TO HAVE IMPACT ON DISCOVERY?





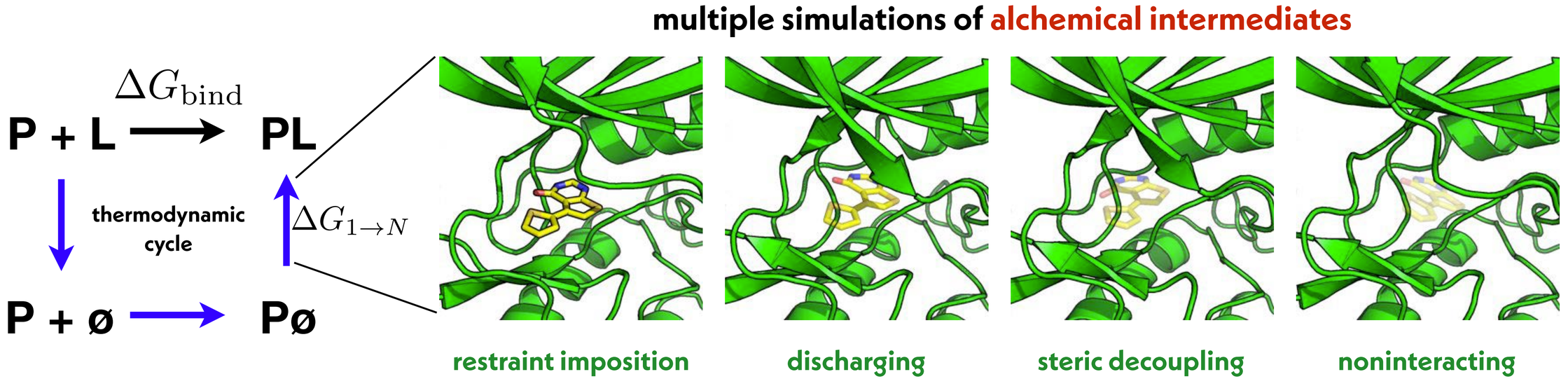
M. R. Shirts, D. L. Mobley and Scott P. Brown. "Free energy calculations in structure-based drug design", in Drug Design: Structure- and Ligand-Based Approaches, pgs. 61-86, 2010.

CYBORG CHEMISTS WILL DEFEAT HUMAN CHEMISTS



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.

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



Includes all contributions from enthalpy and entropy of binding to a flexible receptor

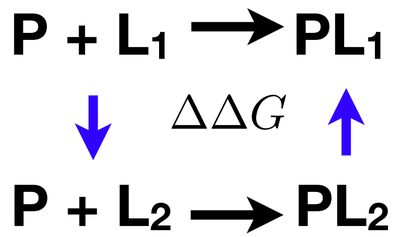
$$\Delta G_{1\to N} = -\beta^{-1} \ln \frac{Z_N}{Z_1} = -\beta^{-1} \ln \frac{Z_2}{Z_1} \cdot \frac{Z_3}{Z_2} \cdots \frac{Z_N}{Z_{N-1}}$$

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

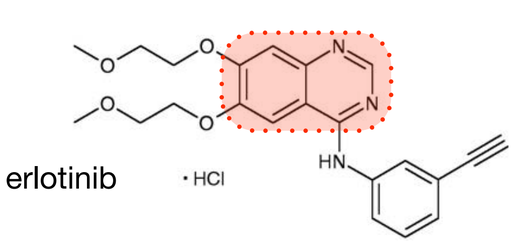
Pioneering work from many: McCammon, van Gunsteren, Kollman, Jorgensen, Chipot, Roux, Boresch, Fujitani, Pande, Shirts, Swope, Christ, Mobley, Schrödinger, and many more

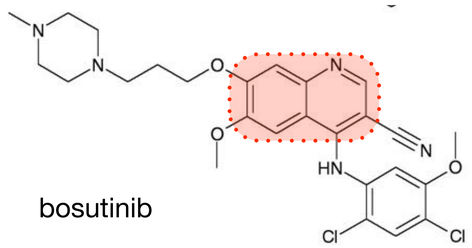
ALCHEMICAL FREE ENERGY CALCULATIONS COME IN TWO FLAVORS: RELATIVE AND ABSOLUTE

RELATIVE



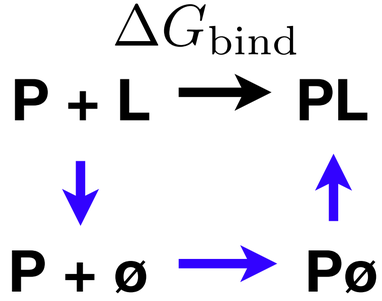
capable of **transforming a few atoms**
good for comparing **similar ligands**
requires same or **similar scaffolds**
requires **common scaffold to anchor series**



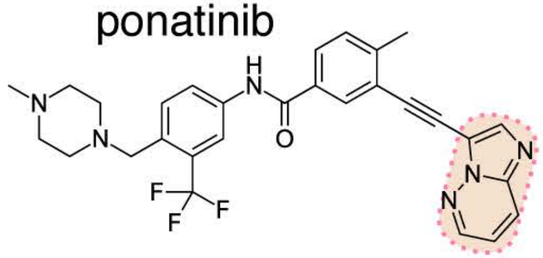


Cournia, Allen, Sherman 2017: http://dx.doi.org/10.1021/acs.jcim.7b00564

ABSOLUTE



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

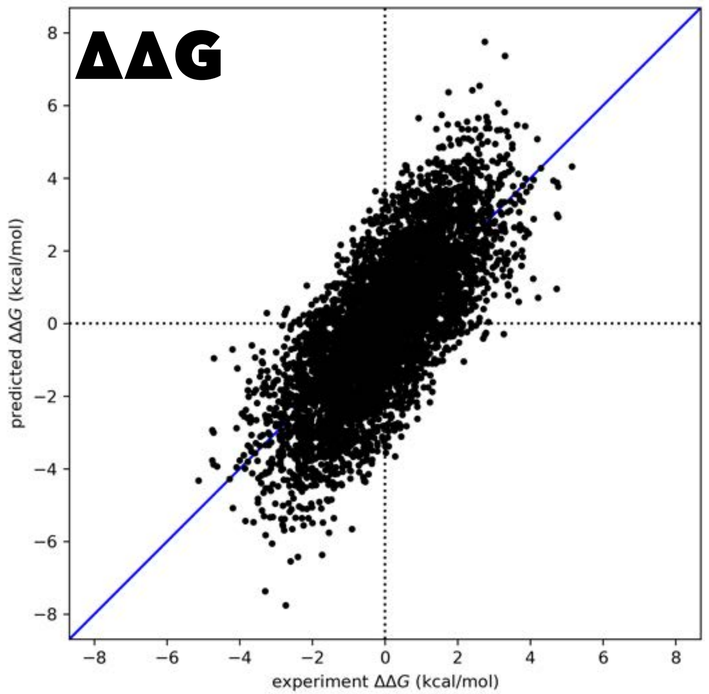


Aldeghi, Bluck, Biggin 2018: https://doi.org/10.1007/978-1-4939-7756-7_11

USEFUL ACCURACY IS SOMETIMES ACHIEVABLE

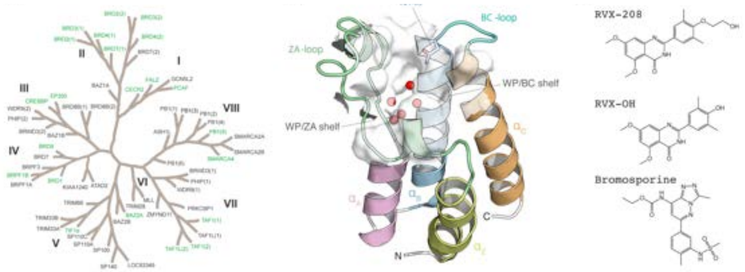
RELATIVE

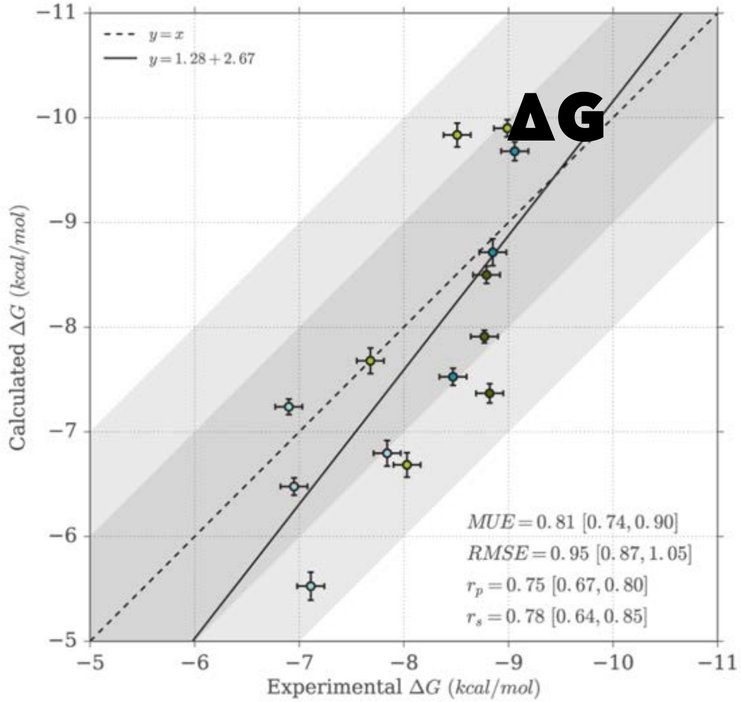
```
                all within-target pairs ΔΔG (N = 5620)
RMSE: OPLS      1.37 [95%:  1.34,  1.39] kcal/mol
MUE : OPLS      1.09 [95%:  1.07,  1.11] kcal/mol
R2  : OPLS      0.10 [95%:  0.06,  0.15] kcal/mol
rho : OPLS      0.73 [95%:  0.72,  0.74] kcal/mol
```



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

ABSOLUTE





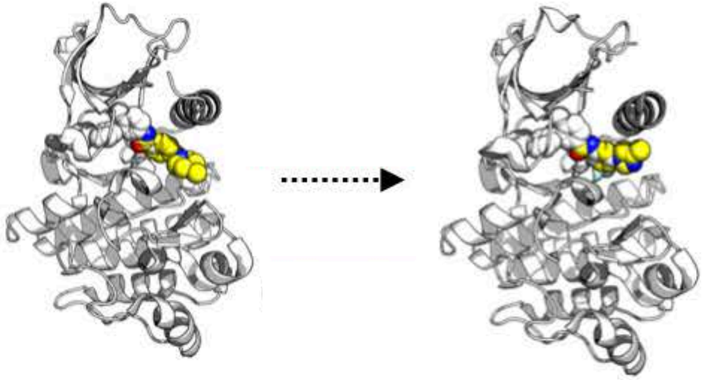
Aldeghi et al. JACS 139:946, 2017.
https://doi.org/10.1021/jacs.6b11467

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

***best-case scenarios!**

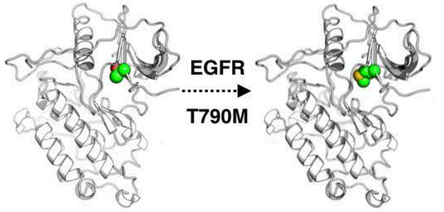
ALCHEMICAL FREE ENERGY CALCULATIONS HAVE THE POTENTIAL TO COMPUTE MULTIPLE PROPERTIES OF INTEREST

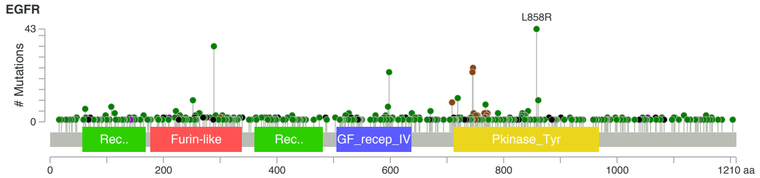
driving affinity / potency



driving selectivity

Moraca, Negri, de Olivera, Abel JCIM 2019
https://doi.org/10.1021/acs.jcim.9b00106
Aldeghi et al. JACS 139:946, 2017.
https://doi.org/10.1021/jacs.6b11467



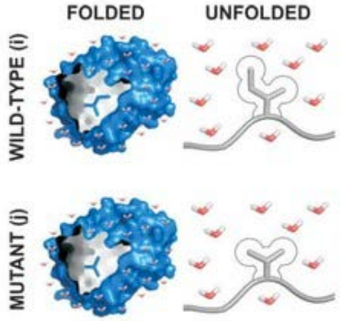


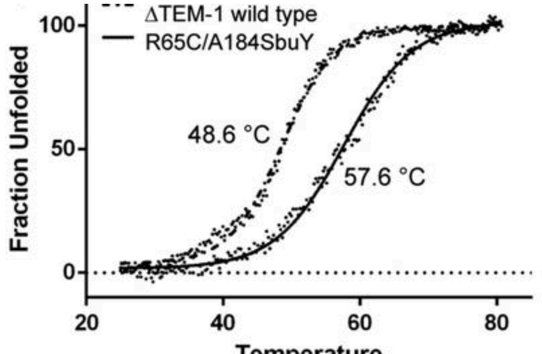
predicting clinical drug resistance/sensitivity

Hauser, Negron, Albanese, Ray, Steinbrecher, Abel, Chodera, Wang.
Communications Biology 1:70, 2018
https://doi.org/10.1038/s42003-018-0075-x
Aldeghi, Gapsys, de Groot. ACS Central Science 4:1708, 2018
https://doi.org/10.1021/acscentsci.8b00717

optimizing thermostability

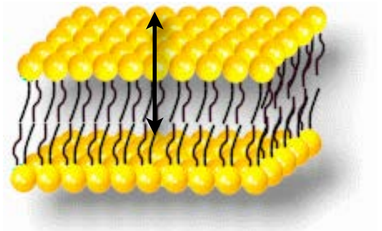
Gapsys, Michielssens, Seeliger, and de Groot. Angew Chem 55:7364, 2016
https://doi.org/10.1002/anie.201510054



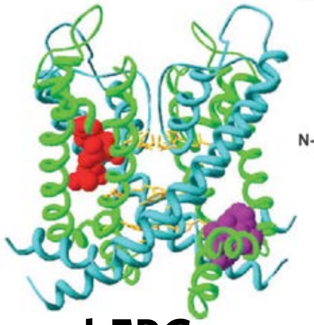


ALCHEMICAL FREE ENERGY CALCULATIONS HAVE THE POTENTIAL TO COMPUTE MULTIPLE PROPERTIES OF INTEREST

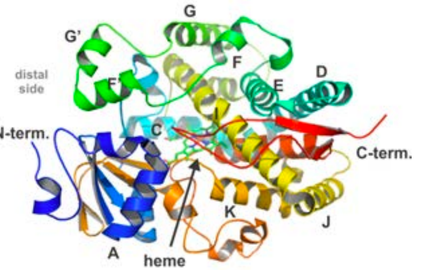
partition coefficients (logP, logD) and permeabilities



structure-enabled ADME/Tox targets

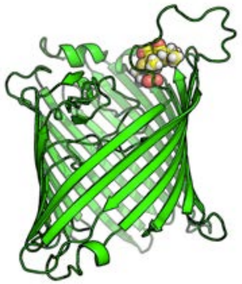


hERG

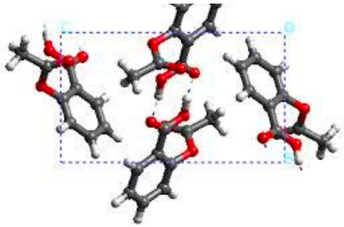


CYP3A4

porin permeation



crystal polymorphs, etc.



YANK: AN OPEN-SOURCE, COMMUNITY-ORIENTED PLATFORM FOR GPU-ACCELERATED ABSOLUTE FREE ENERGY CALCULATIONS

OpenMM

YANK

NVIDIA GTX-1080TI ($700)
11 TFLOP/S SINGLE PRECISION

OpenMM performance on a GTX 1080 Ti

method	natoms	AMBER GPU	OpenMM GPU
GB/SA	2,489	N/A	**902 ns/day**
RF	23,558	N/A	**577 ns/day**
PME	23,558	304 ns/day	**377 ns/day**

http://openmm.org OpenMM 7.4.0 benchmark
AMBER benchmarks from https://ambermd.org/gpus16/benchmarks.htm

http://www.getyank.org

http://openmm.org

YANK

A GPU-accelerated Python framework for exploring algorithms for alchemical free energy calculations

Note

YANK is now in Early Access for its 1.0 release! YAML syntax should be fully operational while we prepare the underlying Python API. The program has not yet been extensively validated. Use at your own risk!

COLLABORATORS:

DAVID MOBLEY

MICHAEL SHIRTS

PETER EASTMAN

A **free, open-source**, extensible platform
for best-practices free energy calculations and ligand design

LEVI NADEN

MERCK

ANDREA RIZZI

YANK ALLOWS FLEXIBLE LEVELS OF CONTROL

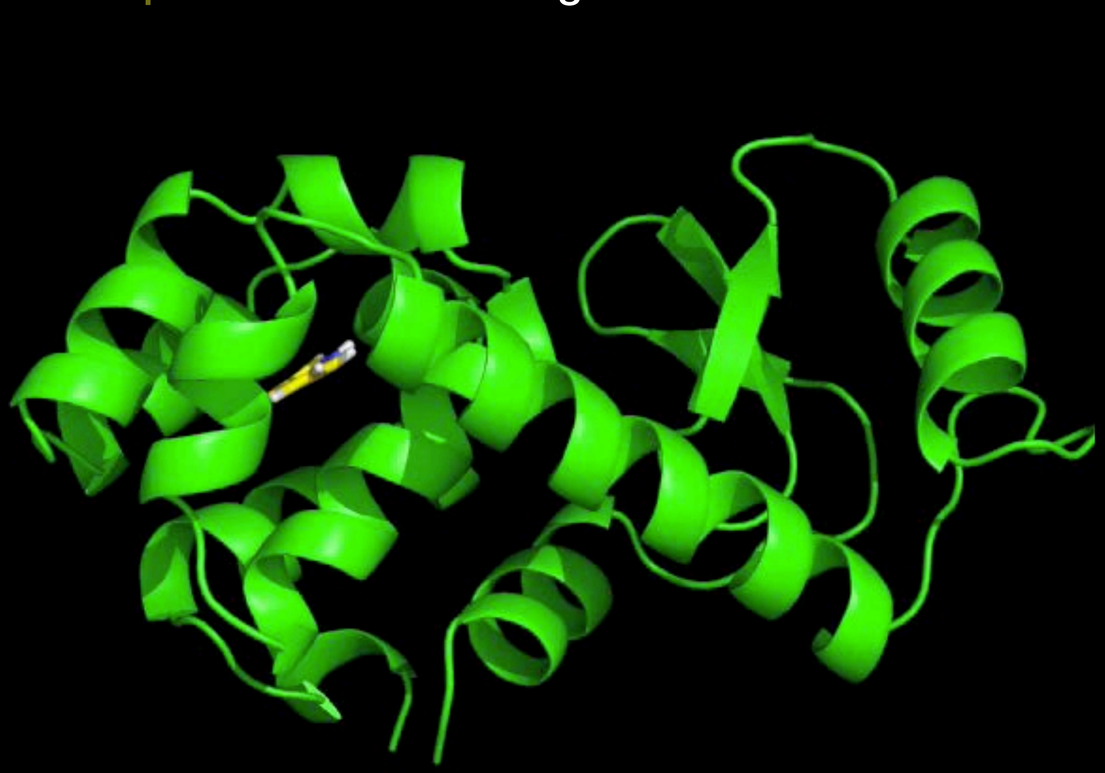
fully automated
according to best practices

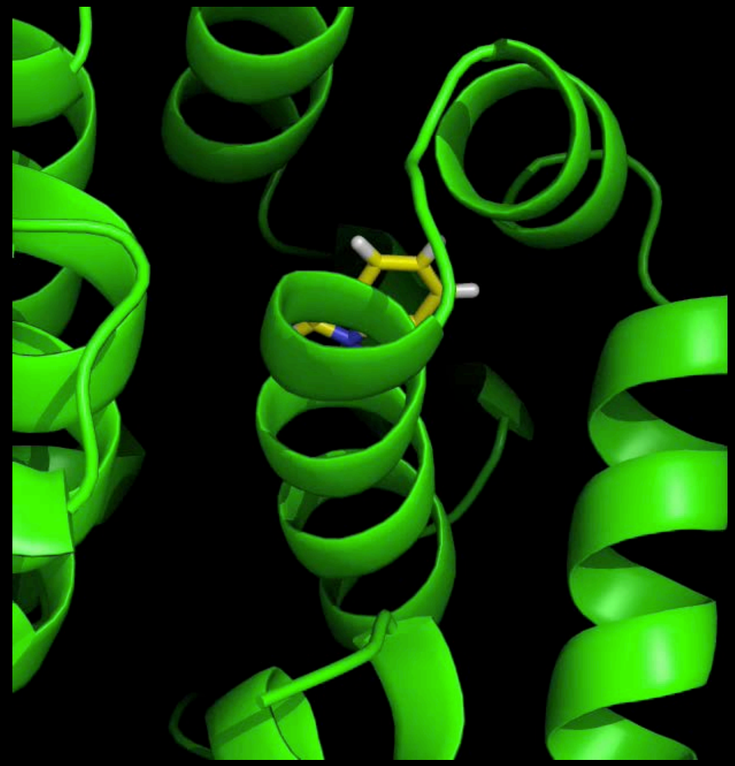
full manual control
fly yourself into the ground

automated evaluation
benchmark all the things

HAMILTONIAN EXCHANGE ALLOWS FOR RAPID DECORRELATION BY EXCHANGE BETWEEN ALCHEMICAL STATES

solid fully interacting
transparent noninteracting



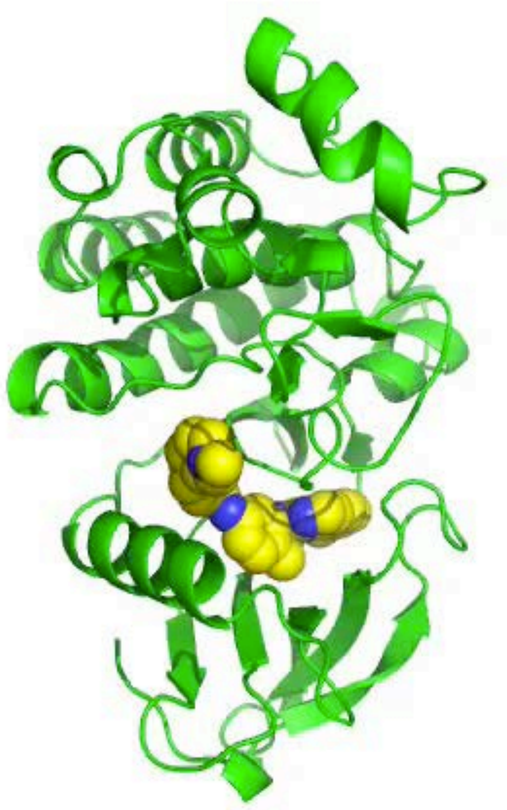


getyank.org

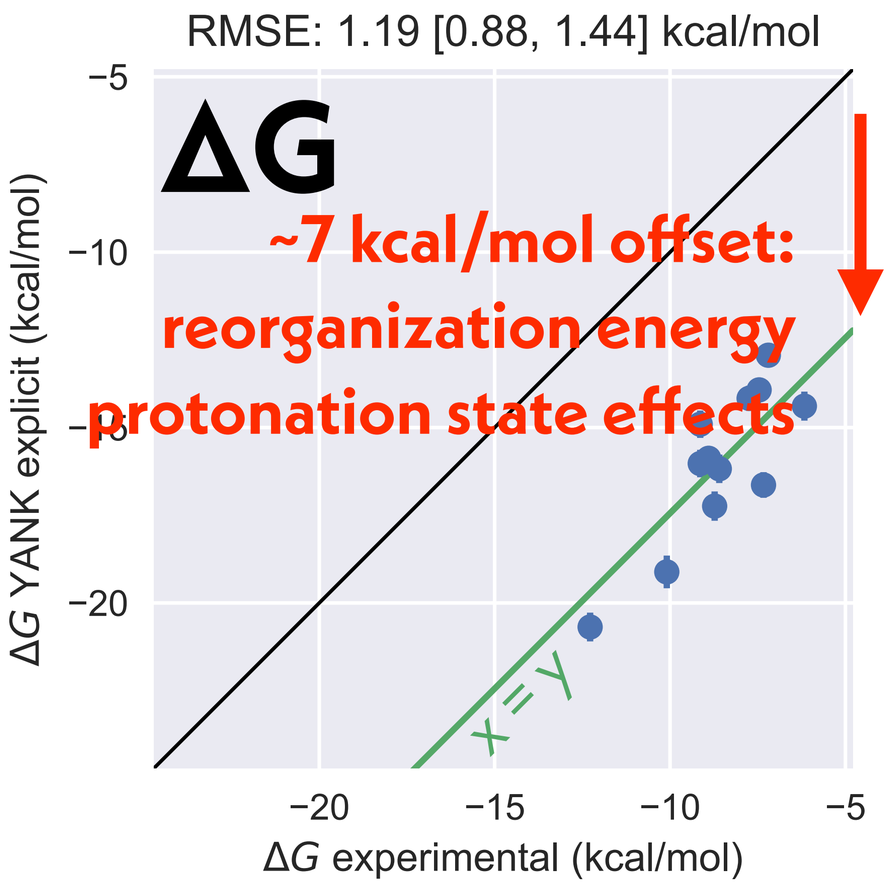
indole binding to T4 lysozyme L99A in GBSA
Hamiltonian exchange with Gibbs sampling

Chodera and Shirts. JCP 135:194110, 2011
Wang, Chodera, Yang, and Shirts. JCAMD 27:989, 2013.
http://github.org/choderalab/yank

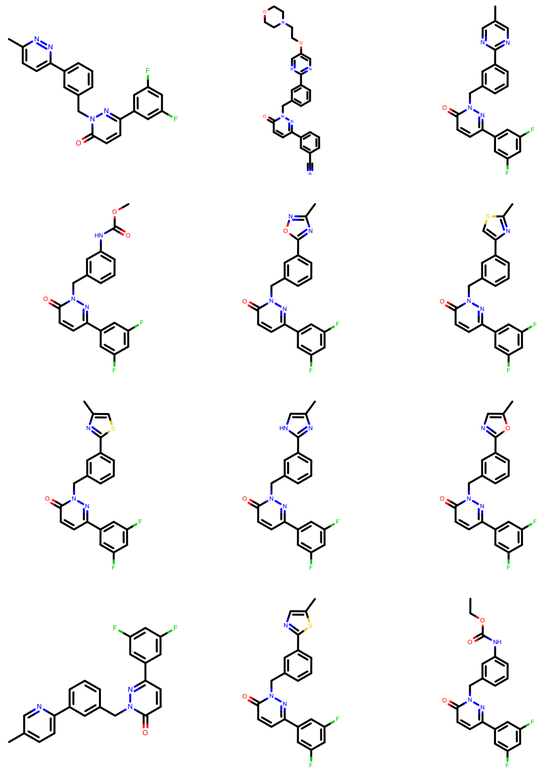
ALCHEMICAL FREE ENERGY CALCULATIONS COULD HELP PRIORITIZE LIGAND SYNTHESIS



c-Met inhibitors from Bioorg. & Med Chem Lett. 25:1597, 2015
https://github.com/choderalab/yank-benchmark



RMSE ~ 1.2 kcal/mol

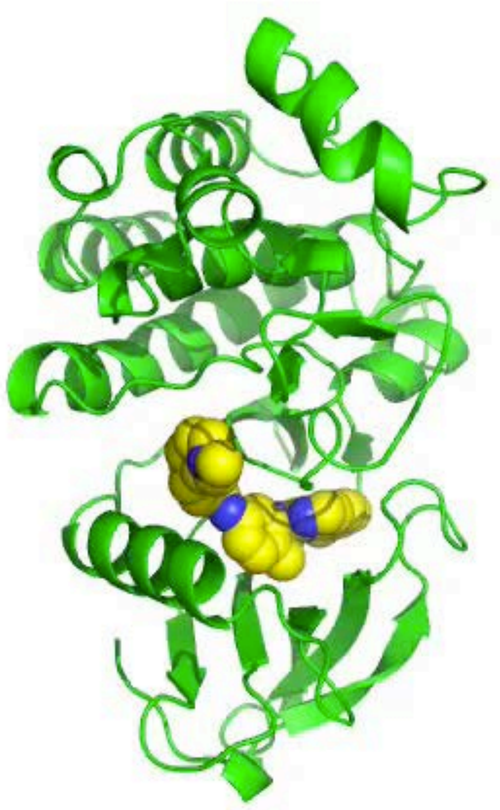


LEVI NADEN
ANDREA RIZZI

MERCK

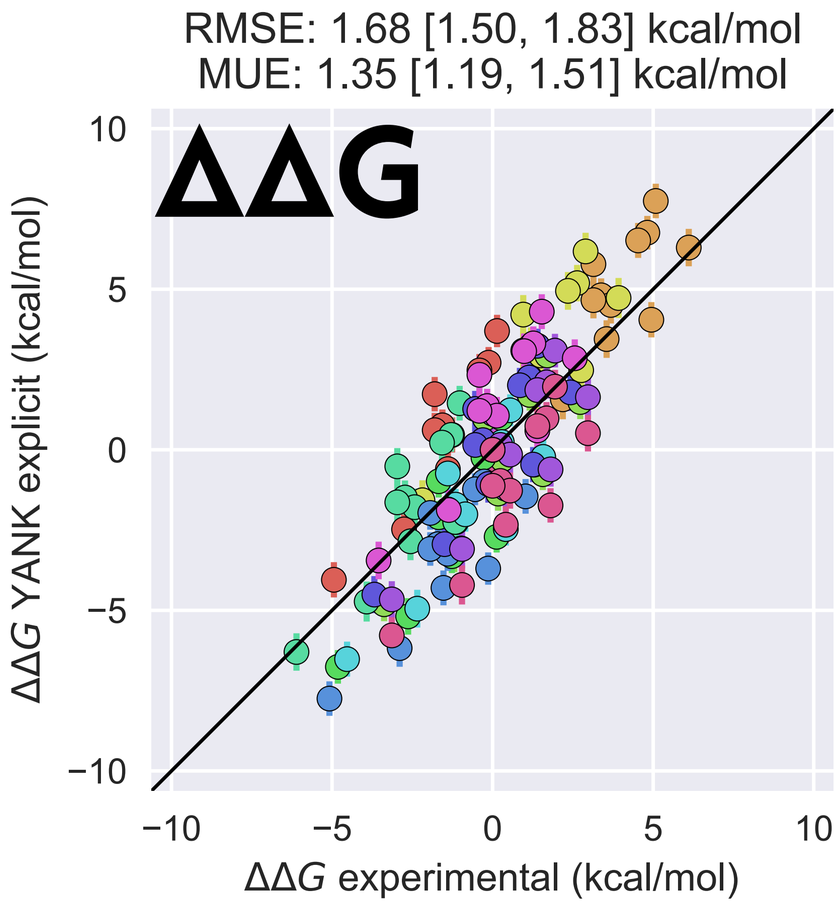
PAUL CZODROWSKI
DANIEL KUHN

ALCHEMICAL FREE ENERGY CALCULATIONS COULD HELP PRIORITIZE LIGAND SYNTHESIS



c-Met inhibitors from Bioorg. & Med Chem Lett. 25:1597, 2015

https://github.com/choderalab/yank-benchmark

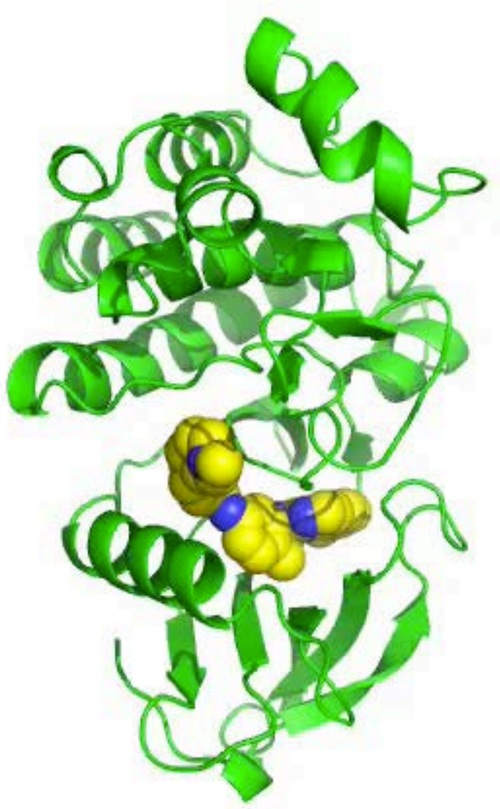


LEVI NADEN
ANDREA RIZZI

MERCK

PAUL CZODROWSKI
DANIEL KUHN

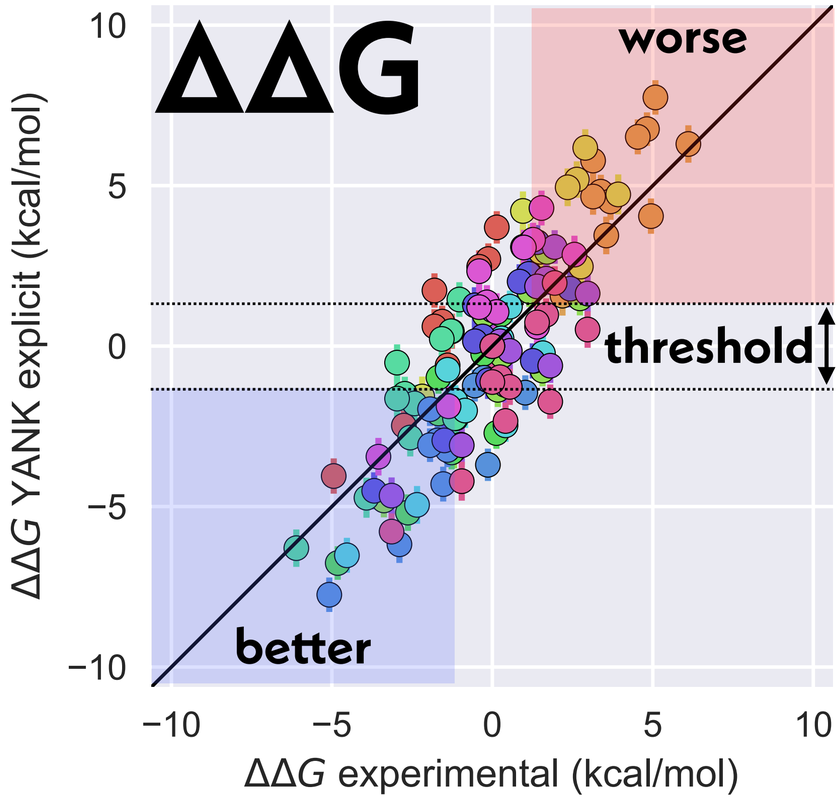
ALCHEMICAL FREE ENERGY CALCULATIONS COULD HELP PRIORITIZE LIGAND SYNTHESIS



c-Met inhibitors from Bioorg. & Med Chem Lett. 25:1597, 2015

https://github.com/choderalab/yank-benchmark

RMSE: 1.68 [1.50, 1.83] kcal/mol
MUE: 1.35 [1.19, 1.51] kcal/mol



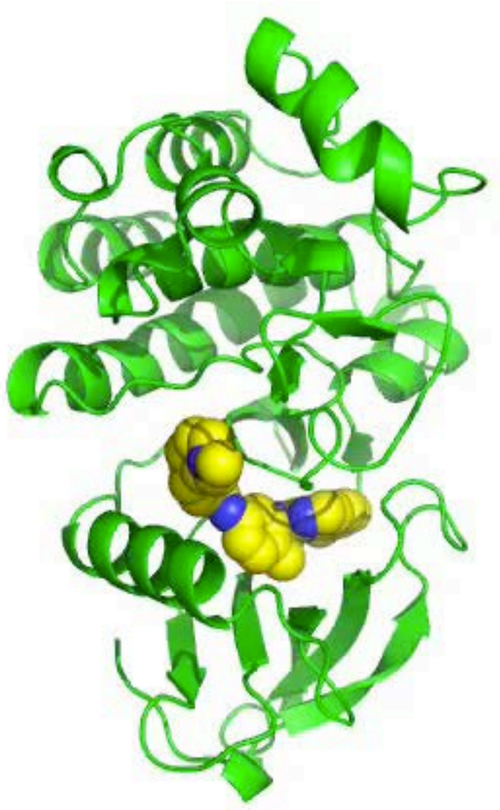
How often can this help us make the right decision about which molecules to synthesize?

LEVI NADEN
ANDREA RIZZI

MERCK

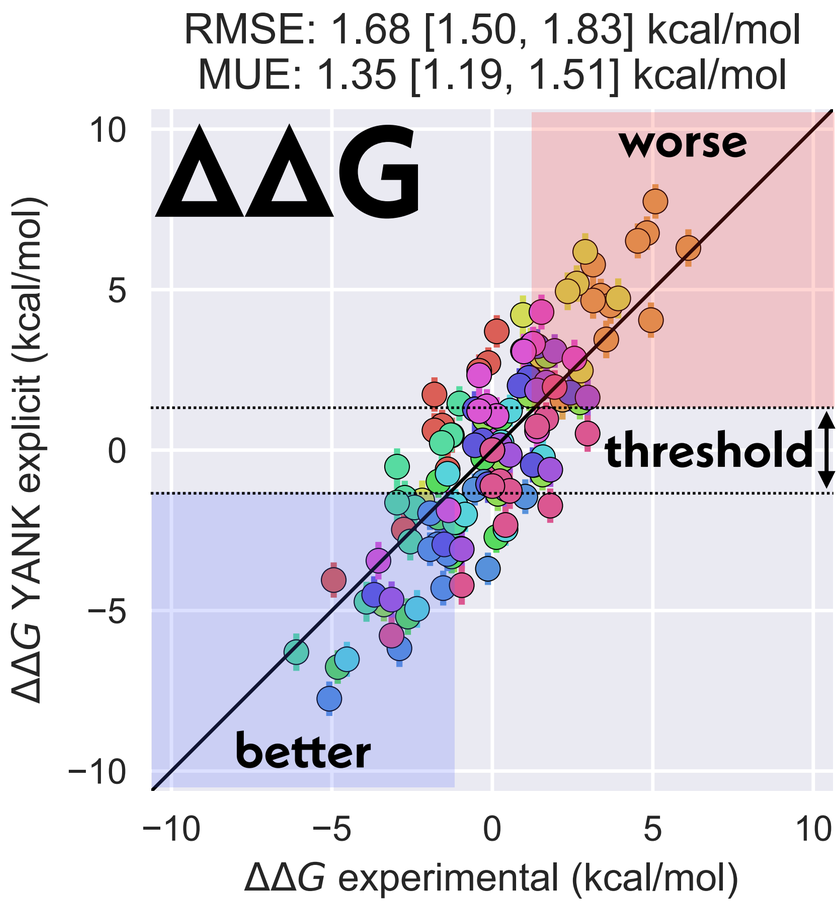
PAUL CZODROWSKI
DANIEL KUHN

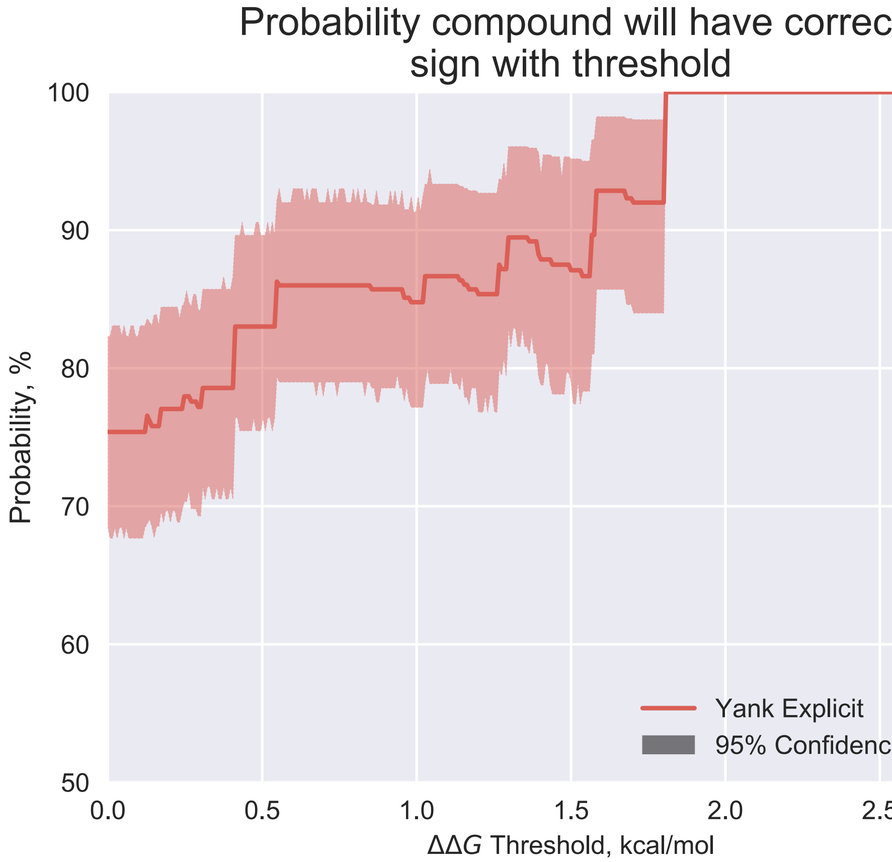
ALCHEMICAL FREE ENERGY CALCULATIONS COULD HELP PRIORITIZE LIGAND SYNTHESIS



c-Met inhibitors from Bioorg. & Med Chem Lett. 25:1597, 2015

https://github.com/choderalab/yank-benchmark



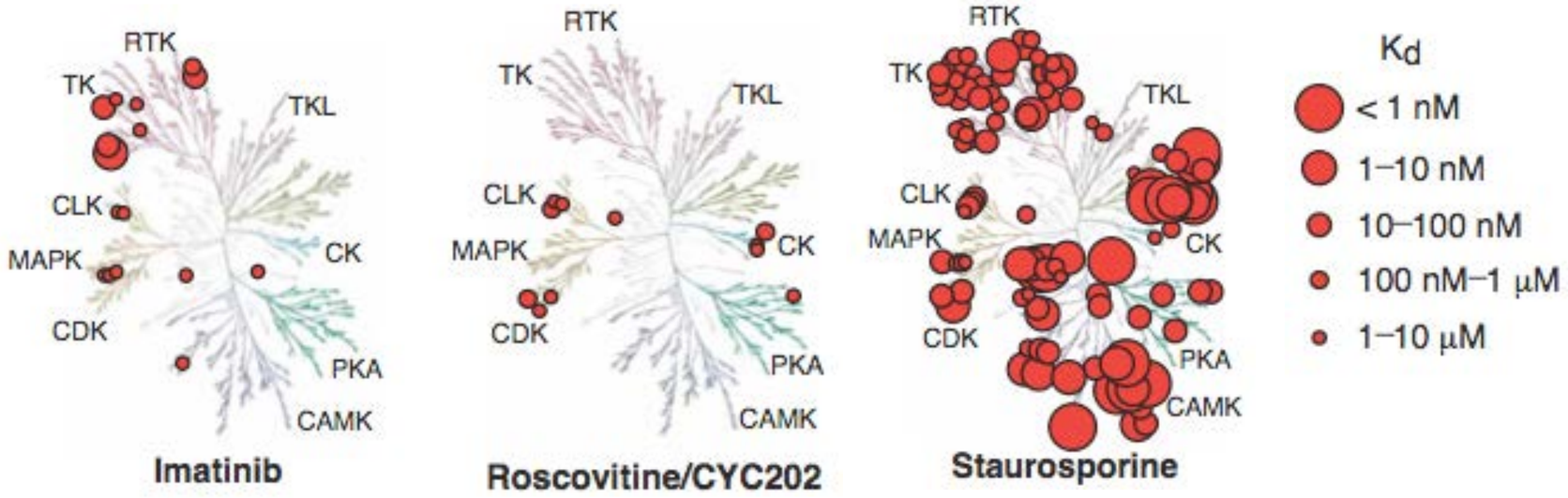


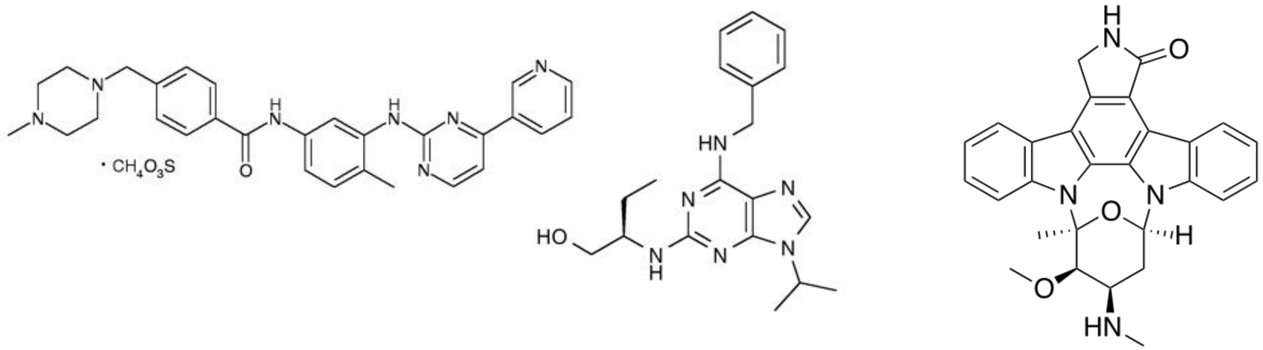
LEVI NADEN
ANDREA RIZZI

MERCK

PAUL CZODROWSKI
DANIEL KUHN

OK, SO WE CAN COMPUTE AFFINITIES. WHAT ABOUT SELECTIVITIES? ISN'T THAT MUCH HARDER?

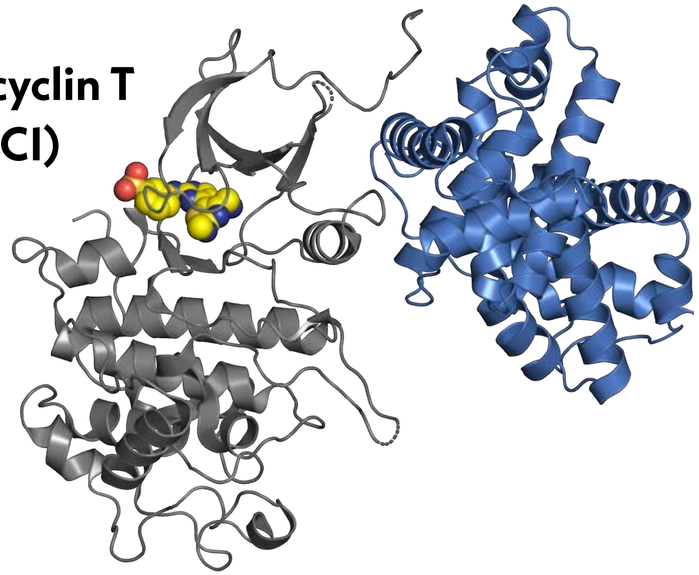




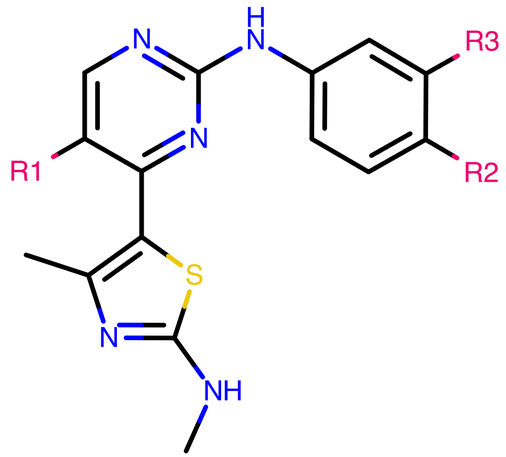
Nature Biotech 26:127, 2008

LET'S LOOK AT SOME REAL SELECTIVITY DATA

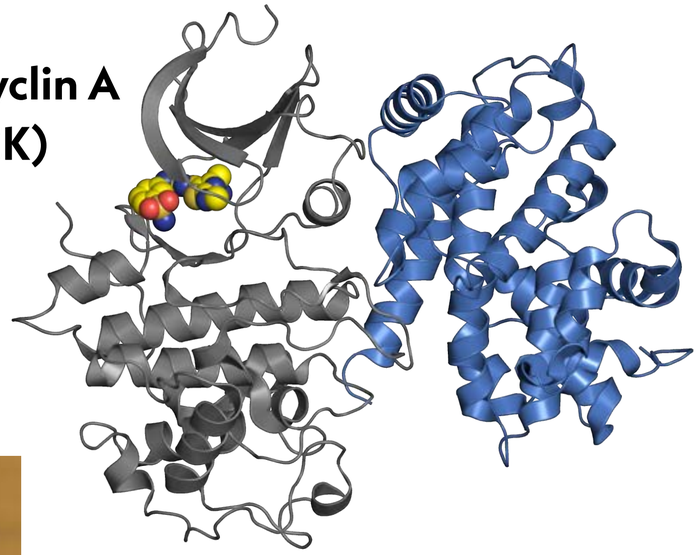
CDK9/cyclin T (4BCI)



inhibition **reinstates apoptosis** in cancer cells

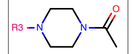
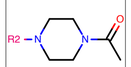
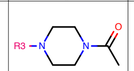
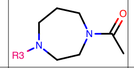


CDK2/cyclin A (4BCK)



essential for S-phase progression

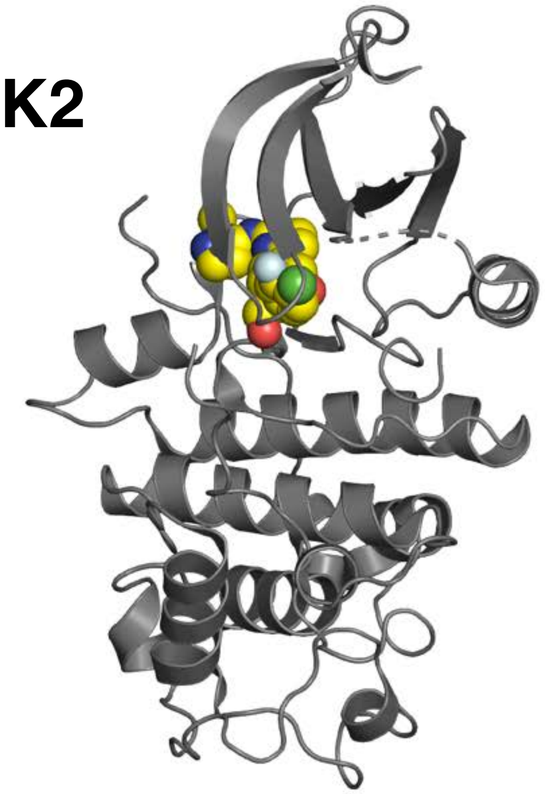
STEVEN ALBANESE

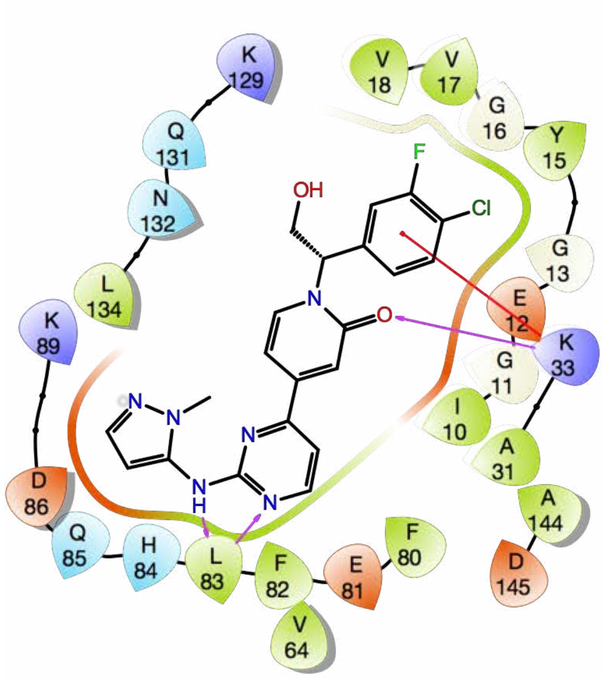
Ligand	R1	R2	R3	ΔG CDK2 (kcal/mol)	ΔG CDK9 (kcal/mol)	ΔΔG (kcal/mol)
12a	CN	H		-12.27	-11.21	-1.64
12b	OH	H		-7.23	-8.22	-1.57
12c	CN	H		-11.45	-11.21	-1.57
12e	F	H		-11.62	-11.45	-1.57
12f	Cl	H		-10.91	-10.85	-2.36
12g	Methyl	H		-10.18	-11.32	-1.97
12h	Ethyl	H		-8.28	-9.56	-2.37
12j	CN	H		-10.04	-11.12	-1.56
12l	CN		H	-10.34	-10.44	-1.34
12n	CN	H		-10.06	-10.97	-2.47
12o	F	H		-10.06	-11.12	-0.75
12q	F	H		-10.91	-11.62	-2.31
12t	CN	H		-9.38	-11.12	-1.91
1a	H	H		-11.62	-11.86	-2.77
1b	H	H		-11.45	-11.86	-1.77

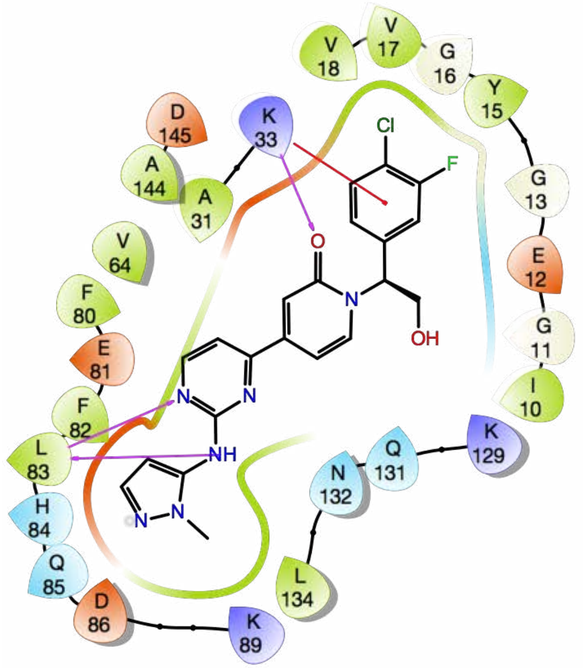
Shao et *al.*, J Med Chem 56(3), 640–659

HOW MUCH DOES CANCELLATION OF ERROR BETWEEN SIMILAR BINDING SITES HELP SELECTIVITY PREDICTION?

CDK2

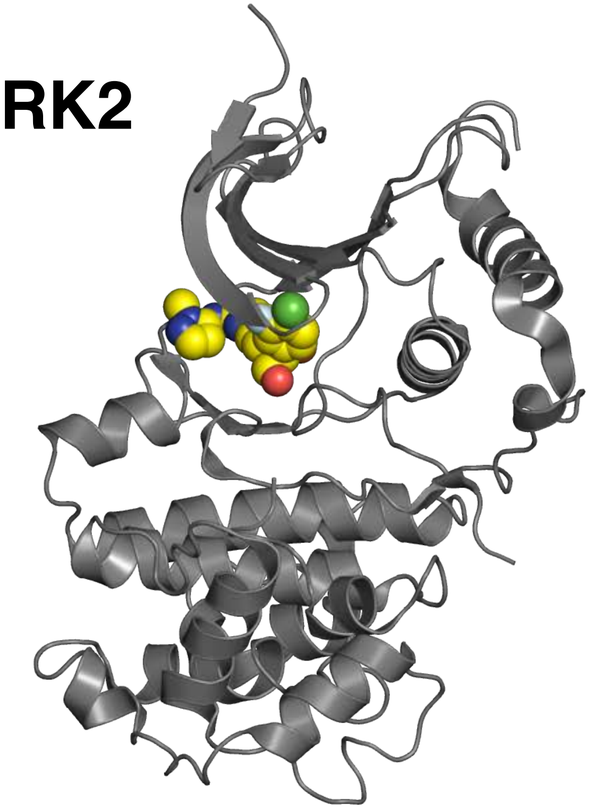






Charged (negative) Charged (positive) Glycine Hydrophobic Polar H-bond Pi-cation Solvent exposure

ERK2



Quantify via the **correlation coefficient**

$$\rho \equiv \frac{\mathrm{cov}(\epsilon_1, \epsilon_2)}{\sqrt{\mathrm{var}(\epsilon_1)\mathrm{var}(\epsilon_2)}}$$

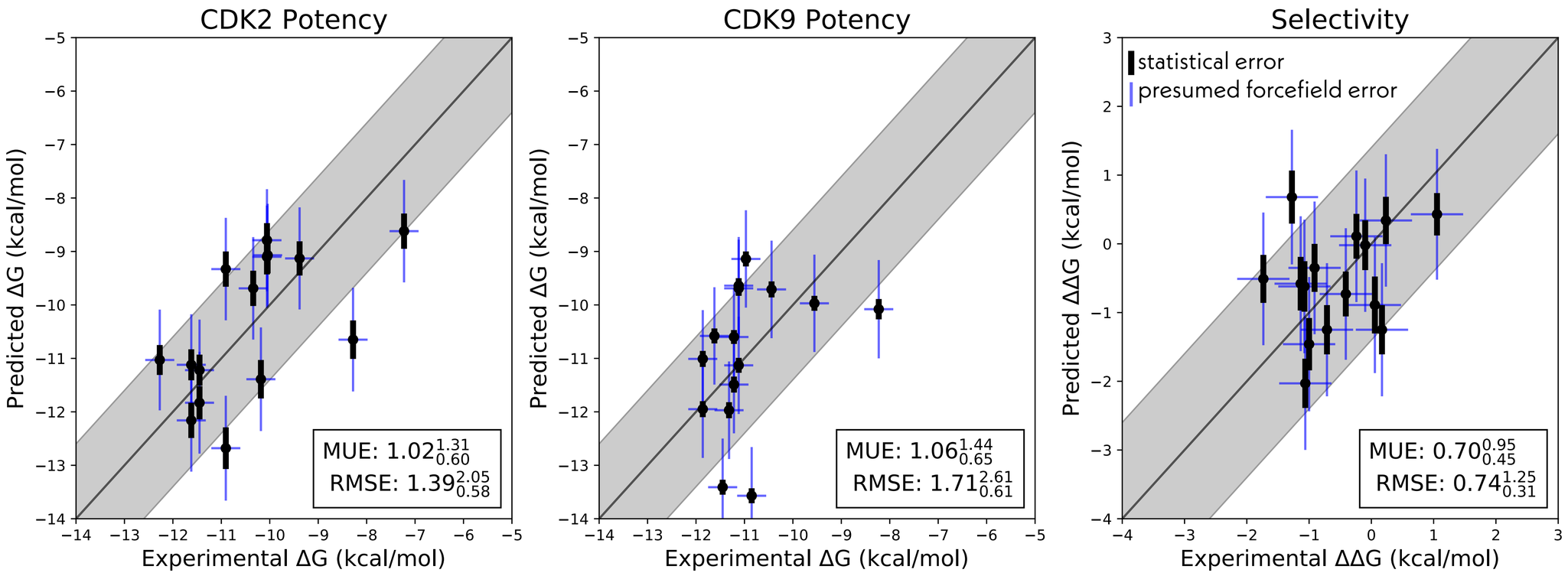
of the **error**

$$\epsilon_* \equiv \Delta\Delta G_*^{\mathrm{FEP}} - \Delta\Delta G_*^{\mathrm{exp}}$$

ALCHEMICAL METHODS CAN ACCURATELY PREDICT BINDING AFFINITIES TO INDIVIDUAL CDKS

ΔG

ΔΔG (CDK9 - CDK2)



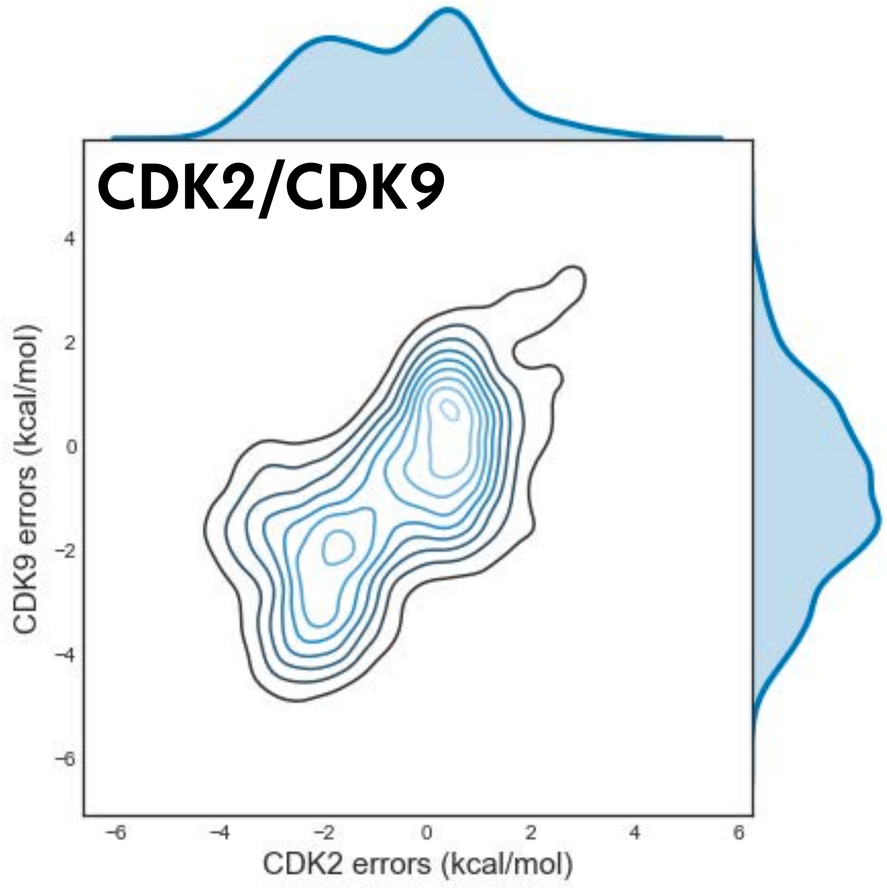
STEVEN ALBANESE

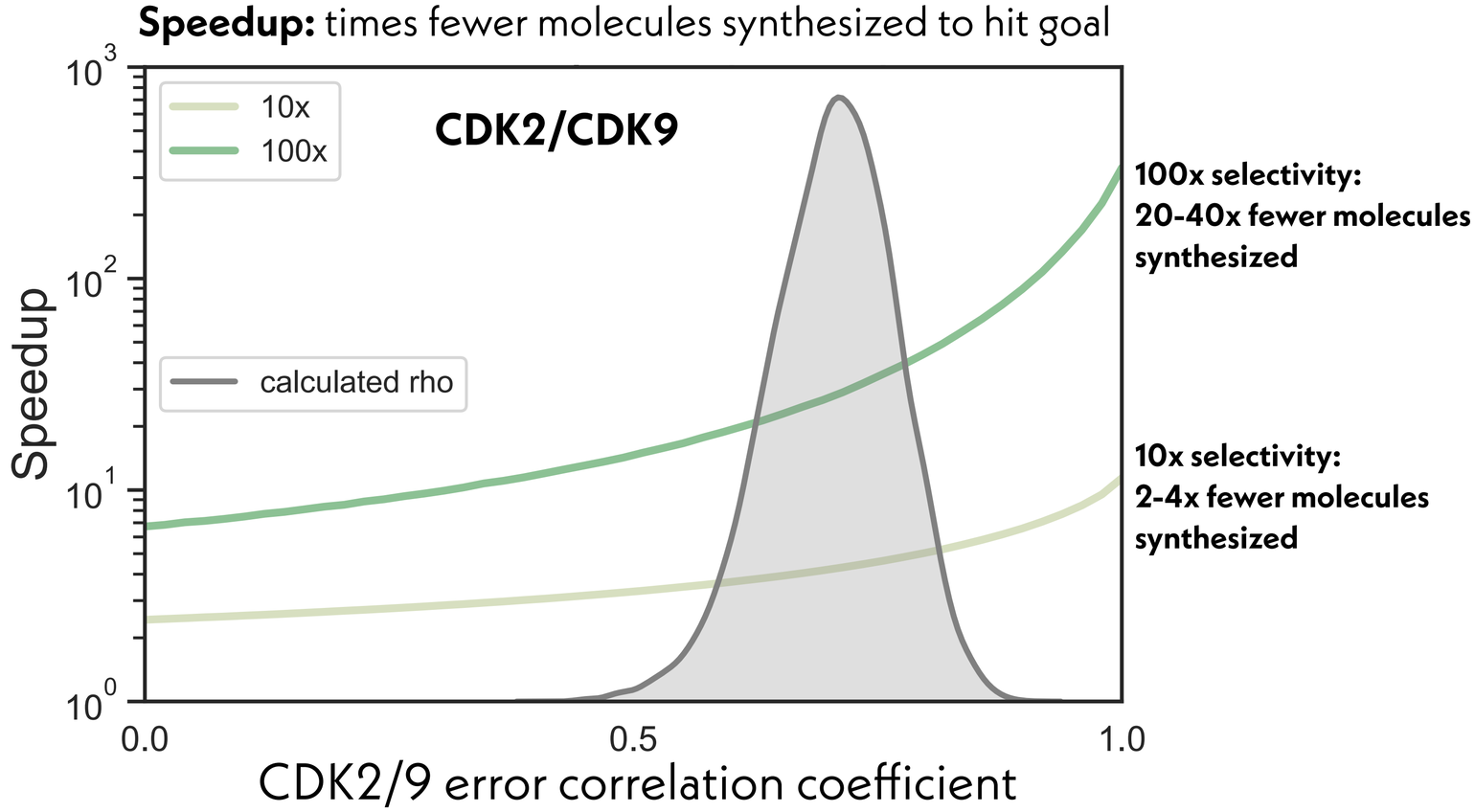
Individual affinities predicted confidently, but what does this mean for selectivity?

FEP+/OPLS3

LINGLE WANG
SCHRÖDINGER

HOW MUCH CAN FREE ENERGY CALCULATIONS ACCELERATE SELECTIVE INHIBITOR DISCOVERY?





STEVEN ALBANESE

Achieving 100x selectivity is difficult,
but predictive modeling can have substantial impact.

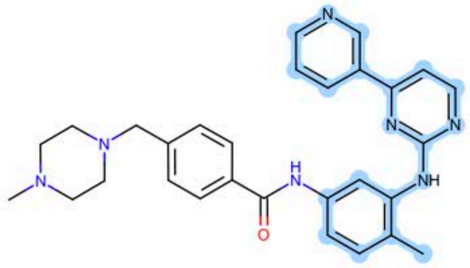
FEP+/OPLS3

LINGLE WANG
SCHRÖDINGER

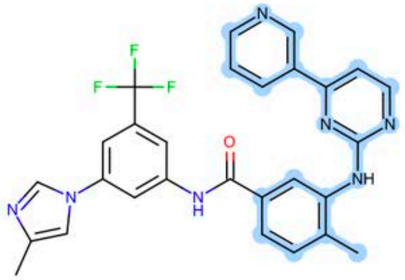
PERSES: A PLATFORM FOR RELATIVE ALCHEMICAL FREE ENERGY CALCULATIONS

Propose new molecules with common scaffold via MCSS

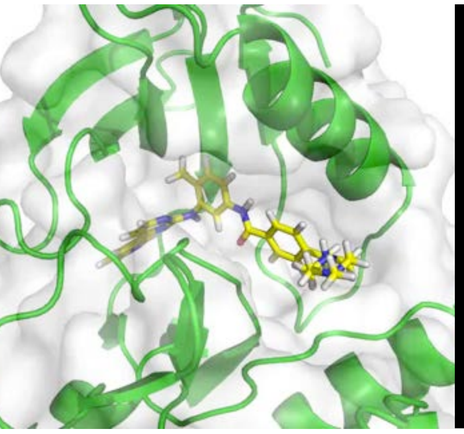
Imatinib

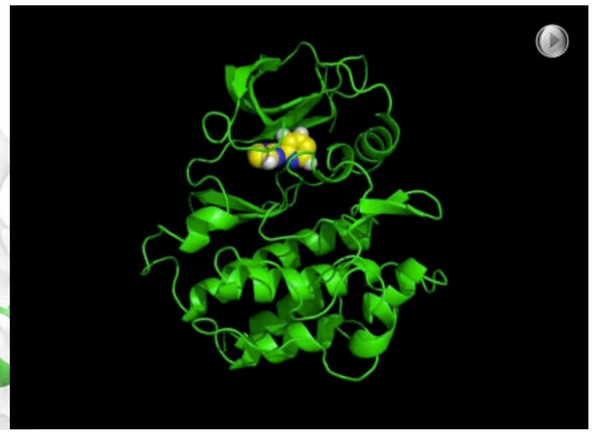


Nilotinib

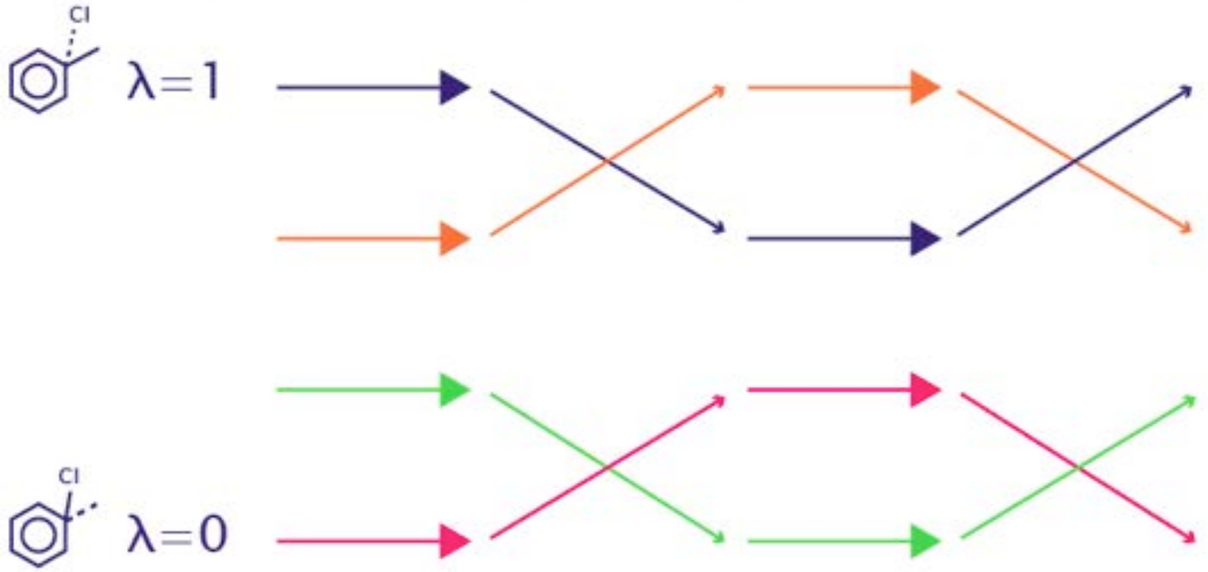


Build in new atoms with reversible-jump Monte Carlo





replica exchange among Hamiltonians

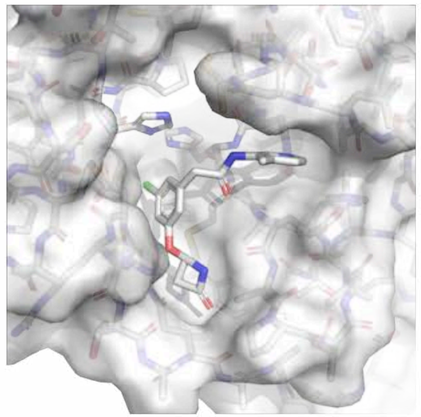


HANNAH BRUCE MACDONALD
DOMINIC RUFA
PATRICK GRINAWAY

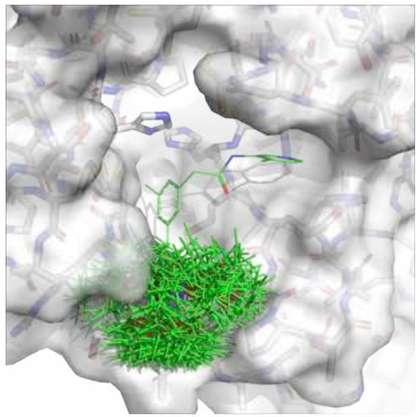
SILICON Therapeutics BRYCE ALLEN
WOODY SHERMAN

WE RECENTLY USED PERSES TO RUN THOUSANDS OF FREE ENERGY CALCULATIONS/WEEK FOR THE COVID MOONSHOT

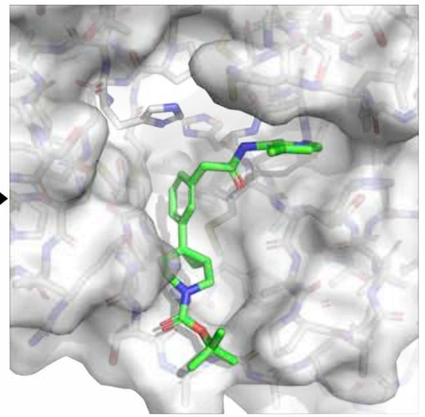
X-ray structure as reference



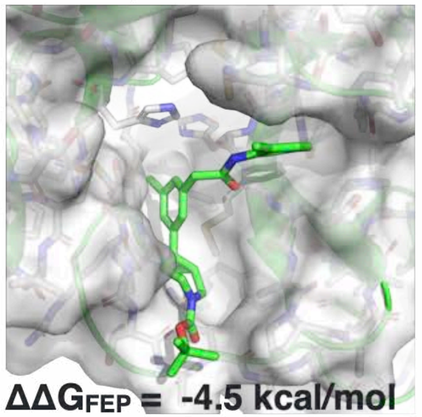
constrained enumeration of poses for proposed molecule



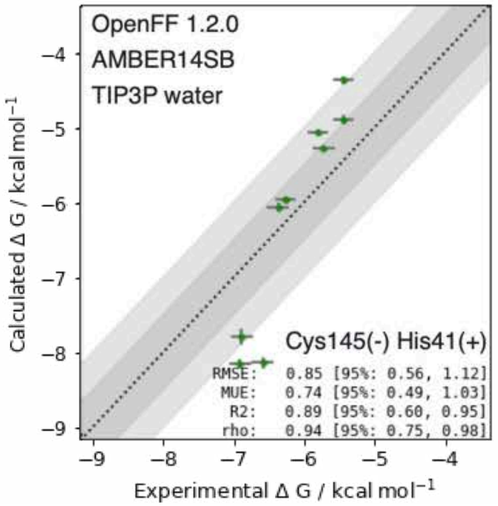
selection of pose with best docking score



nonequilibrium alchemical free energy calculation final posed structure



retrospective performance on 3-aminopyridine lead series

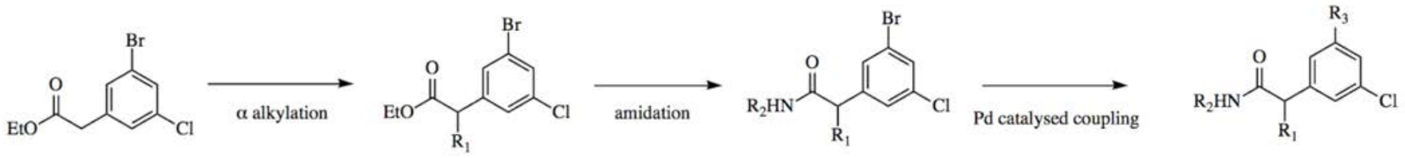


perses: open source relative alchemical free energy calculations
http://github.com/choderalab/perses
Open Force Field Initiative OpenFF (“Parsley”) small molecule force field
http://openforcefield.org

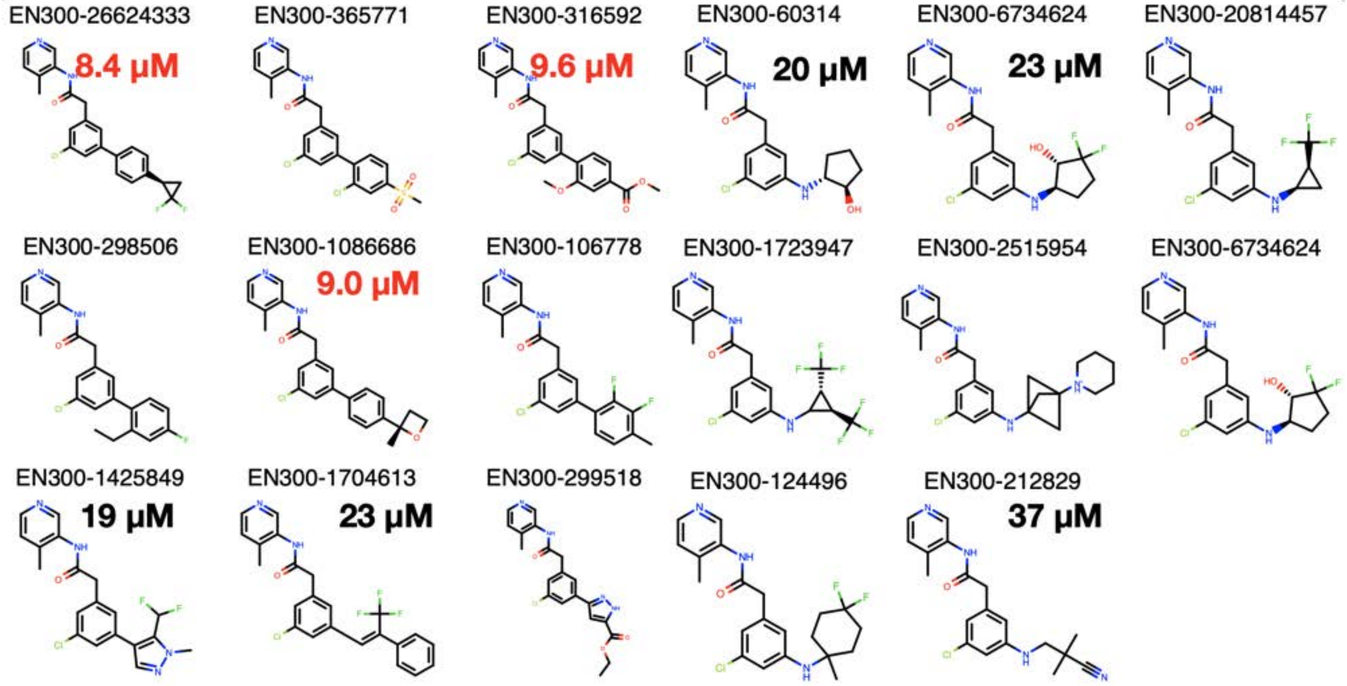
Dominic Rufa, Hannah Bruce Macdonald, William Glass, Matt Wittman, David Dotson
+ The Folding@home and COVID Moonshot contributors

FREE ENERGY CALCULATIONS CAN RAPIDLY PRIORITIZE COMPOUNDS FROM LARGE VIRTUAL SYNTHETIC LIBRARIES

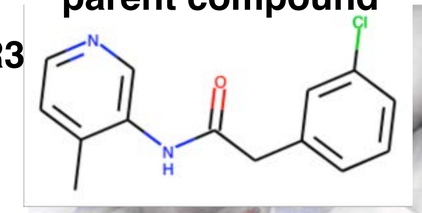
Can we engage S4 from this 5,000-compound virtual synthetic library varying R3



Top free energy calculation compounds and experimental affinity measurements:



parent compound



IC_{50} = 25 µM

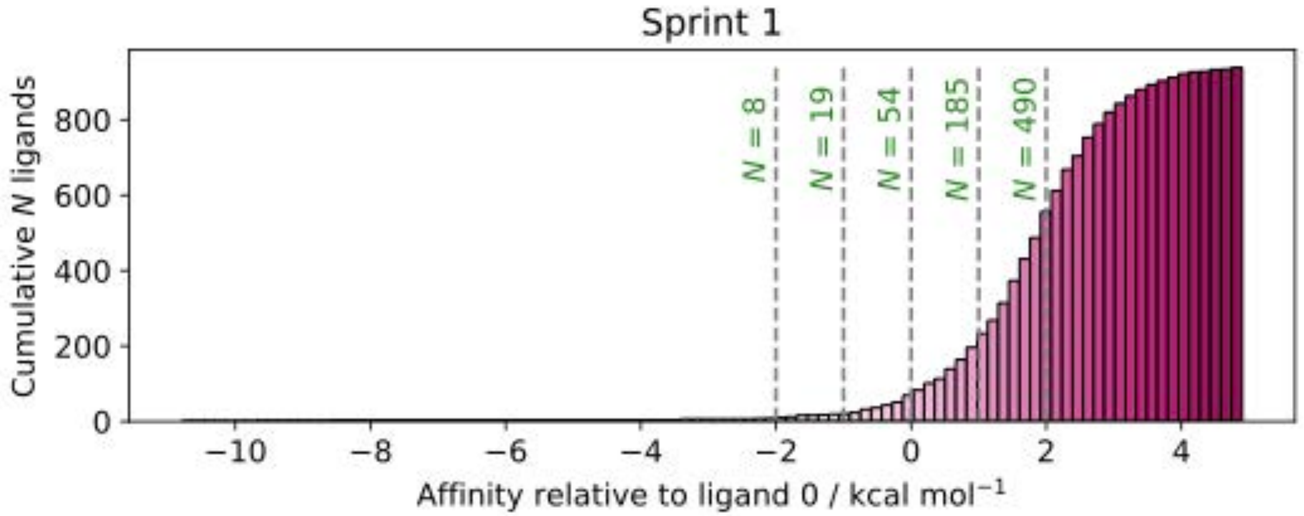
TRY-UNI-714a760b-6

top compounds from free energy calculations

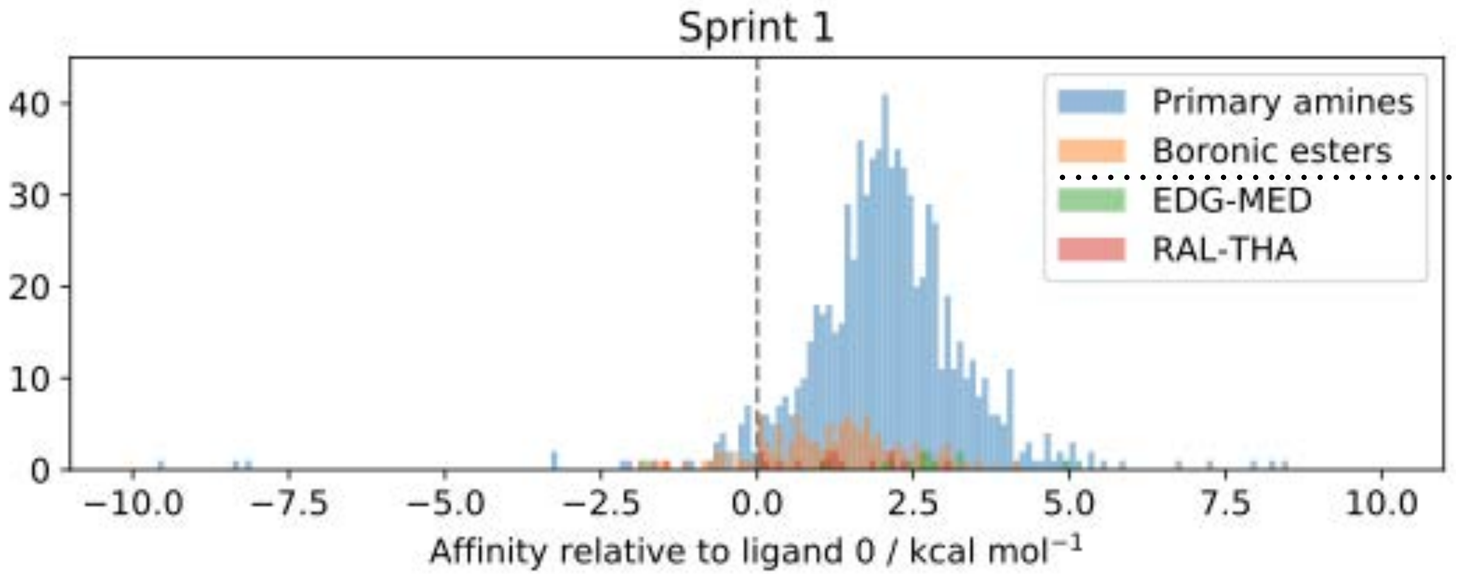
COVID Moonshot: [Moonshot] [Fragalysis] [Dashboard]

UNSURPRISINGLY, MOST IDEAS ARE BAD IDEAS

better **worse**



HUMAN CHEMISTS MAY BE BIASED TOWARD BETTER COMPOUNDS, BUT THE COMPUTER CAN GENERATE AND SCORE MORE IDEAS

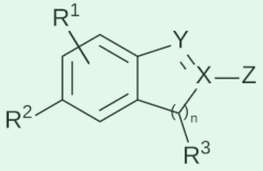


computer

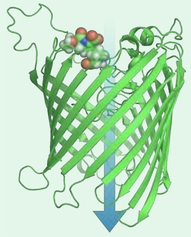
humans

OPEN SOFTWARE FRAMEWORKS ALLOW US TO BUILD POWERFUL APPLICATIONS

yank



perses



iapetus

ROCSAlt

targeted domain-specific applications (Python)
http://github.com/choderalab

APPLICATIONS

openmmtools

high-level simulation algorithms, alchemical tools (Python to enable rapid development)
http://github.com/choderalab/openmmtools

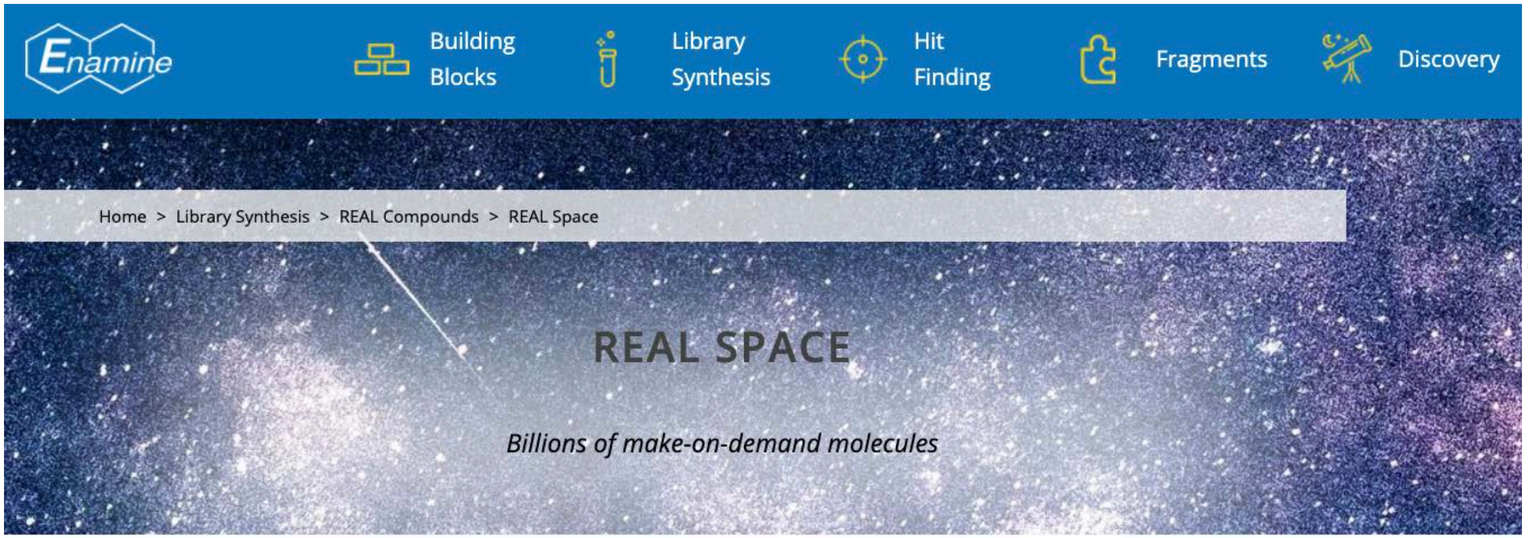
ALGORITHMS

OpenMM

general GPU-accelerated MD simulation engine (C++/CUDA/OpenCL with Python API)
http://openmm.org

CORE

HOW CAN WE SCALE UP FURTHER?



The *REAL* Space comprises over 17 billion make-on-demand molecules and is currently the largest offer of commercially available compounds. The *REAL* compounds in the Space are assembled via more than 150 well-validated parallel synthesis protocols applied to over 123 000 qualified reagents and building blocks. The synthetic protocols include standard and advanced one-pot procedures. They differ in the number of steps, type of purification, and compound handling, and therefore in the effort required to deliver the products. Please contact us at libraries@enamine.net for more details.

Conveniently purchaseable compound space is already ~17B compounds

https://enamine.net/library-synthesis/real-compounds/real-database

GRAPH CONVOLUTIONAL NETWORKS ARE PARTICULARLY WELL-SUITED TO CHEMISTRY

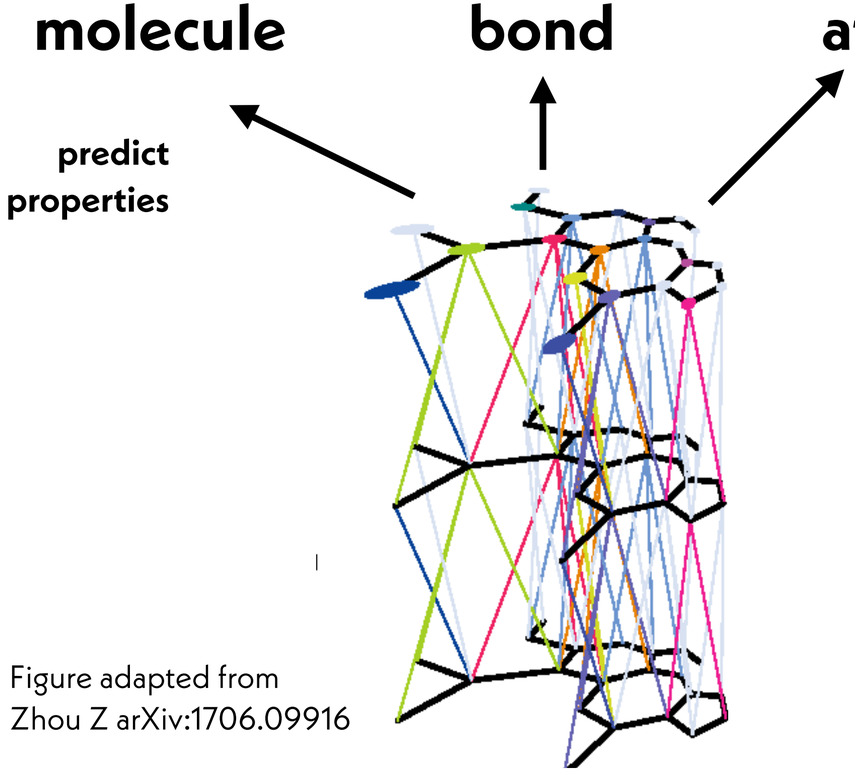


Figure adapted from
Zhou Z arXiv:1706.09916

$$\mathbf{e}_k^{(t+1)} = \phi^e(\mathbf{e}_k^{(t)}, \sum_{i \in \mathcal{N}_k^e} \mathbf{v}_i, \mathbf{u}^{(t)}), \quad \text{(edge update)}$$

$$\bar{\mathbf{e}}_i^{(t+1)} = \rho^{e \to v}(E_i^{(t+1)}), \quad \text{(edge to node aggregate)}$$

$$\mathbf{v}_i^{(t+1)} = \phi^v(\bar{\mathbf{e}}_i^{(t+1)}, \mathbf{v}_i^{(t)}, \mathbf{u}^{(t)}), \quad \text{(node update)}$$

$$\bar{\mathbf{e}}^{(t+1)} = \rho^{e \to u}(E^{(t+1)}), \quad \text{(edge to global aggregate)}$$

$$\bar{\mathbf{v}}^{(t+1)} = \rho^{v \to u}(V^{(t)}), \quad \text{(node to global aggregate)}$$

$$\mathbf{u}^{(t+1)} = \phi^u(\bar{\mathbf{e}}^{(t+1)}, \bar{\mathbf{v}}^{(t+1)}, \mathbf{u}^{(t)}), \quad \text{(global update)}$$

YUANQING WANG

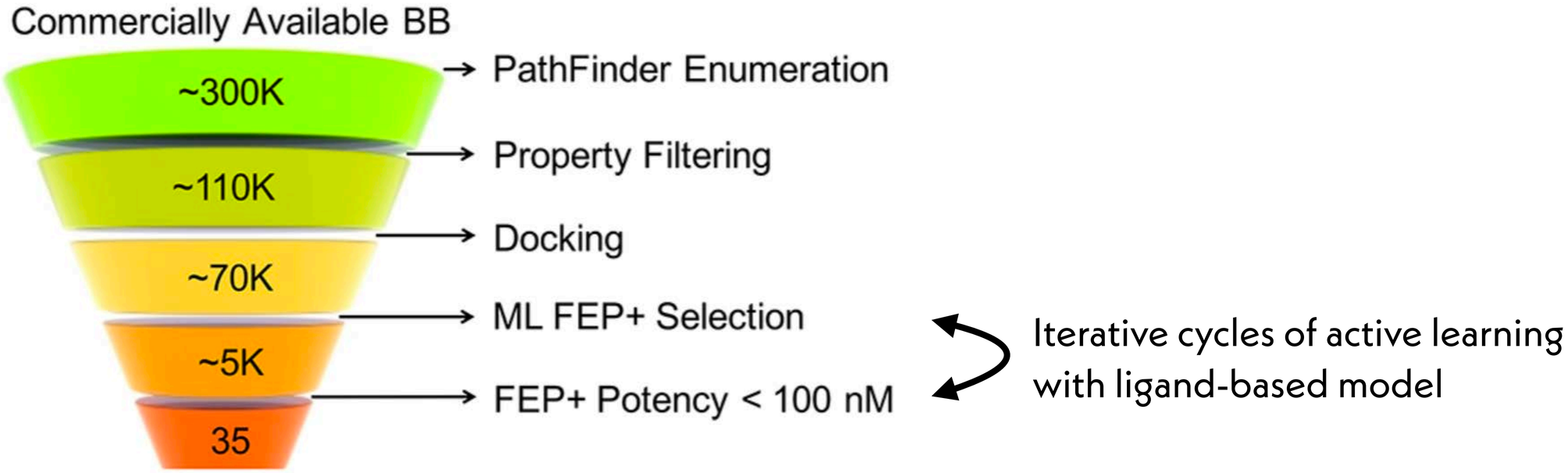
Gimlet

Graph Inference on MoLEcular Topology

http://github.com/choderalab/gimlet

- `gin/` the core (and fun) part of the package.
  - `i_o/` reading and writing popular molecule embedding/representing structures.
  - `deterministic/` property predictions, conformer and charge generations.
  - `probabilistic/` molecular machine learning through graph networks.
- `lime/` auxiliary scripts.
  - `for_biologists/` ready-to-use modules and scripts.
  - `architectures/` off-the-shelf model architectures developed elsewhere.
  - `scripts/` fun scripts we used to generate data and hypothesis.
  - `trained_models/` *Nomen est omen.*

ACTIVE LEARNING OF FREE ENERGY CALCULATIONS IS ALREADY HERE



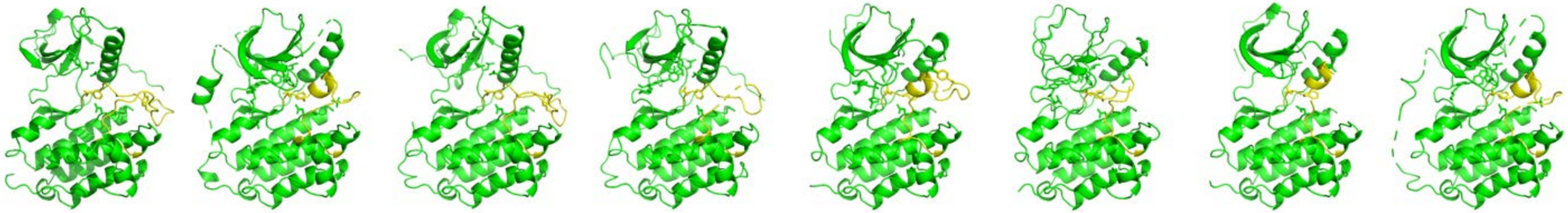
But how can we generalize to related targets (mutants, other superfamily members?)

Konze et al. JCIM 59:3782, 2019

BLENDING PHYSICAL MODELS AND ML IS THE FUTURE

automated modeling
of mutant conformations

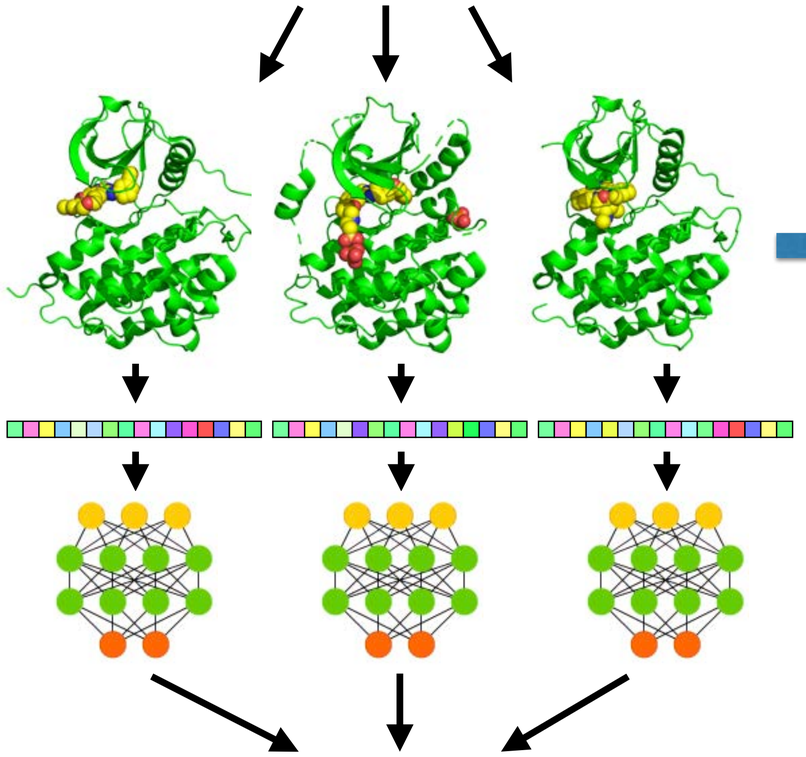
distinct conformations of *apo* receptor

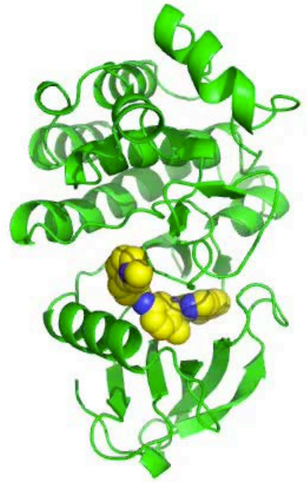


hybrid docking
shape overlay and
physical docking

featurize
sequence, chemical,
structural features

deep learning
to predict conformation/
pose specific affinity





prioritize conformations,
poses for detailed
alchemical free energy
calculations

Boltzmann pooling across
conformations/poses
to predict affinities

$$\Delta G = -k_B T \ln \sum_i e^{-\beta(\Delta G_i^{\mathrm{conf}} + \Delta G_i^{\mathrm{bind}})}$$

ANDREA VOLKAMER

TALIA KIMBER

JAIME RODRIGUEZ GUERRA

FREE ENERGY CALCULATIONS ARE A FIELD IN TRANSITION FROM SCIENCE TO ENGINEERING

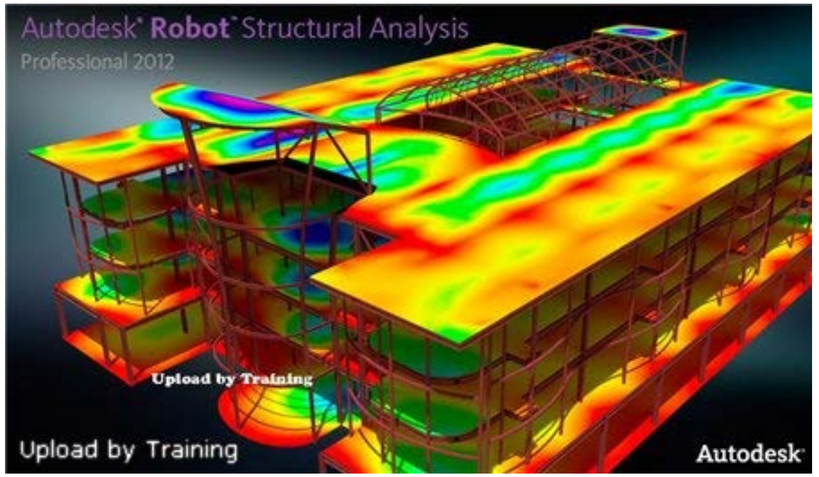
SCIENCE

We got it to work once!
Let's publish it in Nature!

ENGINEERING

We do this regularly
with near 100% success.

HOW CAN WE TRANSITION FROM A RESEARCH FIELD TO AN ENGINEERING FIELD?



STRUCTURAL ENGINEERING WASN'T ALWAYS SO SUCCESSFUL

There were 250 bridge failures in the US and Canada between 1878-1888.

*"**The subject of mechanical pathology is relatively as legitimate and important a study to the engineer as medical pathology is to the physician.** While we expect the physician to be familiar with physiology, without pathology he would be of little use to his fellow-men, and it [is] as much within the province of the engineer to **investigate causes, study symptoms, and find remedies for mechanical failures** as it is to direct the sources of power in nature for the use and convenience of man."*

- George Thomson, 1888

THE DOMAIN OF APPLICABILITY OF FREE ENERGY CALCULATIONS IS CURRENTLY LIMITED

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

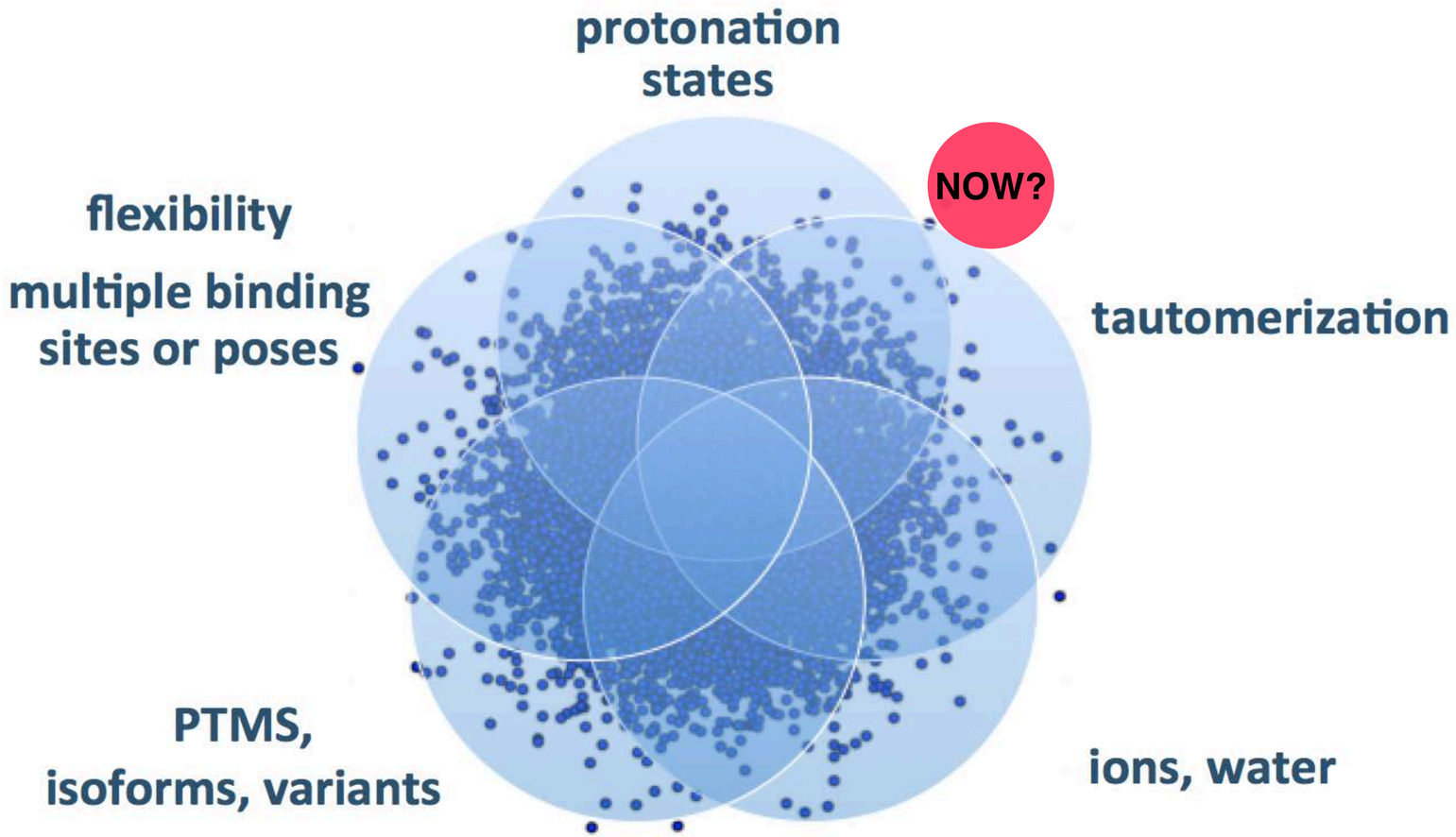
One well-specified, well-resolved isoform/species

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

No exotic chemistries

No metals or prosthetic groups

No membranes?



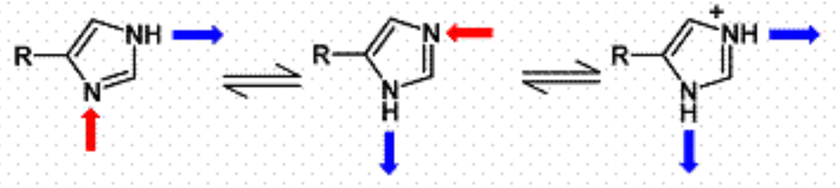
TARGETS
CHALLENGES

FREE ENERGY CALCULATIONS FAIL FOR THREE MAIN REASONS

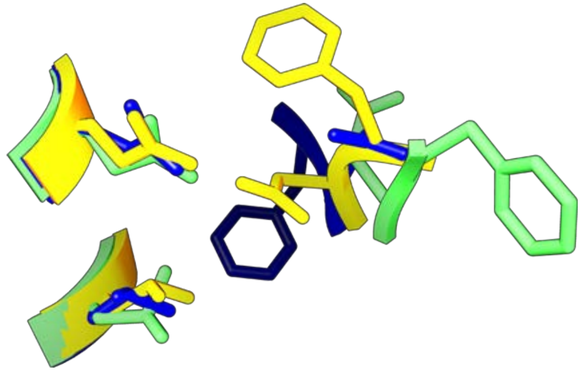
1. The forcefield does a poor job of modeling the physics of our system

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

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

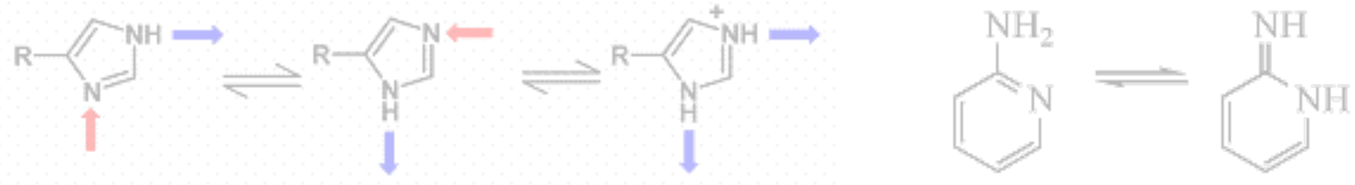


FREE ENERGY CALCULATIONS FAIL FOR THREE MAIN REASONS

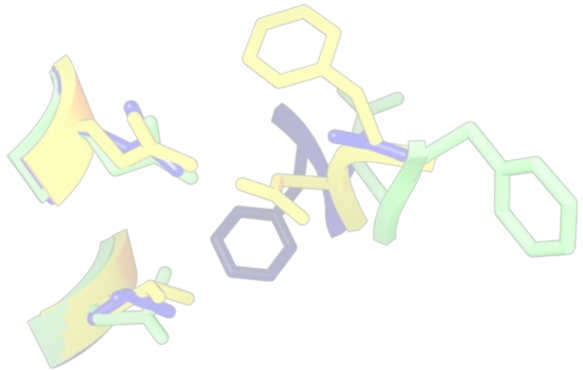
1. The forcefield does a poor job of modeling the physics of our system

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

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

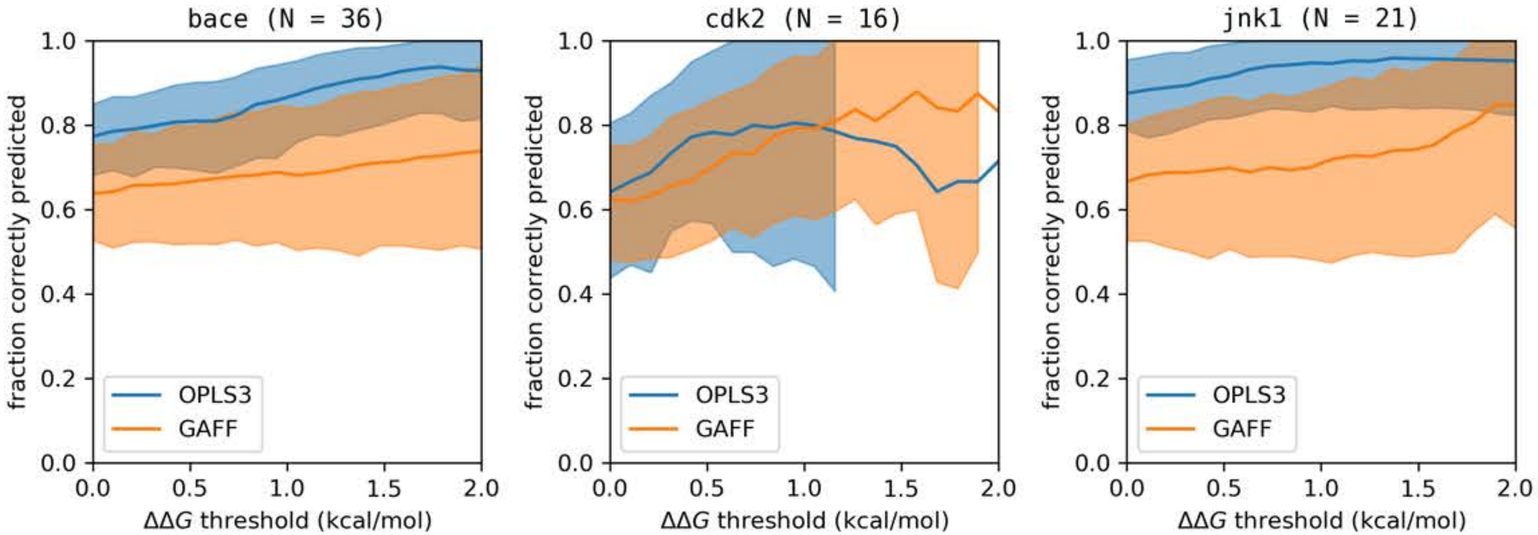


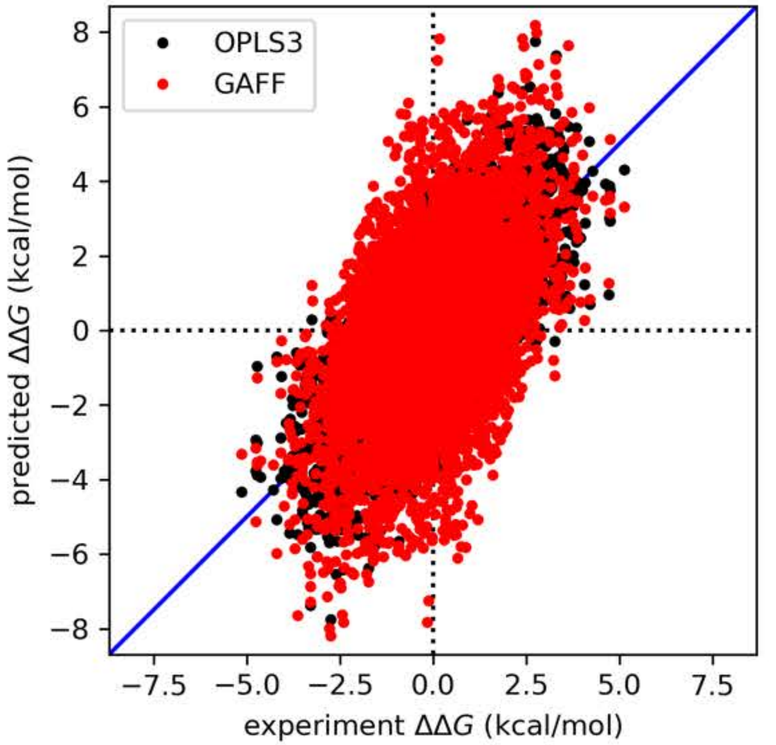
PUBLIC FORCE FIELDS NEED TO CATCH UP

Validation of AMBER/GAFF for Relative Free Energy Calculations
Lin Song Tai-Sung Lee Chun Zhu Darrin M. York Kenneth M. Merz Jr.
chemRxiv preprint 2 Feb 2019
https://chemrxiv.org/articles/Validation_of_AMBER_GAFF_for_Relative_Free_Energy_Calculations/7653434

		all within-target pairs ΔΔG (N = 5620)
RMSE:	OPLS3	1.37 [95%: 1.34, 1.39] kcal/mol
RMSE:	GAFF	1.97 [95%: 1.93, 2.01] kcal/mol
MUE :	OPLS3	1.09 [95%: 1.07, 1.11] kcal/mol
MUE :	GAFF	1.55 [95%: 1.51, 1.58] kcal/mol
R2 :	OPLS3	0.10 [95%: 0.06, 0.15] kcal/mol
R2 :	GAFF	-0.87 [95%: -0.96, -0.76] kcal/mol
rho :	OPLS3	0.73 [95%: 0.72, 0.74] kcal/mol
rho :	GAFF	0.53 [95%: 0.51, 0.55] kcal/mol

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





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

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.

http://openforcefield.org

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

MolSSI

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 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

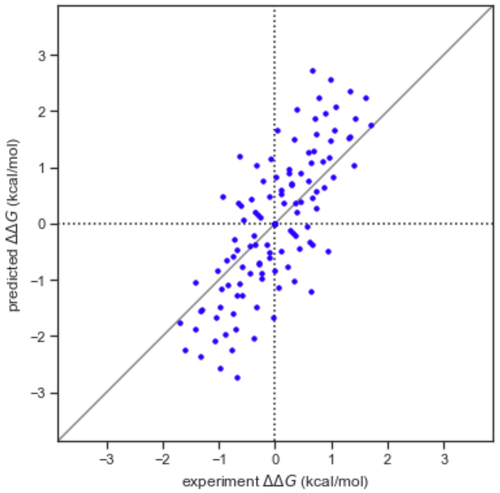
https://openforcefield.org/news/introducing-openforcefield-1.0/

21+ researchers across 7+ sites

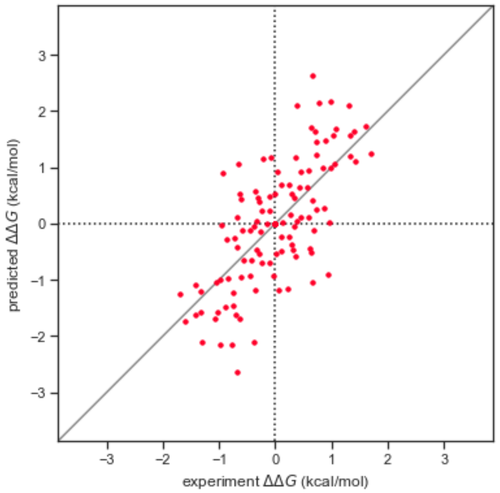
WE'VE MADE SIGNIFICANT PROGRESS IN JUST A YEAR

Open Force Field Initiative →

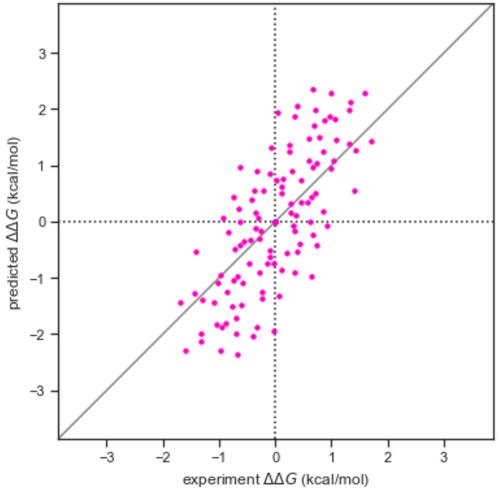
GAFF 1 (1999)



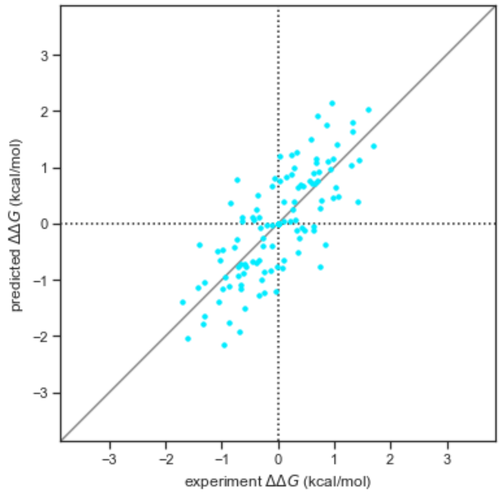
OPLS2.1 (2015)



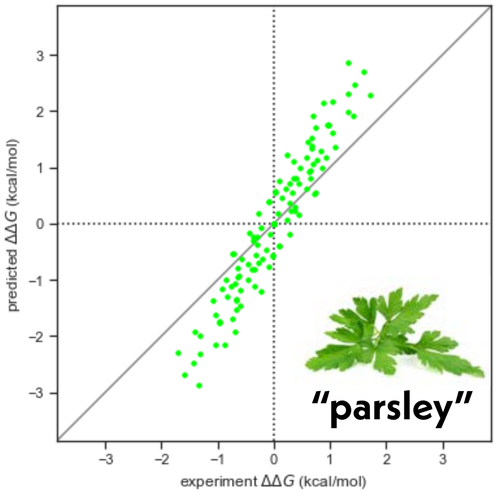
GAFF 2 (2016)



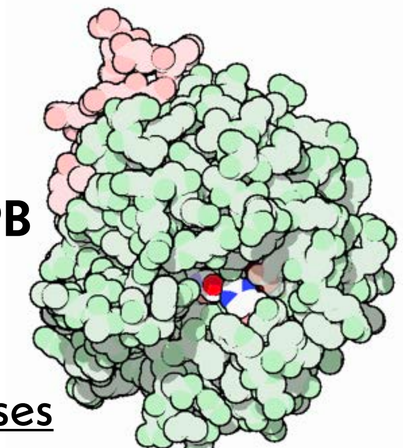
smirnoff99Frosst (2018)



openff 1.0 (2019)



thrombin
PDB101: 1PPB



HANNAH BRUCE MACDONALD
MSKCC

http://github.com/choderalab/perses

DOMINIC RUFA

CAN WE MAKE FORCE FIELD FITTING AS EASY AS TRAINING A MACHINE LEARNING MODEL IN TENSORFLOW?

training a neural network

```
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

https://www.tensorflow.org/overview

import your tools

grab a standard, curated dataset

define a novel model architecture

declare your objectives in training it

fit it

use it

CAN WE MAKE FORCE FIELD FITTING AS EASY AS TRAINING A MACHINE LEARNING MODEL IN TENSORFLOW?

training a neural network

```python
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

fitting a force field

```python
import openforcefield as off
training_data, benchmark_data = off.datasets.load('2019-Q1')

force_field_model = off.models.ForceFieldModel([
    off.models.forces.HarmonicBond(),
    off.models.forces.HarmonicAngle(),
    off.models.forces.PeriodicTorsion(max_order=6),
    off.models.forces.LennardJones(),
    off.models.forces.BondChargeCorrections(),
])

model.compile(optimizer='L-BFGS',
     loss='error-weighted',
     metrics=['accuracy'])

model.fit(training_data)

model.evaluate(test_data)
```

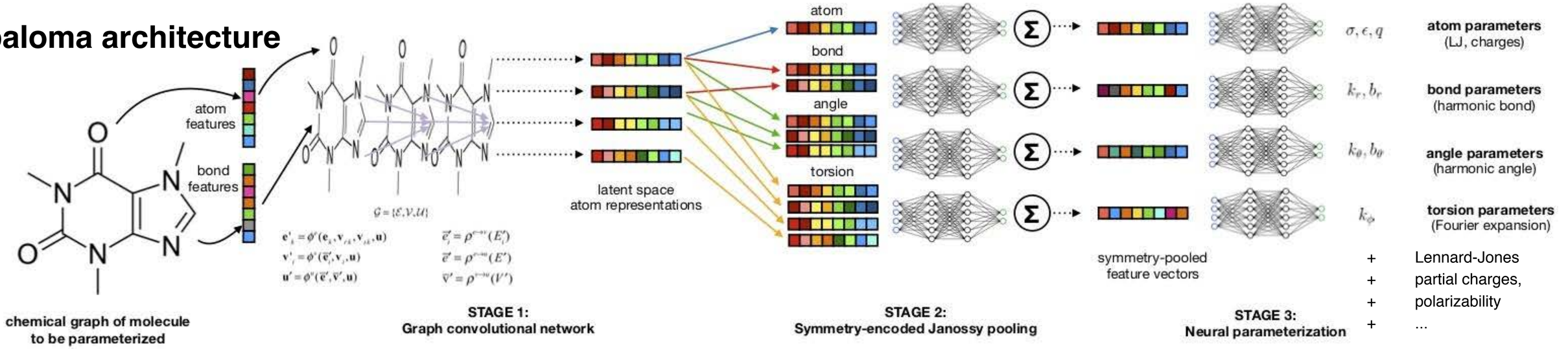
Run code now Try in Google's interactive notebook

https://www.tensorflow.org/overview

This isn't quite this simple yet, but this gives you an idea of where we're headed

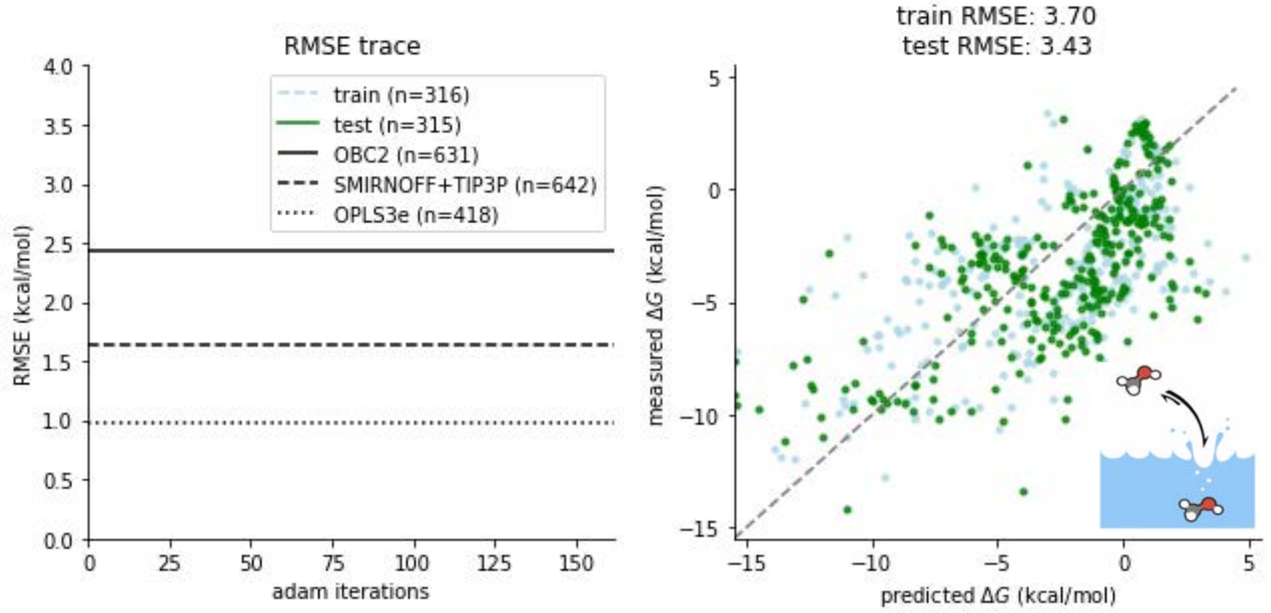
END-TO-END DIFFERENTIABLE MM PARAMETER ASSIGNMENT ENABLES FORCE FIELDS TO BE EASILY REFIT TO EXPERIMENTAL DATA

espaloma architecture



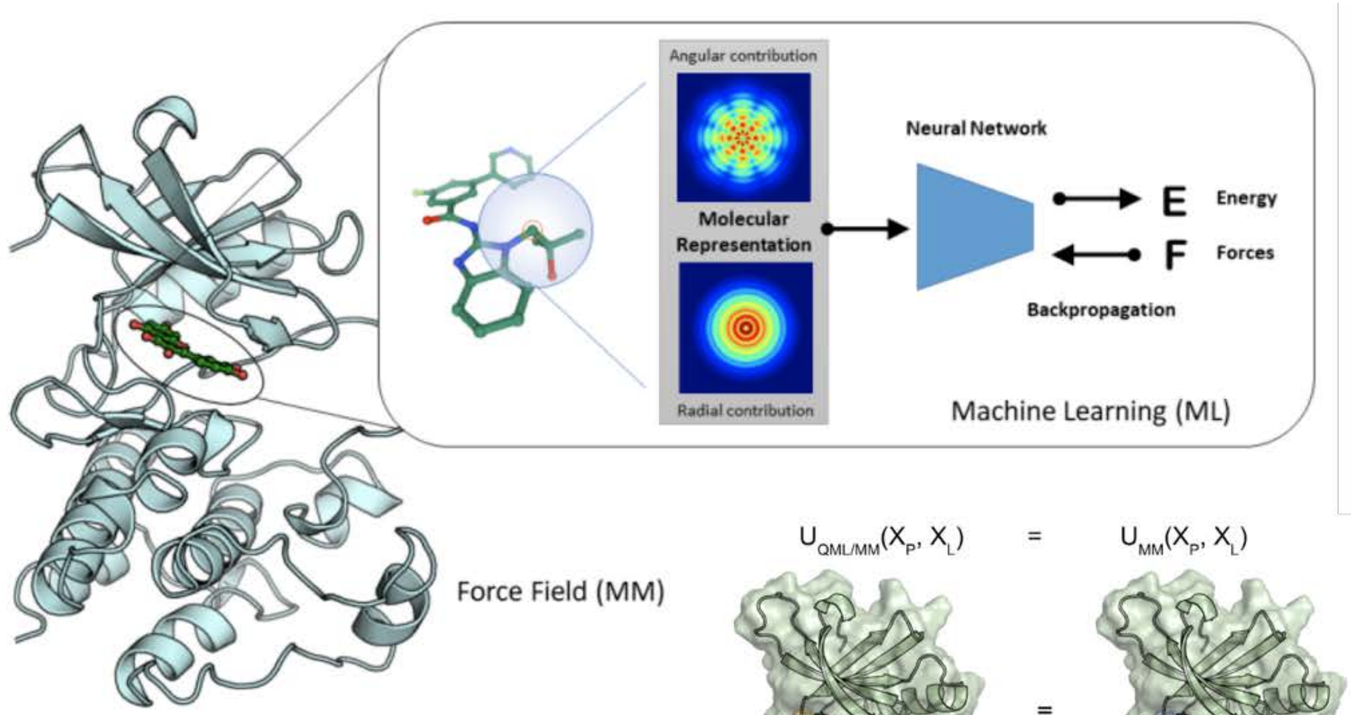
espaloma will self-consistently assign parameters to small molecules, proteins, biopolymers, lipids, etc.

learning hydration free energies for an implicit solvent model from experimental data

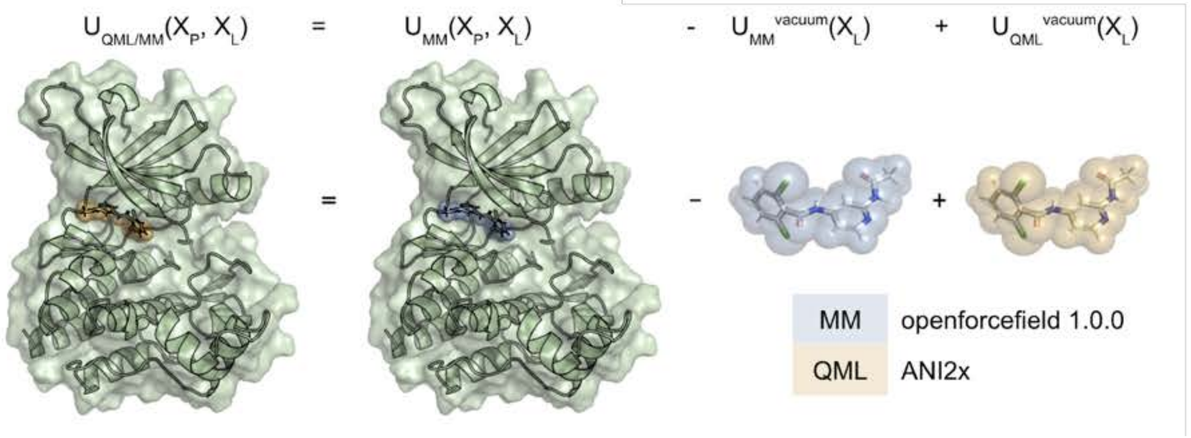


preprint: https://arxiv.org/abs/2010.01196
code: https://github.com/choderalab/espaloma

HYBRID QUANTUM MACHINE LEARNING / MOLECULAR MECHANICS (QML/MM) FREE ENERGY CALCULATIONS CUT ERROR IN HALF



many QML/MM formulations possible, including those that use QML for protein-ligand interactions



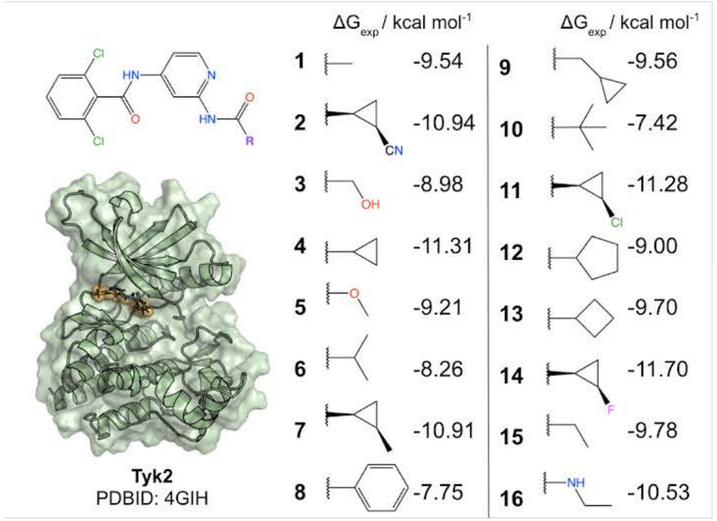
Rufa, Bruce Macdonald, Fass, Wieder, Grinaway, Roitberg, Isayev, and **Chodera**.
preprint: https://doi.org/10.1101/2020.07.29.227959
code: https://github.com/choderalab/qmlify

HYBRID QUANTUM MACHINE LEARNING / MOLECULAR MECHANICS (QML/MM) FREE ENERGY CALCULATIONS CUT ERROR IN HALF

MM (OPLS2.1 + CM1A-BCC charges)
Missing torsions from LMP2/cc-pVTZ(-f) QM calculations
SPC water

MM (OpenFF 1.0.0 "Parsley")
AMBER14SB protein force field
TIP3P; Joung and Cheatham ions

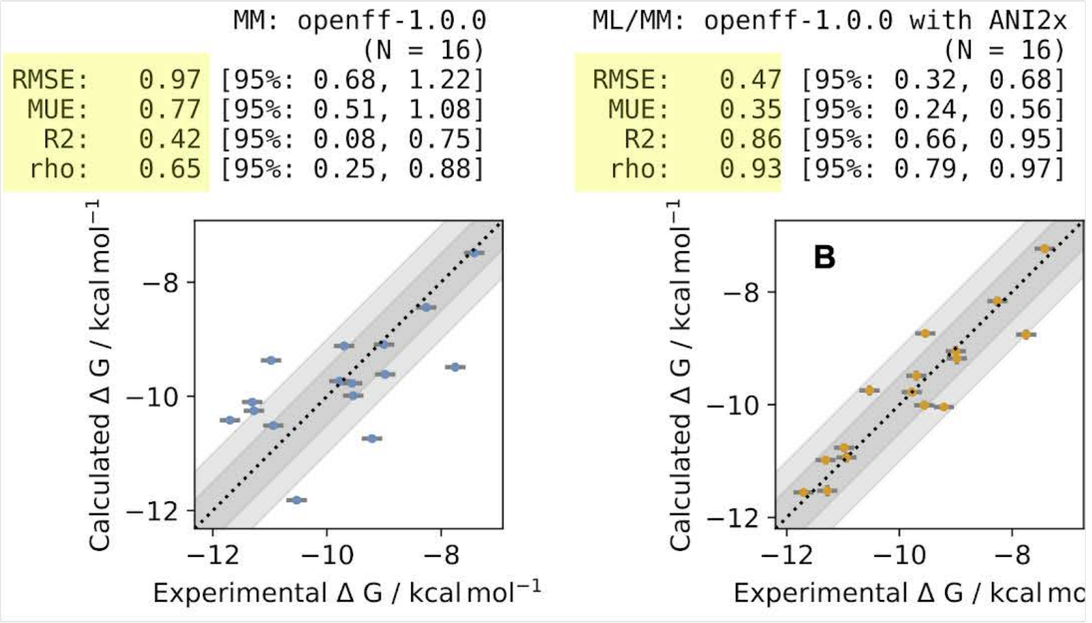
QML/MM (OpenFF 1.0.0 + ANI2x)
AMBER14SB protein force field
TIP3P; Joung and Cheatham ions



	ΔG_{exp} / kcal mol^{-1}		ΔG_{exp} / kcal mol^{-1}
1	-9.54	9	-9.56
2	-10.94	10	-7.42
3	-8.98	11	-11.28
4	-11.31	12	-9.00
5	-9.21	13	-9.70
6	-8.26	14	-11.70
7	-10.91	15	-9.78
8	-7.75	16	-10.53

	Tyk2
no. of compds	16
binding affinity range (kcal/mol)	4.3
crystal structure	4GIH
series ref	52,53
no. of perturbations	24
MUE FEP	0.75 ± 0.11
RMSE FEP	0.93 ± 0.12

Free energies are in units of kilocalories per mole.



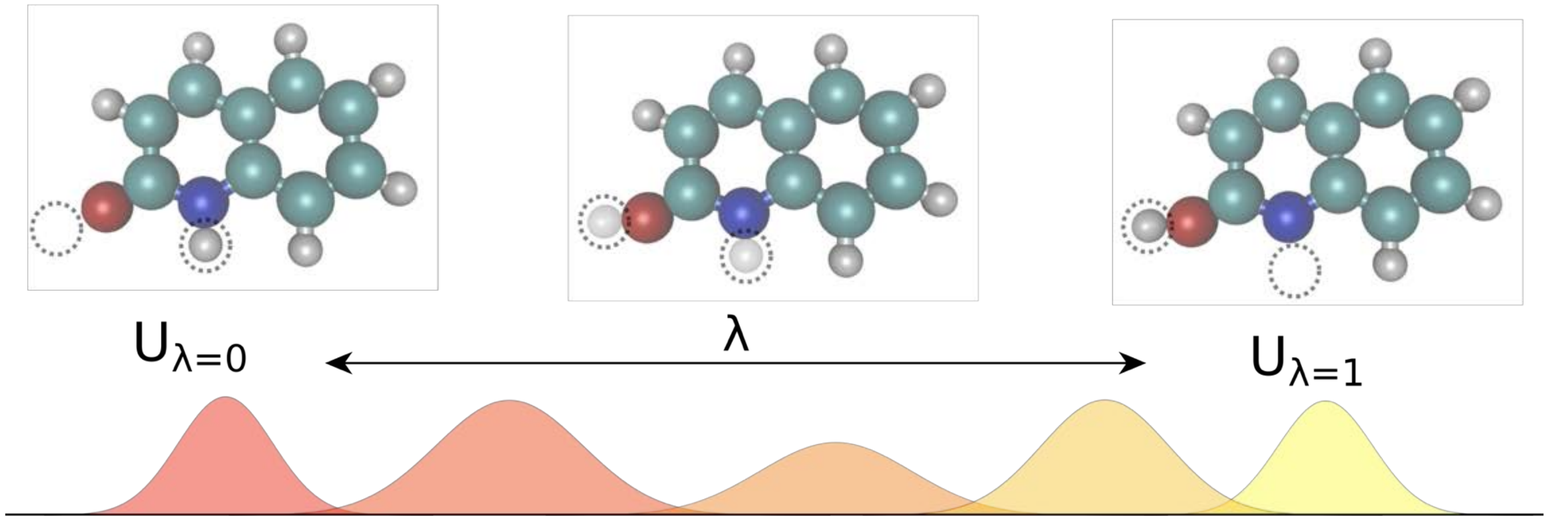
Tyk2 benchmark system from Wang et al. JACS 137:2695, 2015
replica-exchange free energy calculations with solute tempering (FEP/REST)

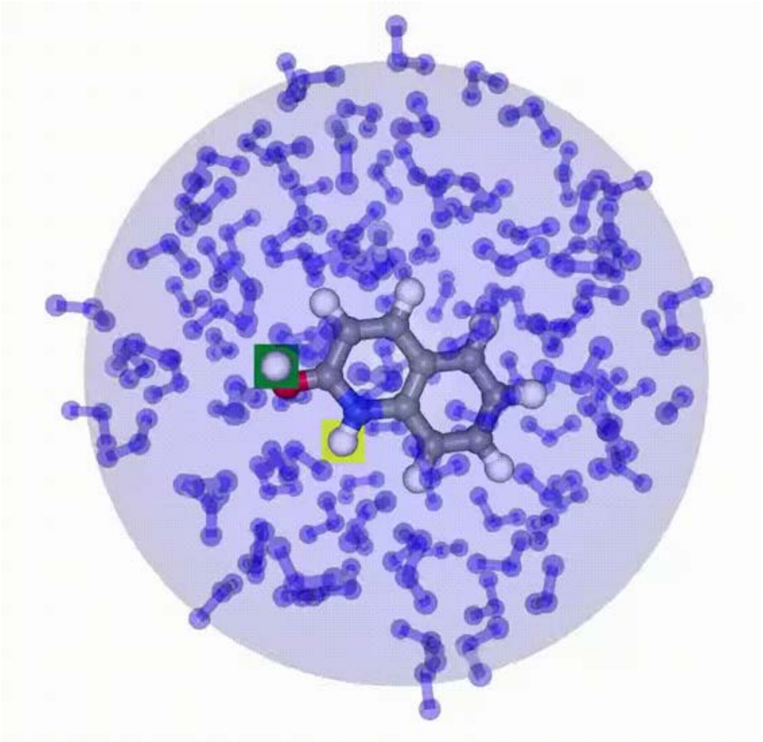
replica-exchange free energy calculations with perses
preprint: https://doi.org/10.1101/2020.07.29.227959
code: https://github.com/choderalab/perses
https://github.com/choderalab/qmlify

PURE QUANTUM MACHINE LEARNING (QML) POTENTIALS CAN BE USED TO COMPUTE FREE ENERGY DIFFERENCES BETWEEN CHEMICAL SPECIES

Potentials are free of singularities, so **simple linear alchemical potentials** can robustly compute alchemical free energies

$$U(x;\lambda) = (1-\lambda)U_{\lambda=0}(x) + \lambda U_{\lambda=1}(x)$$



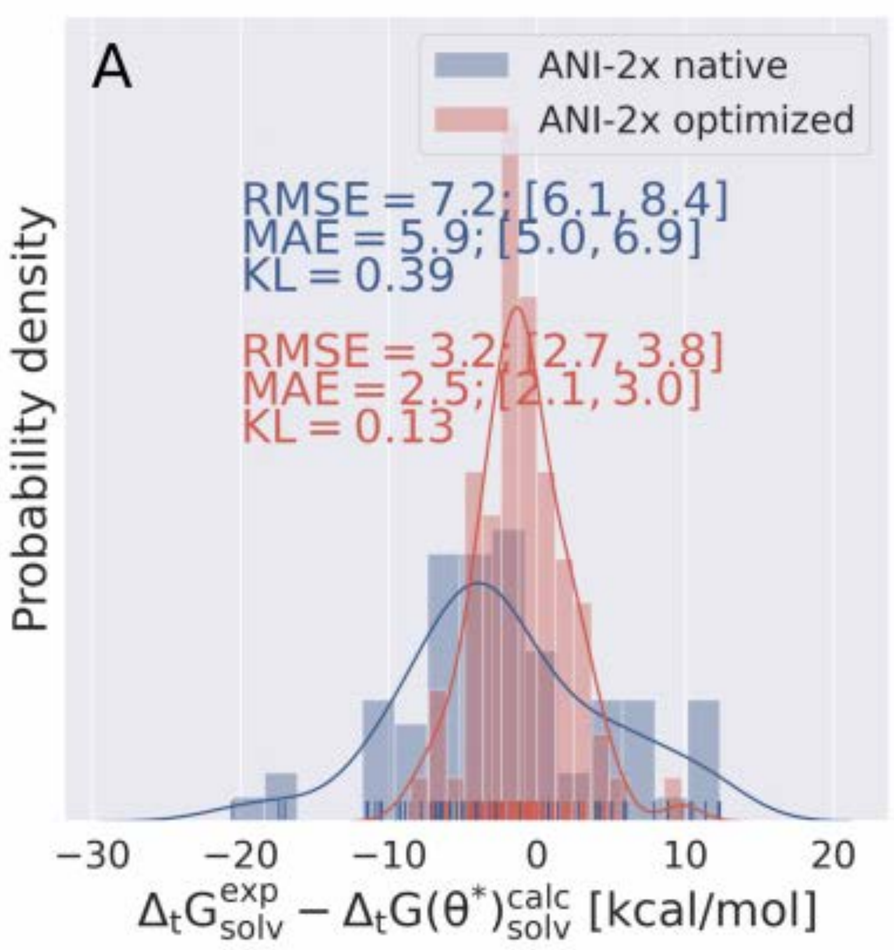


Simple atomic restraints can be used to improve efficiency by preventing atoms from flying away

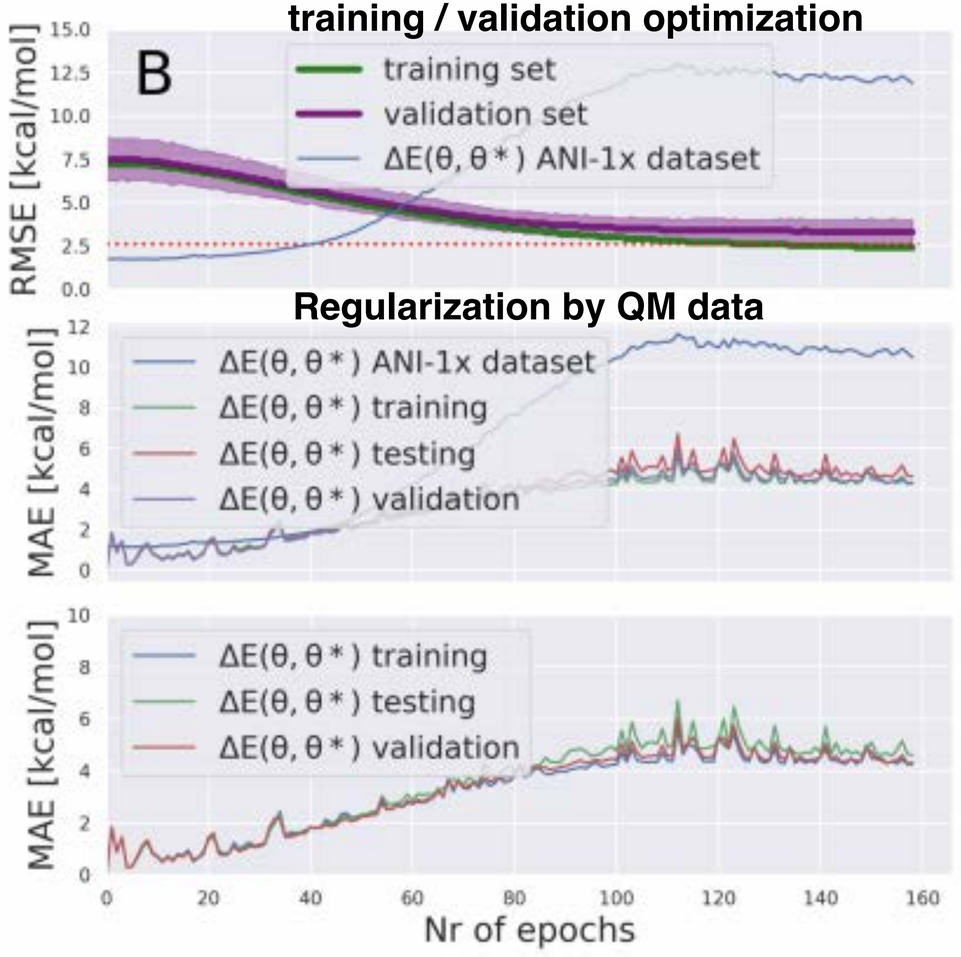
preprint: https://doi.org/10.1101/2020.10.24.353318
code: https://github.com/choderalab/neutromeratio

PURE QUANTUM MACHINE LEARNING (QML) POTENTIALS CAN BE TUNED/RETRAINED BY FREE ENERGIES, REGULARIZED BY QM DATA

test set performance

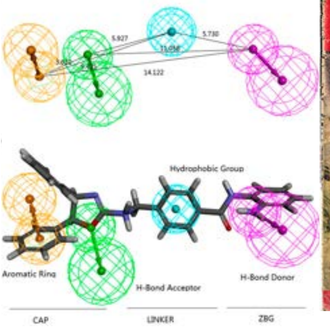


training / validation optimization



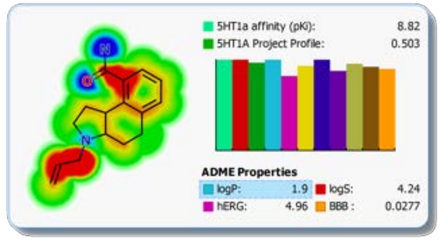
preprint: https://doi.org/10.1101/2020.10.24.353318
code: https://github.com/choderalab/neutromeratio

Fast on-the-fly reweighting enables inexpensive loss/gradient computation without repeating expensive free energy calculation

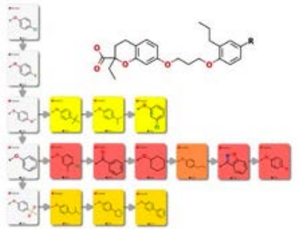


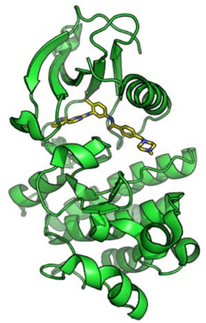
pharmacophore models

physical modeling

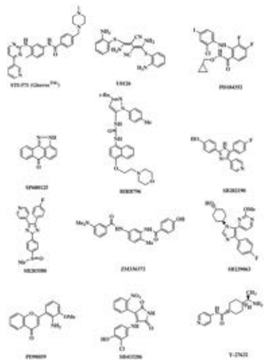


ADME/Tox models





structures



known ligands

PROPOSE MOLECULES

PERFORM ASSAYS

ASSESS MODEL UNCERTAINTIES

EVALUATE HYPOTHESES

ENUMERATE HYPOTHESES

EXIT CYCLE

?

Several excellent graphics taken from http://www.optibrium.com/

AUTOMATED LABORATORIES ARE CHANGING WETLAB BIOLOGY

CHODERA LAB, Z1745D
ZUCKERMAN RESEARCH CENTER

http://choderalab.org/resources

AUTOMATED CLOUD LABORATORIES ARE TRANSFORMING WETLAB BIOLOGY

TRANSCRIPTIC, MENLO PARK CA

https://www.transcriptic.com

AUTOMATED CLOUD LABORATORIES ARE TRANSFORMING WETLAB BIOLOGY

"When we started Transcriptic, we set out with the goal of giving the life sciences the same structural advantages that web has enjoyed, **making it possible for two postdocs with a laptop in a coffee shop to run a drug company** without the need for millions of dollars in capital equipment or lab space."

- **Max Hodak**, Transcriptic (Founder and President)

http://blog.transcriptic.com/adding-potential-energy-transcriptics-series-a

INEXPENSIVE WETLAB ROBOTS ARE HERE

$5000

OPENTRONS OT-2

http://opentrons.com

AUTOMATION BRINGS STANDARDIZATION

AUTOPROTOCOL .org

An open standard for life science experimental design and automation.

open standard for specifying wetlab protocols

encodes experimental protocols **unambiguously**

extensible through open community process

Python tools enable **metaprogramming experiments:**
code that designs new experiments based on previous data

AUTOPROTOCOL .org

An open standard for life science experimental design and automation.

```
import json
from autoprotocol.protocol import Protocol

#instantiate new Protocol object
p = Protocol()

# append refs (containers) to Protocol object
bacteria = p.ref("bacteria", cont_type="96-pcr", storage="cold_4")
media = p.ref("media", cont_type="micro-1.5", storage="cold_4")
reaction_plate = p.ref("reaction_plate", cont_type="96-flat", storage="warm_37")

# distribute media from 1.5mL tube to reaction wells
p.distribute(media.well(0).set_volume("1000:microliter"),
             reaction_plate.wells_from(0,4), ["140:microliter",
             "130:microliter", "120:microliter", "100:microliter"])

# transfer bacteria from source wells to reaction wells
p.transfer(bacteria.wells_from(0,4), reaction_plate.wells_from(0,4),
           ["10:microliter", "20:microliter", "30:microliter", "40:microliter"])

# cover plate
p.cover(reaction_plate)

# incubate bacteria at 37 degrees for 5 hours
p.incubate(reaction_plate, "warm_37", "5:hour", shaking=True)

# read absorbance of the first four wells on the reaction plate at 600 nanometers
p.absorbance(reaction_plate, reaction_plate.wells_from(0,4).indices(), "600:nanometer",
             "OD600_reading_01092014")

print json.dumps(p.as_dict(), indent=2)
```

provision labware

distribute media

transfer bacteria

cover plate

incubate 5h @ 37C

measure OD600

WHY IS THIS TRANSFORMATIVE?

With an machine-readable way to **describe** the experimental protocol, we can also explain to the machine how to **interpret** the experiment.

BAYESIAN INFERENCE ALLOWS US TO REASON FROM DATA

$$\underbrace{p(\theta|\mathcal{D})}_{\text{posterior}} \propto \underbrace{p(\mathcal{D}|\theta)}_{\text{likelihood}} \underbrace{p(\theta)}_{\text{prior}}$$

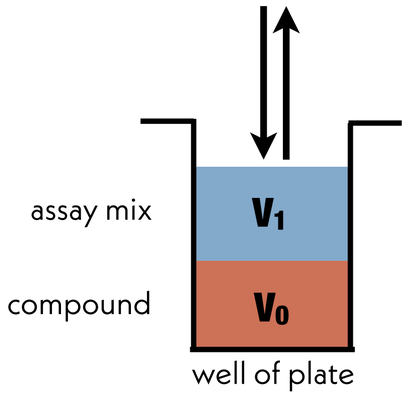
$\mathcal{D}$	data
θ	model parameters
$p(\theta\|\mathcal{D})$	posterior
$p(\mathcal{D}\|\theta)$	sampling distribution (model)
$p(\theta)$	prior

We need good **data likelihood models** that capture sources of experimental error or uncertainty

Where do we get the data likelihood functions?

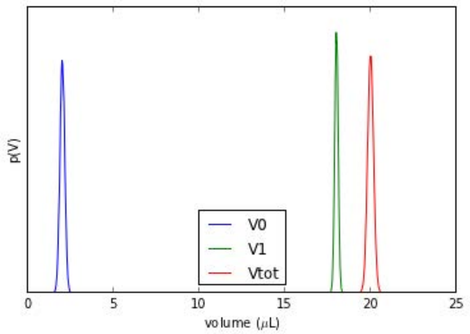
All experiments are contaminated with **experimental error** which introduces uncertainty into what we learn

fluorescence measurement



$$
\begin{aligned}
V_0 &= 2 \quad \mu L \\
V_1 &= 18 \quad \mu L \\
K_d &= 100 \quad nM \\
C_0 &= 10 \quad mM \\
C_1 &= 1 \quad \mu M
\end{aligned}
$$

imprecision and inaccuracy in pipetting

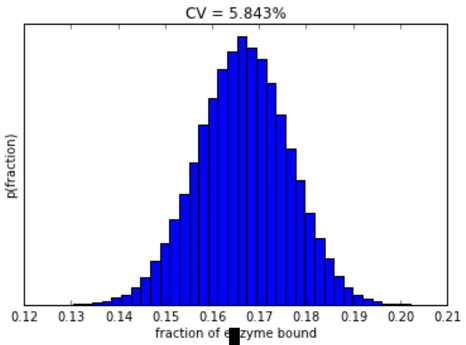


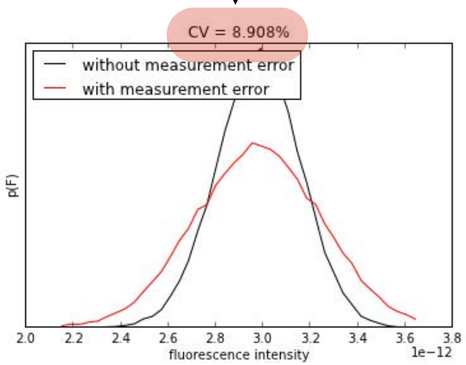
Biomek FX ***Pipetting Performance Specifications***

Span-8 Systems

Transfer Volume	Span-8 Syringe Volume	Tip Types	Accuracy ± %	Precision < %
0.5 µL	250 µL	P20, Fixed 60 mm	5	10
1 µL	250, 500, 1000 µL	P20, P50, Fixed 60 mm	3	7
5 µL	250, 500, 1000 µL	P20, Fixed 60 mm	3	5
10 µL	500 µL	P50, P250	3	5

variation in protein:ligand complex concentration

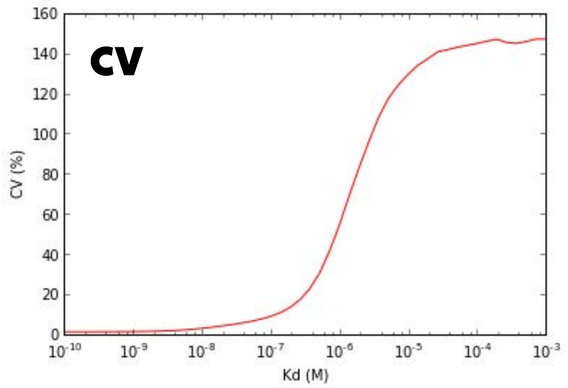




signal broadening due to measurement error

Is the expected CV small enough to be useful?

Modeling an experiment can ensure it will yield useful data or uncover unexpected issues

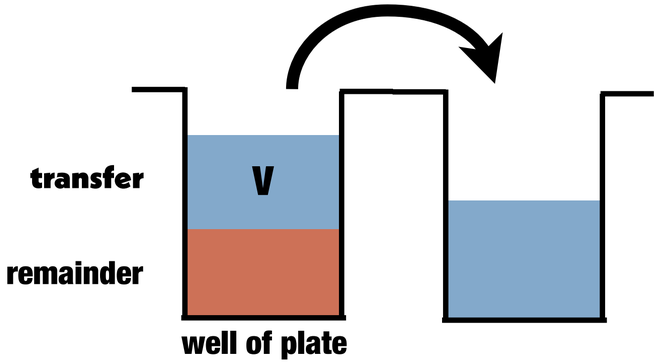


Hanson, Ekins, Chodera. JCAMD 29:1073, 2015.

https://github.com/choderalab/dispensing-errors-manuscript

Sources of error in individual assay steps can be explicitly modeled

pipetting error



$$V \sim \mathcal{N}(\mu, \sigma^2)$$

imprecision (CV)

inaccuracy (bias)

Hanson, Ekins, Chodera. JCAMD 29:1073, 2015.

Autoprotocol enables automated construction of **Bayesian graphical models** to describe the accumulation of experimental error

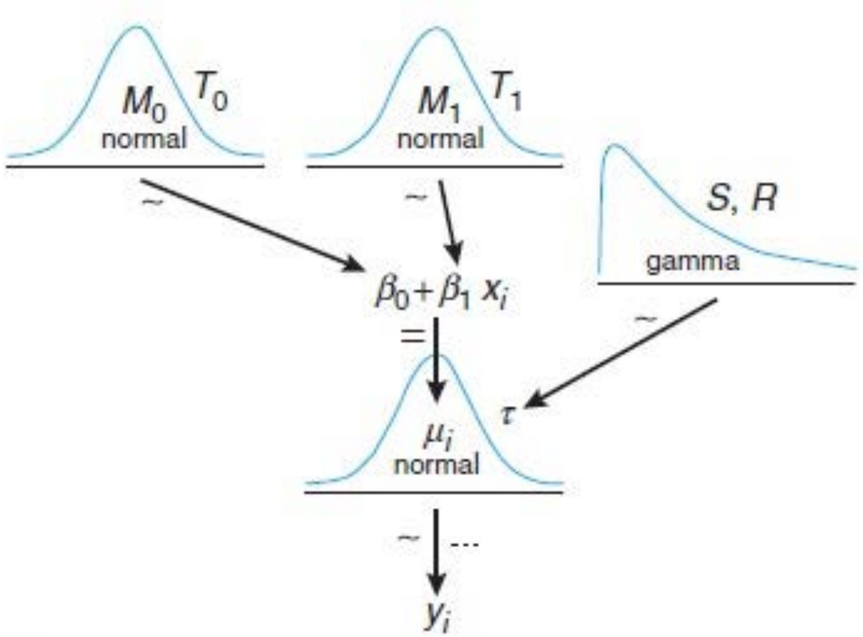


FIGURE 16.3

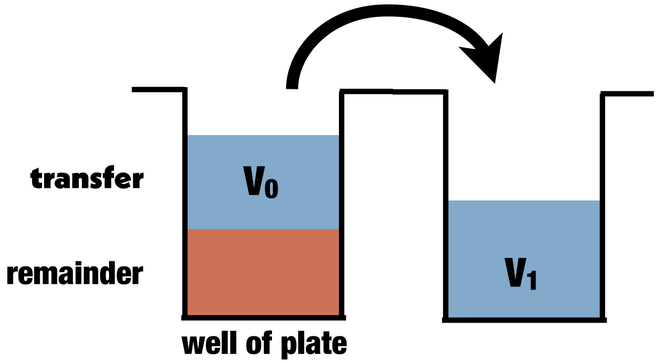
$$\beta_0 \sim \mathcal{N}(M_0, T_0)$$
$$\beta_1 \sim \mathcal{N}(M_1, T_1)$$
$$\tau \sim \Gamma(S, R)$$
$$y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \tau)$$

Figure from Doing Bayesian Data Analysis

Liquid transfers are modeled by **lognormal** distributions

autoprotocol:

```
protocol.transfer(plate.well('A1'), plate.well('A2'), ["10:microliter"])
```

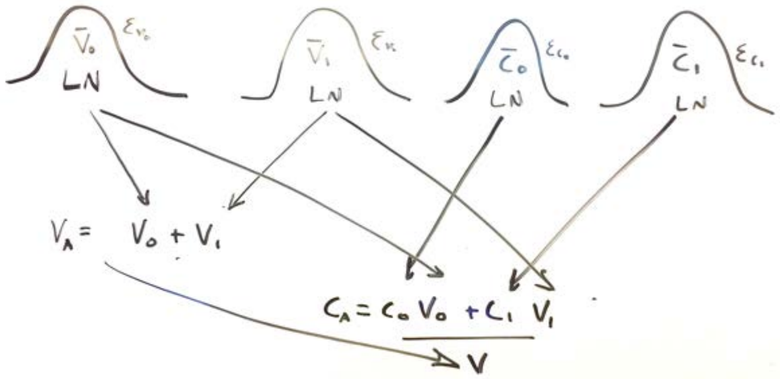


equations:

$$V_A = V_0 + V_1$$

$$C_A = \frac{C_0 V_0 + C_1 V_1}{V}$$

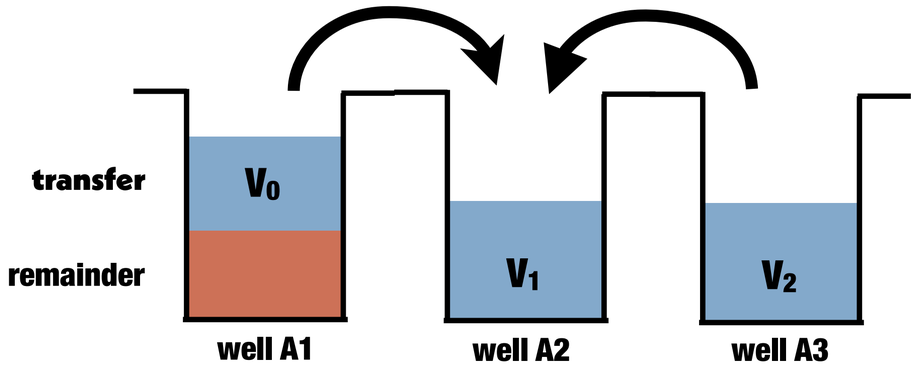
Bayesian graphical model:



Each operation **adds new nodes** to the graphical model

autoprotocol:

```
protocol.transfer(plate.well('A1'), plate.well('A2'), ["10:microliter"])
protocol.transfer(plate.well('A3'), plate.well('A2'), ["10:microliter"])
```



equations:

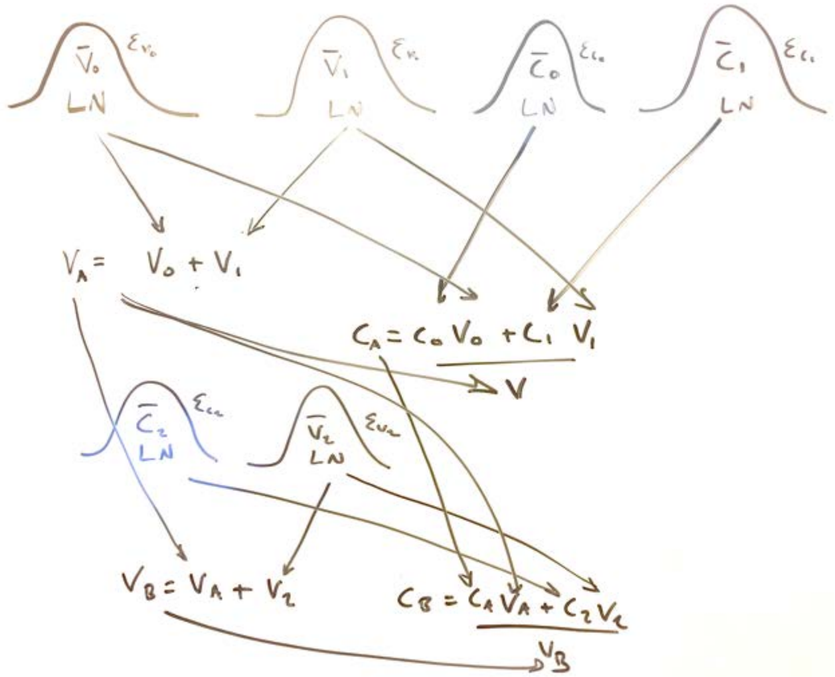
$$V_A = V_0 + V_1$$

$$C_A = \frac{C_0 V_0 + C_1 V_1}{V}$$

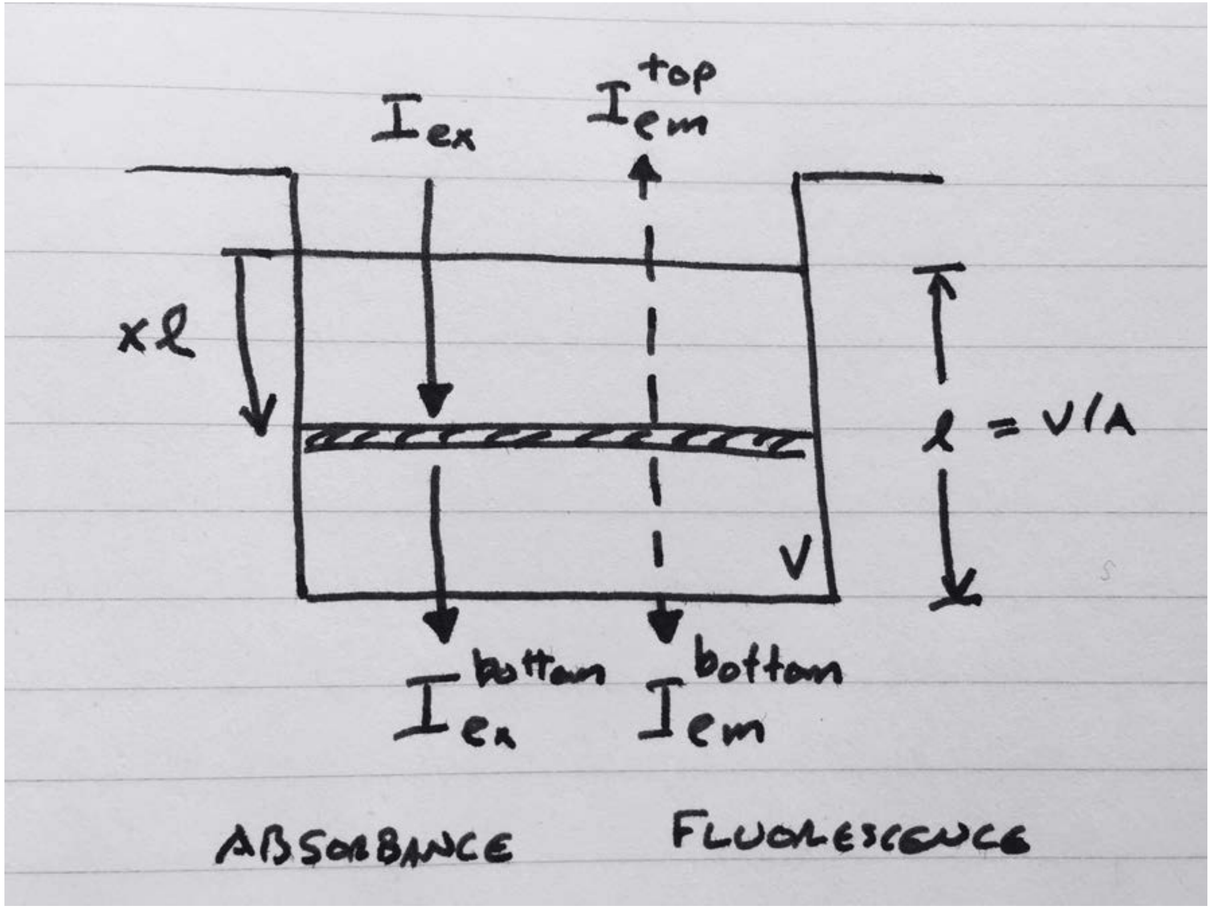
$$V_B = V_A + V_2$$

$$C_B = \frac{C_A V_A + C_2 V_2}{V_B}$$

Bayesian graphical model:



We can describe biophysical measurements using simple **forward models**




```
# read absorbance at 600 nanometers
p.absorbance(plate, plate.wells_from(0,4).indices(),
     "600:nanometer",
     dataref="absorbance")
```

$$A = 1 - e^{-\epsilon \cdot l \cdot [L]}$$

$$A^{\text{obs}} \sim N(A, \sigma^2_{\text{abs}})$$

```
# read fluorescence
p.fluorescence(plate, plate.wells_from(0,12),
                excitation="280:nanometer",
                emission="480:nanometer",
                dataref="fluorescence")
```

$$F_{\text{top}} = I_{ex}\left[\sum_i q_i(ex, em)[X_i] + lF_{\text{buffer}} + F_{\text{plate}}\right]$$

$$F^{\text{obs}}_{\text{top}} \sim N(f_{\text{top}}, \sigma^2_{\text{top}})$$

https://assaytools.readthedocs.io
https://github.com/choderalab/assaytools

Priors capture uncertainties in parameters or initial quantities

dispensed masses

transferred volumes

stock solution concentrations

extinction coefficients and quantum yields

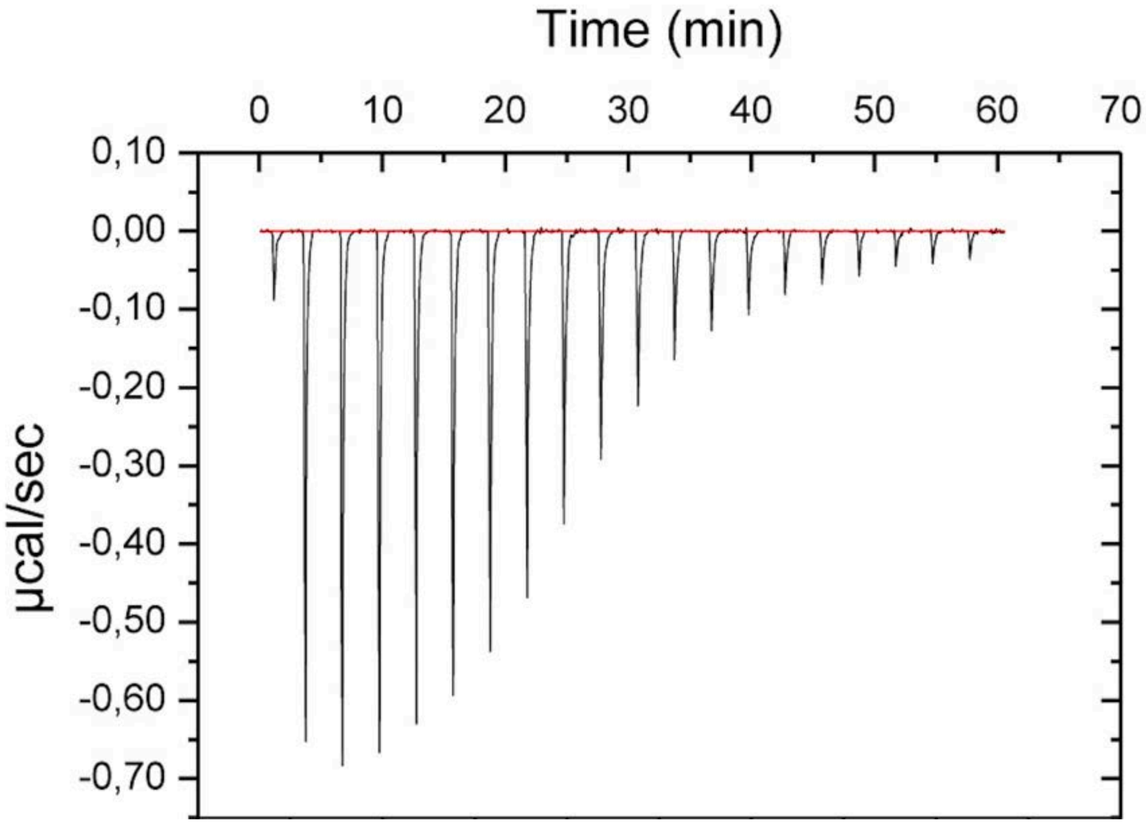
WE OFTEN NEED TO CHOOSE AMONG SEVERAL PHYSICAL BINDING MECHANISMS OR MODELS

simple 1:1 association

$$R + L \overset{K_a}{\leftrightharpoons} RL$$

sequential binding

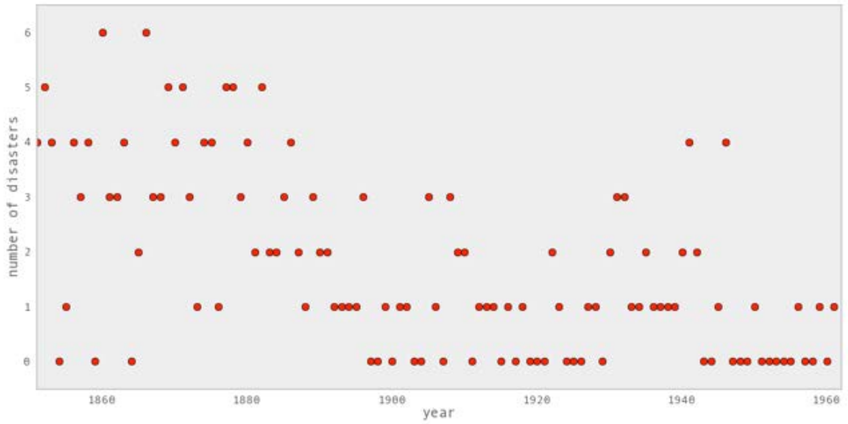
$$R + L \overset{K_1}{\leftrightharpoons} RL \overset{K_2}{\leftrightharpoons} RL_2 + \overset{K_3}{\leftrightharpoons} RL_3 \cdots$$



Bayesian model selection methods can do this in a principled way!

PROBABILISTIC PROGRAMMING LANGUAGES HAVE POWERFUL ABSTRACTIONS TO MAKE INFERENCE EASY

DATA



MODEL

$$(D_t|s,e,l) \sim \text{Poisson}\,(r_t)\,, \quad r_t = \begin{cases} e & \text{if} \quad t < s \\ l & \text{if} \quad t \geq s \end{cases}, \quad t \in [t_l, t_h]$$

$$s \sim \text{Discrete Uniform}(t_l, t_h)$$

$$e \sim \text{Exponential}(r_e)$$

$$l \sim \text{Exponential}(r_l)$$

- D_t: The number of disasters in year t.
- r_t: The rate parameter of the Poisson distribution of disasters in year t.
- s: The year in which the rate parameter changes (the switchpoint).
- e: The rate parameter before the switchpoint s.
- l: The rate parameter after the switchpoint s.
- t_l, t_h: The lower and upper boundaries of year t.
- r_e, r_l: The rate parameters of the priors of the early and late rates, respectively.

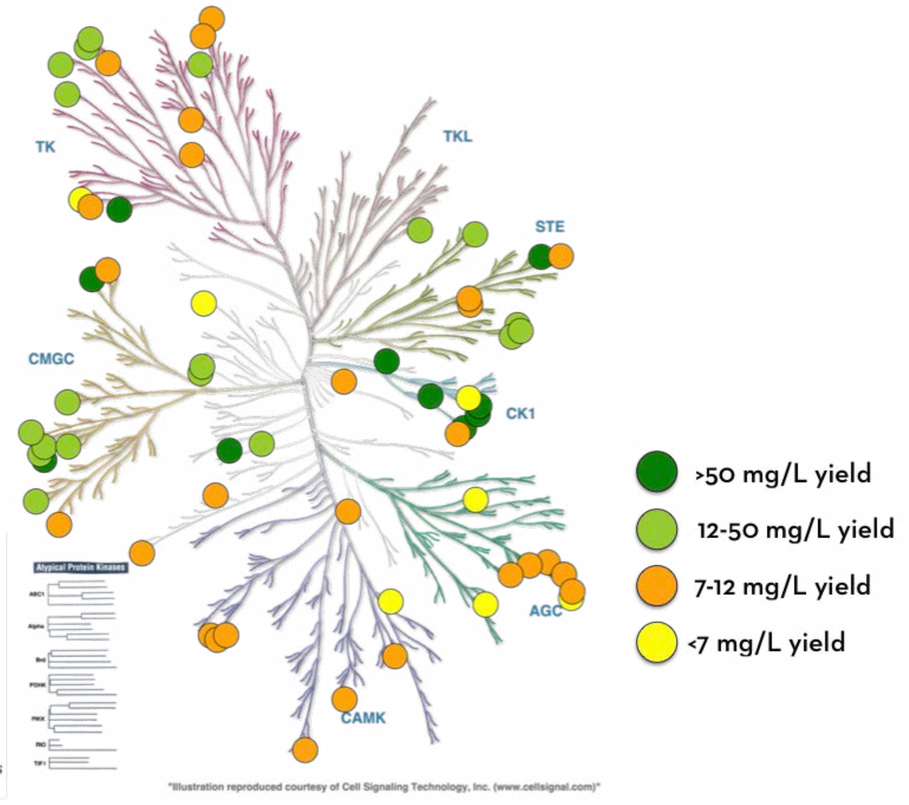
CODE

```
switchpoint = DiscreteUniform('switchpoint', lower=0, upper=110)
early_mean = Exponential('early_mean', beta=1.)
late_mean = Exponential('late_mean', beta=1.)
@deterministic(plot=False)
def rate(s=switchpoint, e=early_mean, l=late_mean):
    ''' Concatenate Poisson means '''
    out = np.empty(len(disasters_array))
    out[:s] = e
    out[s:] = l
    return out
disasters = Poisson('disasters', mu=rate, value=disasters_array, observed=True)
```

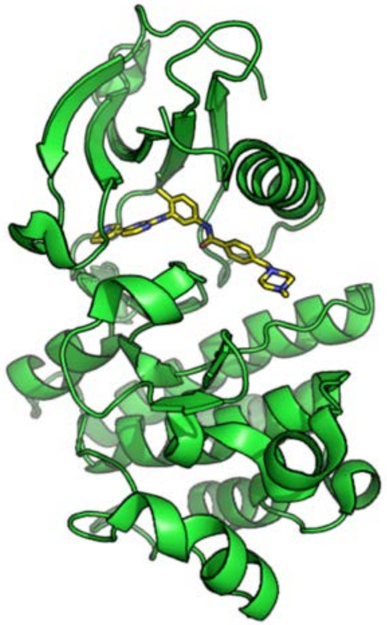
PyMC2: http://pymc-devs.github.io/pymc/tutorial.html#an-example-statistical-model

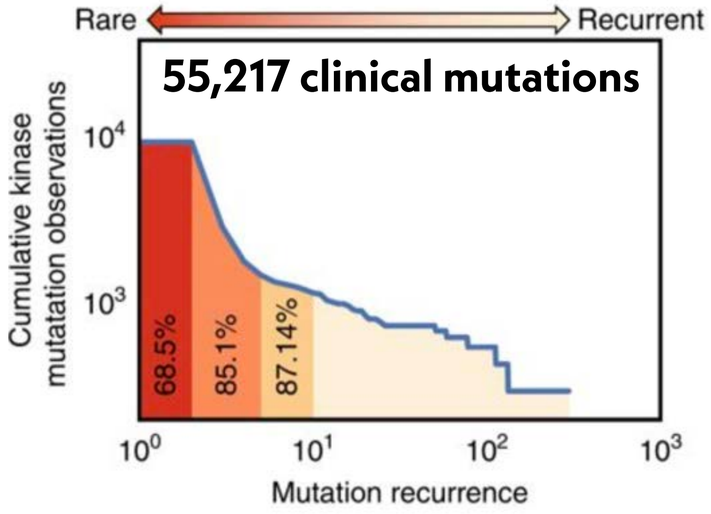
An example: Measuring kinase:inhibitor binding affinities

52 His-tagged kinase domain with good bacterial expression



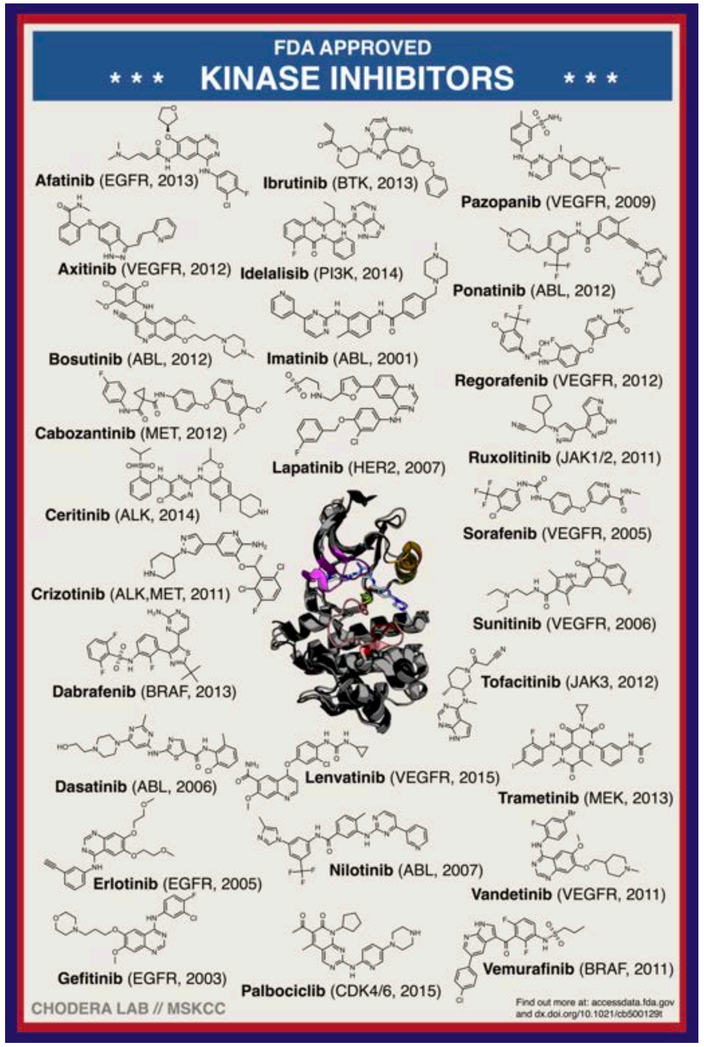
Biochemistry 57:4675, 2018
http://dx.doi.org/10.1021/acs.biochem.7b01081





Communications Biology 1:70, 2018
https://doi.org/10.1038/s42003-018-0075-x

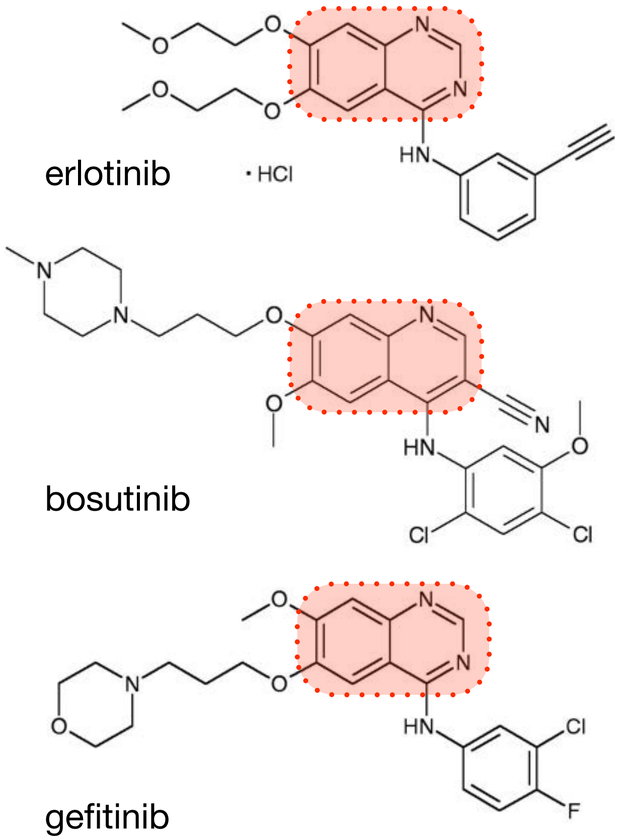
>46 FDA-approved inhibitors

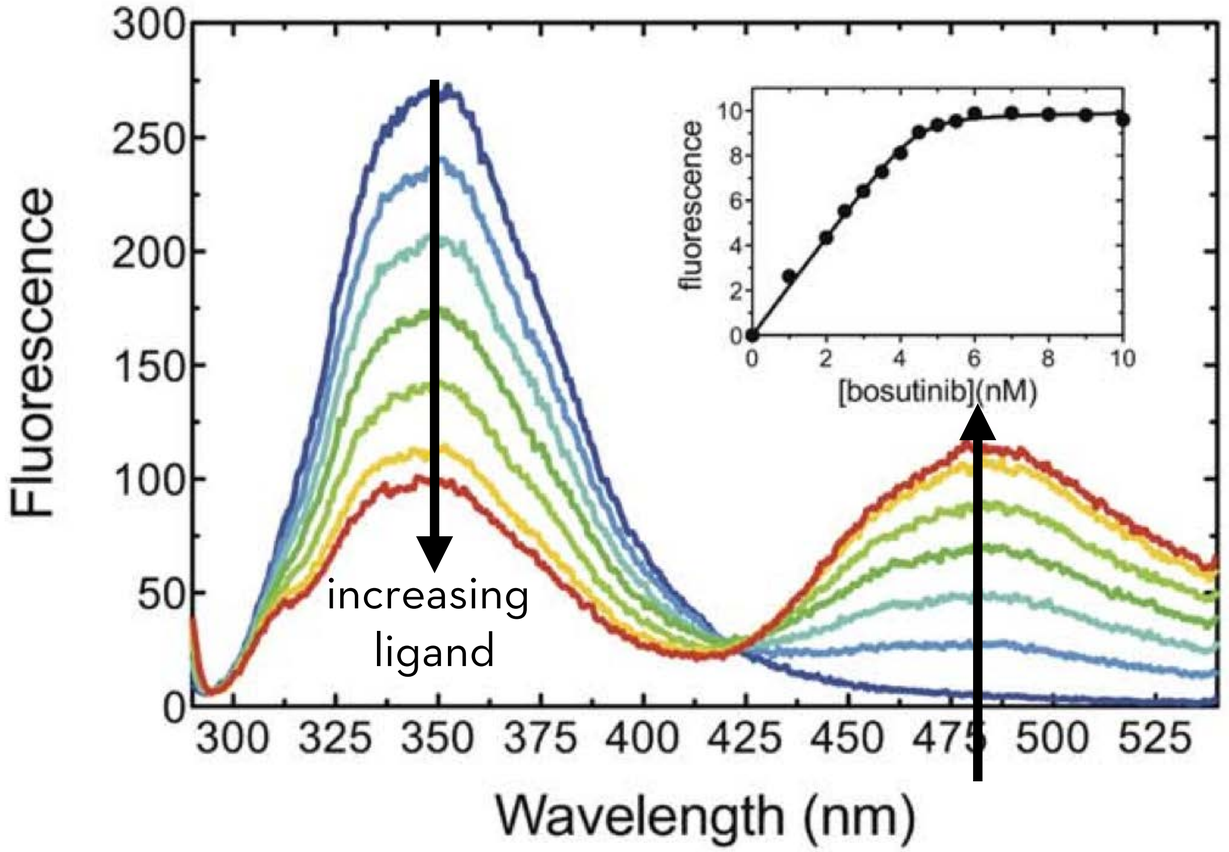


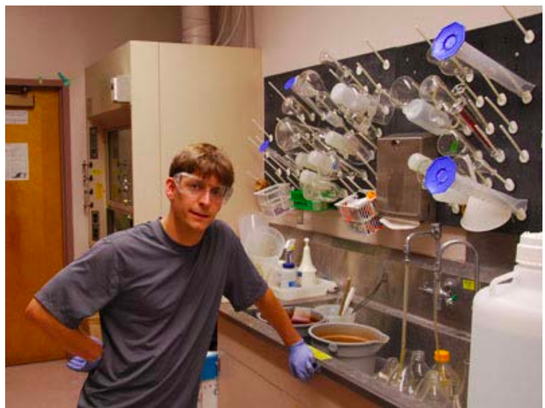
https://bit.ly/2EyX5c2

Many kinase inhibitors fluoresce strongly when they bind to a kinase ATP-binding site

fluorescent inhibitors

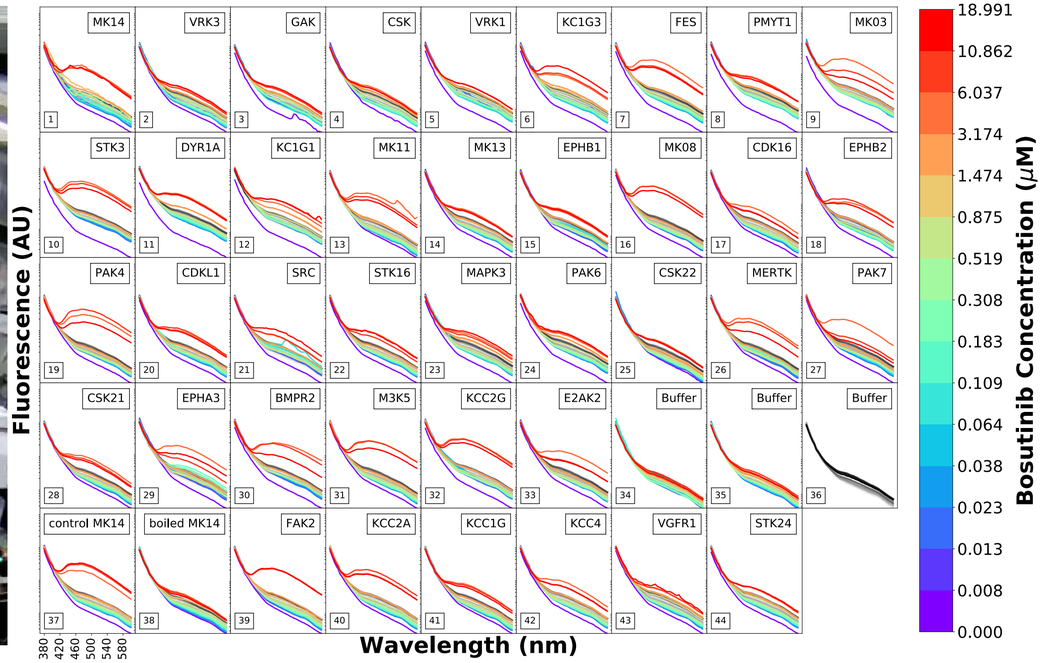






NICK LEVINSON
U MINNESOTA

Assay automation allows us to scale up to highly parallel binding affinity measurements



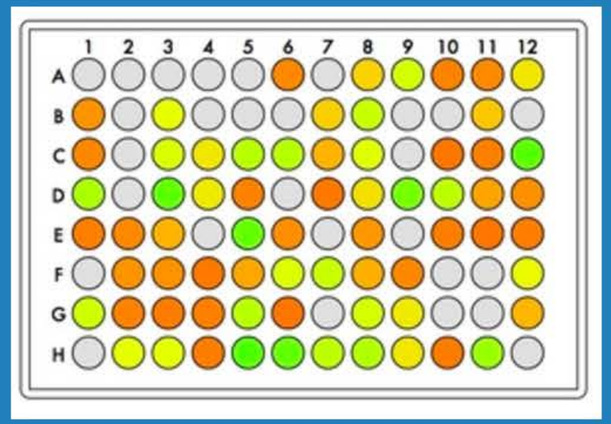
dispense **mix** **spin** **read**

We're scaling up automated measurement of kinase inhibitor binding affinities to clinical cancer mutations in human kinase domains from **cbioportal**

SONYA HANSON **LUCELENIE RODRIGUEZ** **MEHTAP ISIK** **ERIN GRUNDY** **STEVEN ALBANESE**

AssayTools: A toolkit for Bayesian analysis of binding assay data

assaytools



latest

Search docs

Installation

Theory

Analysis Functions

Docs » AssayTools

Edit on GitHub

AssayTools

Assay modeling and Bayesian analysis made easy.

AssayTools is a python library that allows users to model automated assays and analyze assay data using powerful Bayesian techniques that allow complete characterization of the uncertainty in fit models. With AssayTools, you can

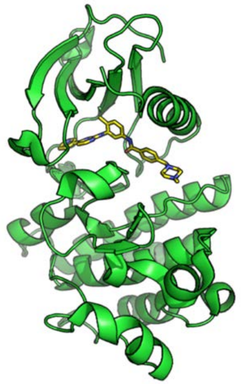
- Create models of experimental assays (e.g. fluorescence or absorbance assays of ligand binding from titration curves prepared by automated liquid handlers)
- Analyze data from these assays using powerful Bayesian techniques
- Derive parameters for these experimental assays from real data to use in modeling
- Model new assay configurations to determine expected accuracy and bias across a range of ligand affinities

http://github.com/choderalab/assaytools

http://assaytools.readthedocs.io

We can automate the construction of Bayesian graphical models from simple experiments

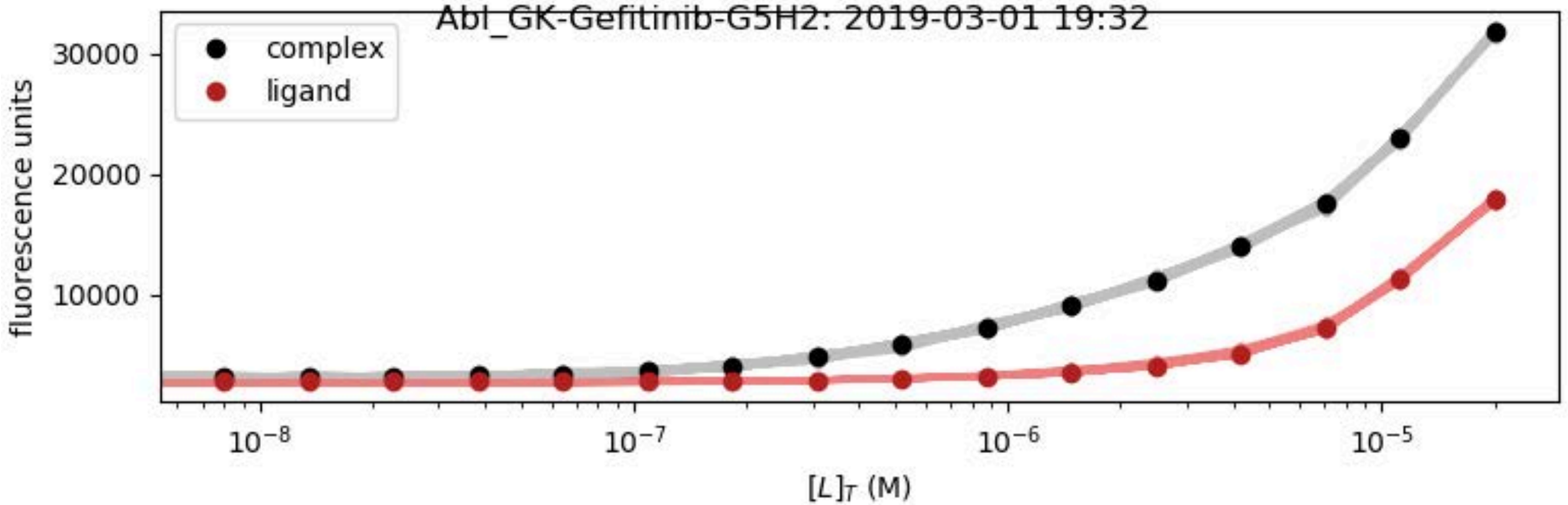
Abl T315I **gefitinib**



+

pymc2 graphical model

excitation @ 280nm
emission @ 480 nm



We can sample from the Bayesian posterior with MCMC

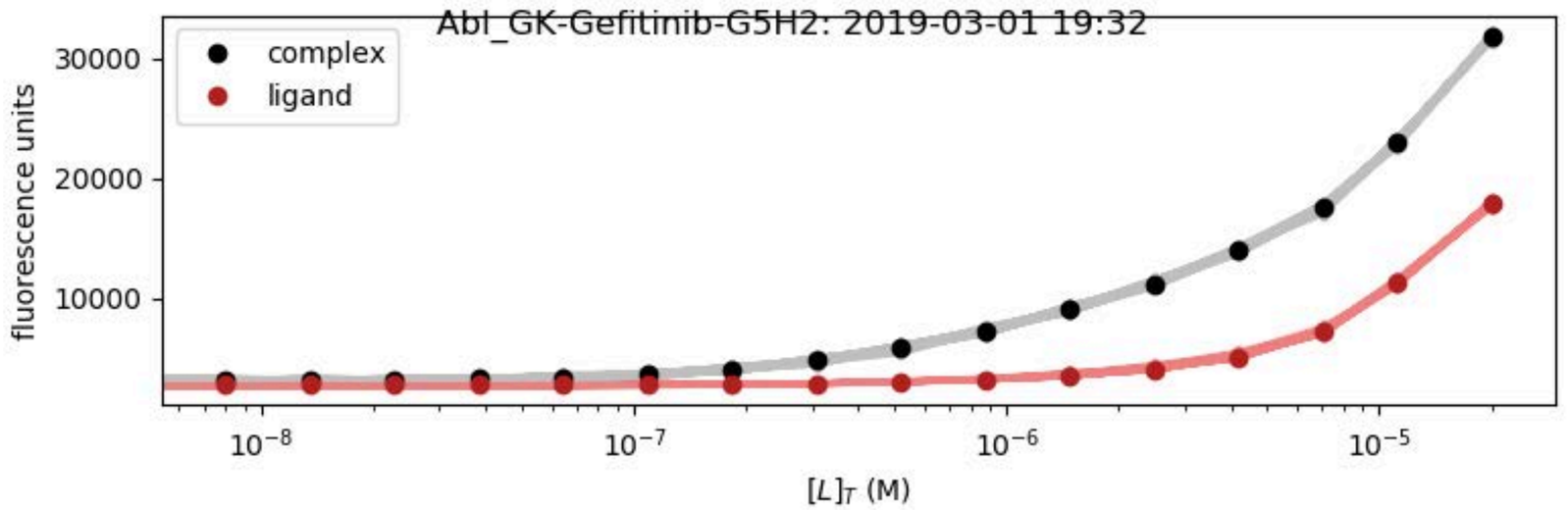
```
# Sample with MCMC
mcmc = pymc.MCMC(pymc_model, db='ram', name='Sampler', verbose=True)
mcmc.sample(iter=100000, burn=10000, thin=50, progress_bar=False)
```

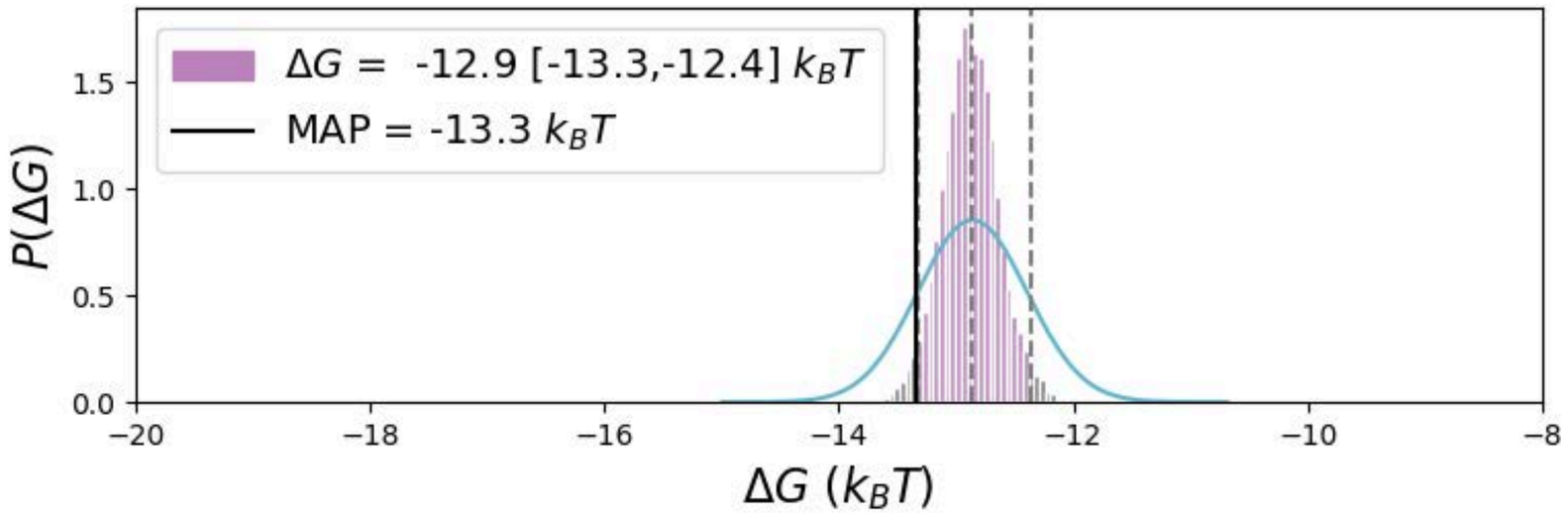
https://github.com/choderalab/assaytools

Sonya Hanson

WE CAN QUANTIFY PARAMETER UNCERTAINTY BY EXAMINING THE MARGINAL DISTRIBUTION OF THE PARAMETERS OF INTEREST

```
# Sample with MCMC
mcmc = pymc.MCMC(pymc_model, db='ram', name='Sampler', verbose=True)
mcmc.sample(iter=100000, burn=10000, thin=50, progress_bar=False)
```

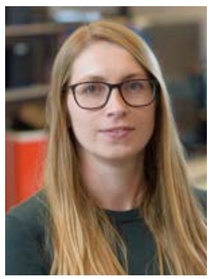




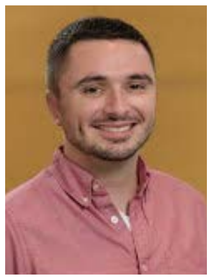
ERIN GRUNDY

MEHTAP ISIK

SONYA HANSON



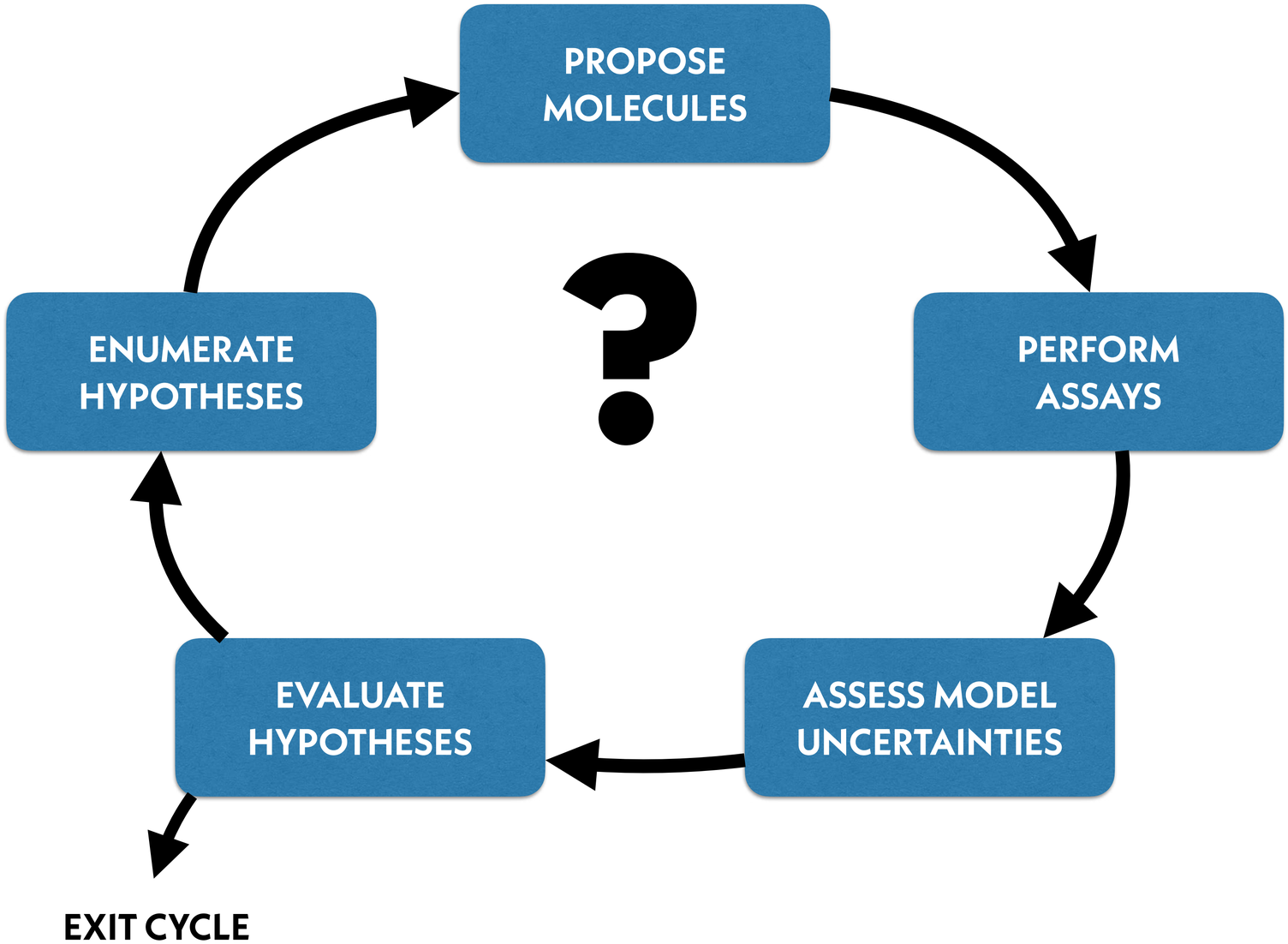
STEVEN ALBANESE



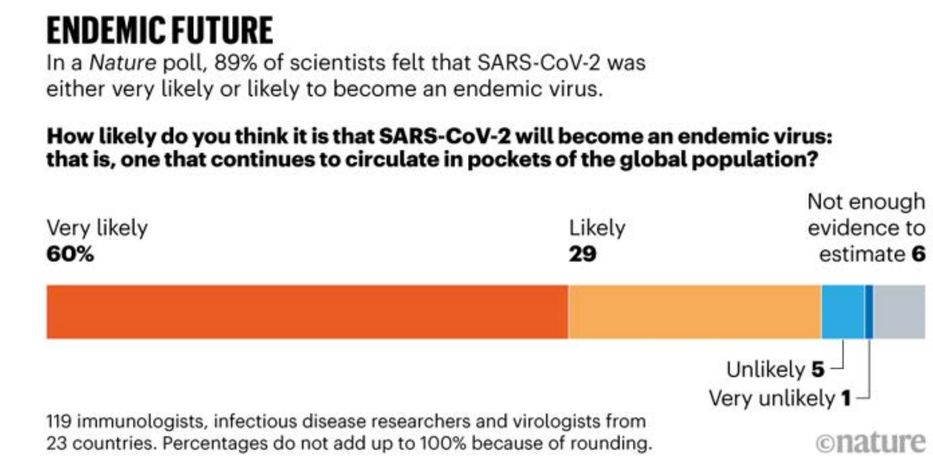
BAYESIAN MODELS OPEN UP NEW POSSIBILITIES FOR AUTOMATED INTERPRETATION OF DATA

automated characterization of **confidence intervals**
inference of **latent variables** (e.g. extinction coefficients)
Bayesian model selection / **hypothesis testing**
automated design of optimal follow-up experiments
feed **knowledge with uncertainty** into our predictive models

DO WE REALLY NEED ALL THIS?



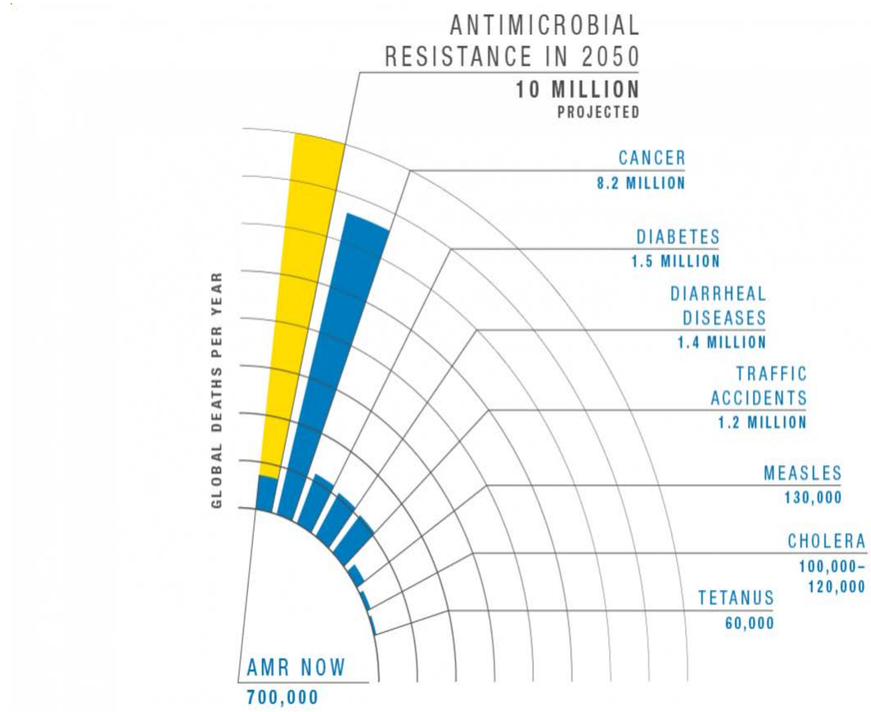
CURRENT DRUG DISCOVERY TIMELINES ARE INCOMPATIBLE WITH THE SURVIVAL OF HUMAN CIVILIZATION



https://doi.org/10.1038/d41586-021-00396-2

global pandemics

antimicrobial resistance



https://www.ucsf.edu/magazine/infectious-disease

WORLD NEWS FEBRUARY 20, 2021 / 7:03 AM / UPDATED A DAY AGO

Russia reports world's first case of human infection with H5N8 bird flu

By Polina Devitt, Gabrielle Tétrault-Farber 3 MIN READ

MOSCOW (Reuters) - Russia has registered the first case of a strain of bird flu virus named A(H5N8) being passed to humans from birds and has reported the matter to the World Health Organization (WHO), Anna Popova, head of consumer health watchdog Rospotrebnadzor, said on Saturday.

Reuters

A Threat to the Global Economy

What can motivate countries to focus on AMR and deploy the comprehensive response this threat demands? Often, economic interest provides an impetus to political action. In this report, we have used World Bank economic simulation tools to quantify the losses that AMR may inflict on the global economy between now and 2050.

Our simulations included two scenarios, corresponding to low AMR impacts and high AMR impacts. AMR impacts were modeled as shocks to labor supply and livestock productivity—a conservative approach that underestimates AMR's full economic effects.

In the optimistic case of low AMR impacts, the simulations found that, by 2050, annual global gross domestic product (GDP) would likely fall by 1.1 percent, relative to a base-case scenario with no AMR effects; the GDP shortfall would exceed $1 trillion annually after 2030. In the high AMR-impact scenario, the world will lose 3.8 percent of its annual GDP by 2050, with an annual shortfall of $3.4 trillion by 2030 (Figure ES1).

Drug-resistant infections: A Threat to Our Economic Future. World Bank Group.

CHODERA LAB

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