REDESIGNING DRUG DESIGN

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

Memorial Sloan Kettering Cancer Center

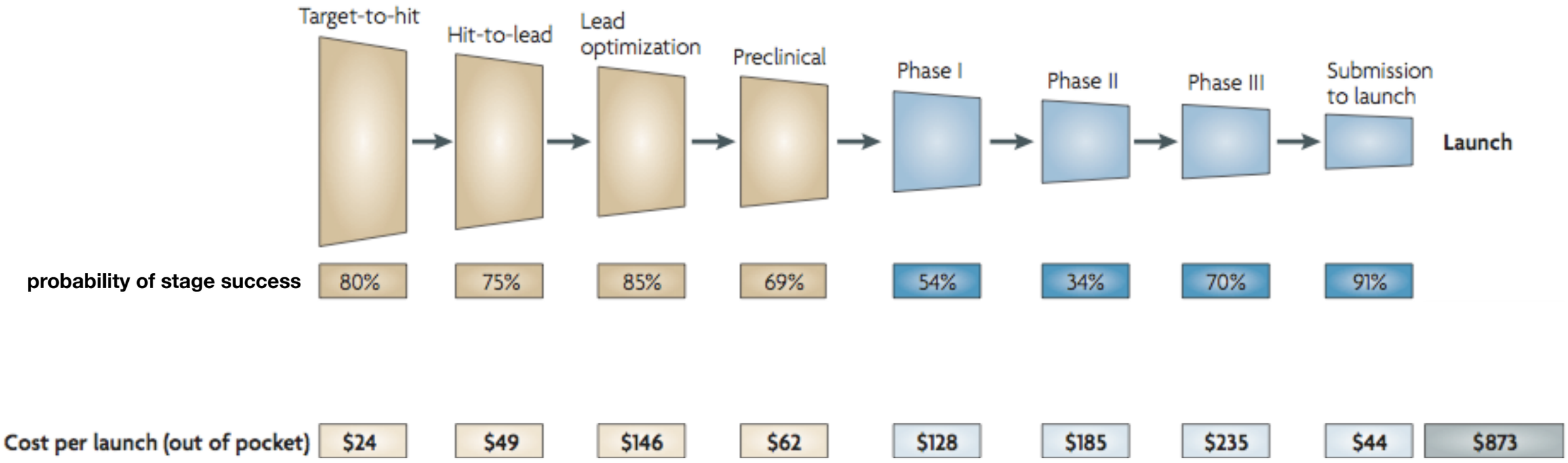
NYSBDG - New York Academy of Sciences - 21 Jan 2015

There is a crisis in drug discovery

Total pharma research spending has doubled to $65.3B over 2000-2010

Number of new molecular entities approved by FDA 2005-2009 is half that from previous five years

Number of truly innovative new molecular entities has remained constant at 5-6/year



Paul et al. Nat. Rev. Drug Discover. 9:203, 2010.
Chodera et al. Curr. Opin. Struct. Biol., 21:150, 2011.

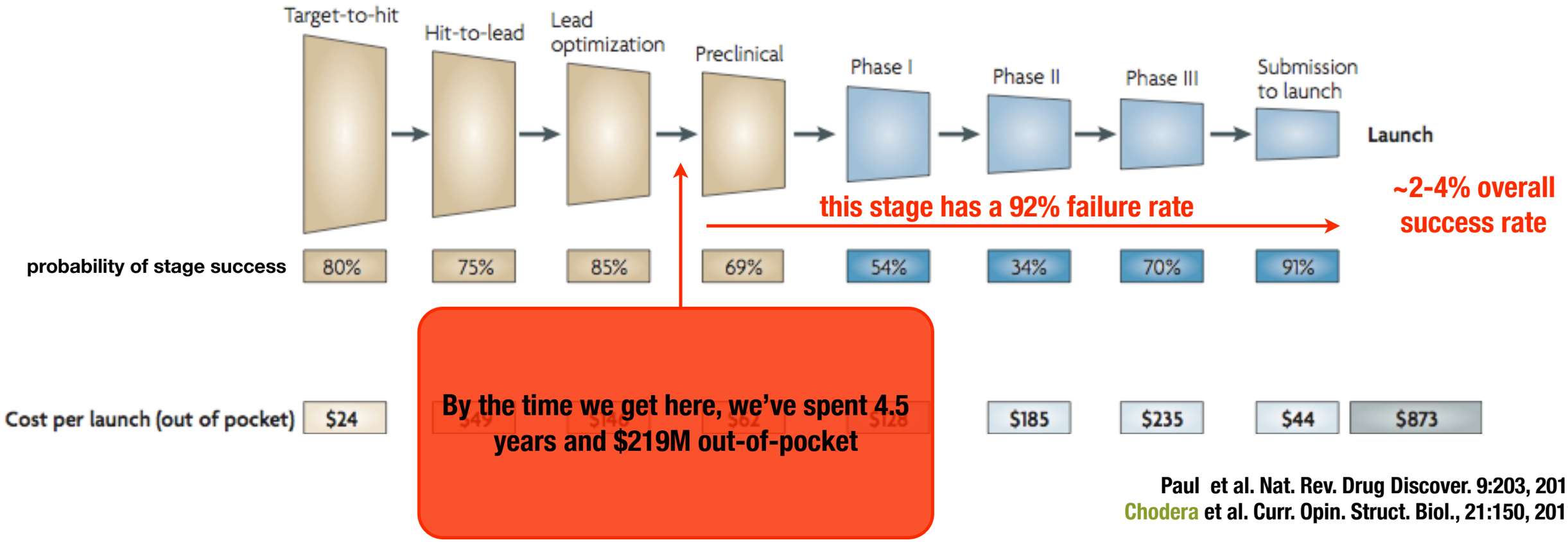
Complex design objectives (efficacy, selectivity, ADME-Tox) make design problem especially complex.

There is a crisis in drug discovery

Total pharma research spending has doubled to $65.3B over 2000-2010

Number of new molecular entities approved by FDA 2005-2009 is half that from previous five years

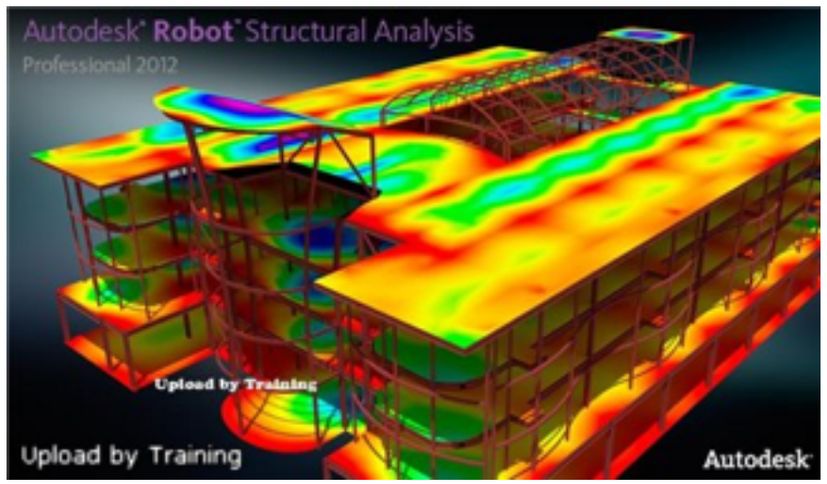
Number of truly innovative new molecular entities has remained constant at 5-6/year



Paul et al. Nat. Rev. Drug Discover. 9:203, 2010.
Chodera et al. Curr. Opin. Struct. Biol., 21:150, 2011.

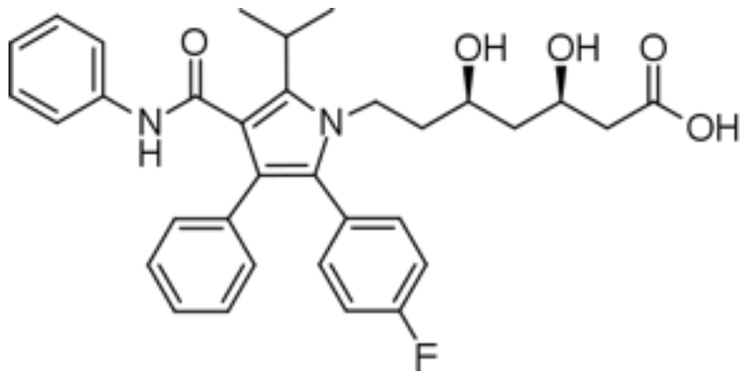
Complex design objectives (efficacy, selectivity, ADME-Tox) make design problem especially complex.

We regularly design planes, bridges, and buildings on computers to satisfy complex design objectives



10^3 - 10^6 parts

Why not small molecule drugs?

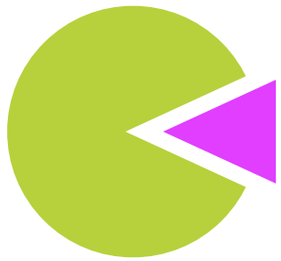


< 10^2 atoms

How can we bring drug design into the 21st century?

To design a small molecule with intended effects, we must predict how it will modulate cellular pathways

To design a small molecule with intended effects, we must predict how it will modulate cellular pathways



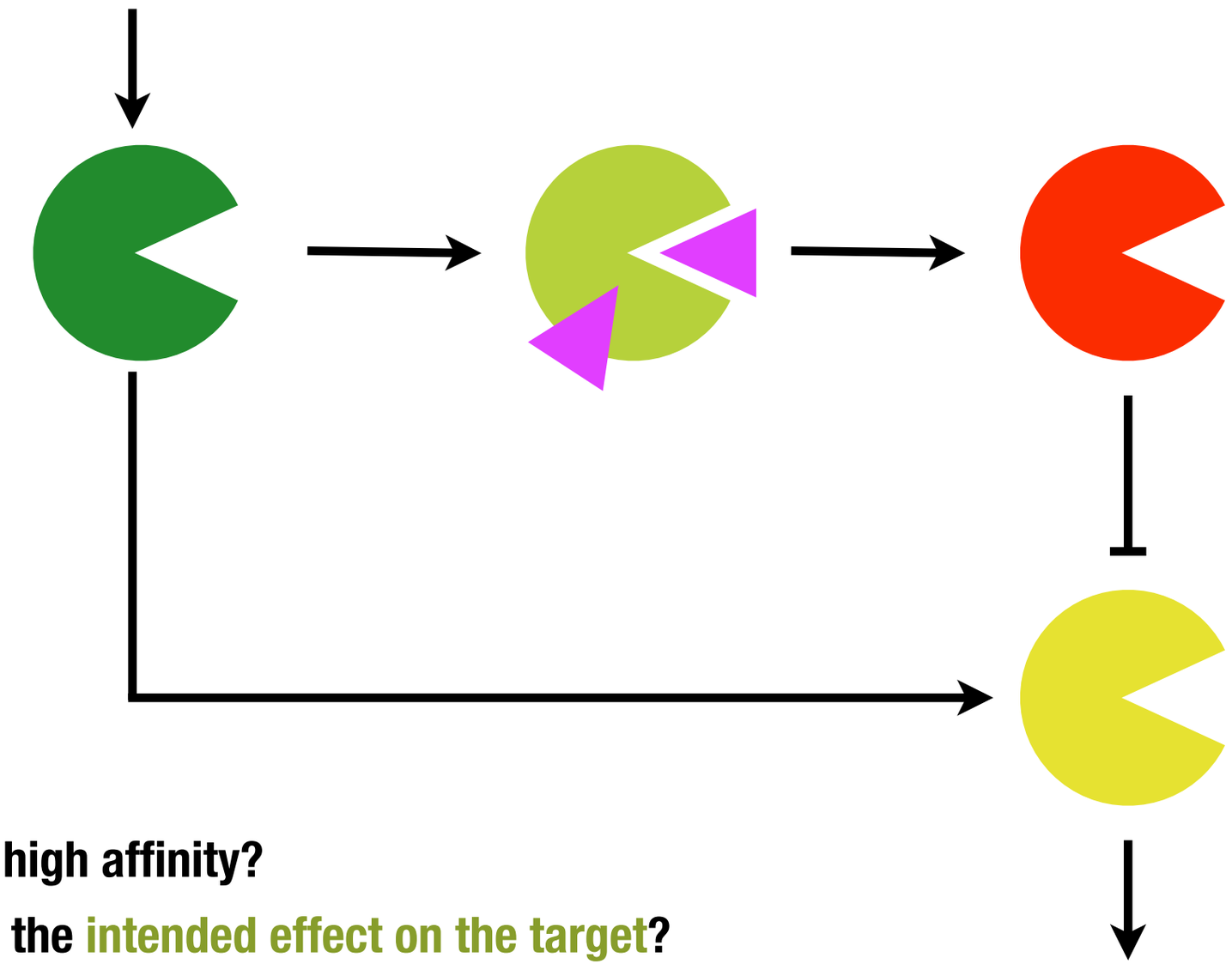
Will it bind the target with high affinity?

To design a small molecule with intended effects, we must predict how it will modulate cellular pathways

Will it bind the target with high affinity?

Will its binding mode have the intended effect on the target?

To design a small molecule with intended effects, we must predict how it will modulate cellular pathways

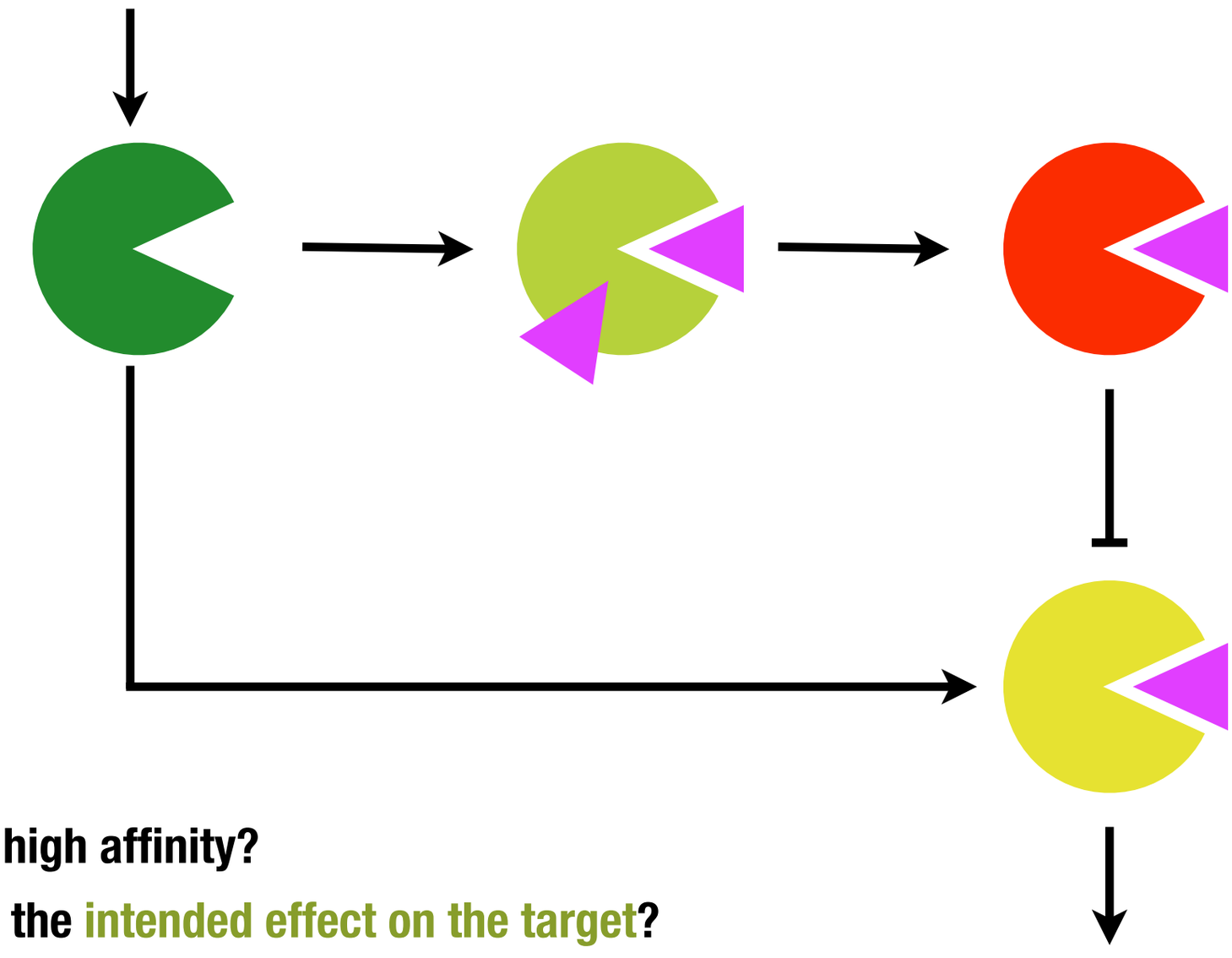


Will it bind the target with high affinity?

Will its binding mode have the intended effect on the target?

Does it produce the desired effect on cellular pathways?

To design a small molecule with intended effects, we must predict how it will modulate cellular pathways



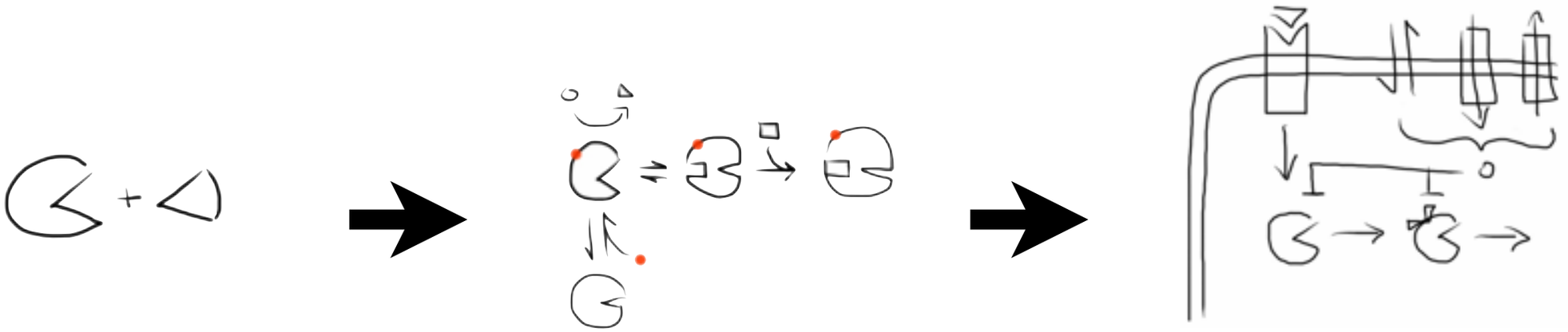
Will it bind the target with high affinity?

Will its binding mode have the intended effect on the target?

Does it produce the desired effect on cellular pathways?

Will it bind unintended targets? Are the resulting effects unacceptably toxic?

Multiscale physical models based on statistical mechanics can potentially aid small-molecule design efforts



physical binding constant

K_d

catalytic life cycle

apparent K_i

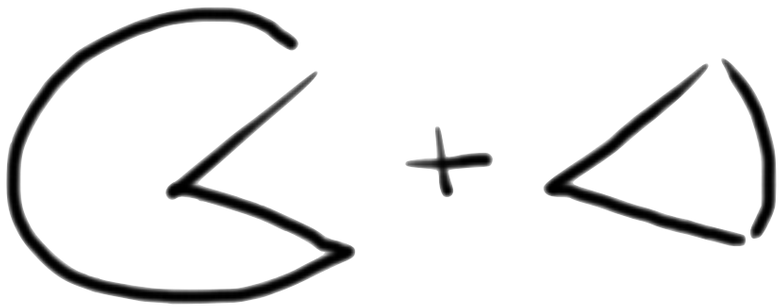
cellular pathways

EC50

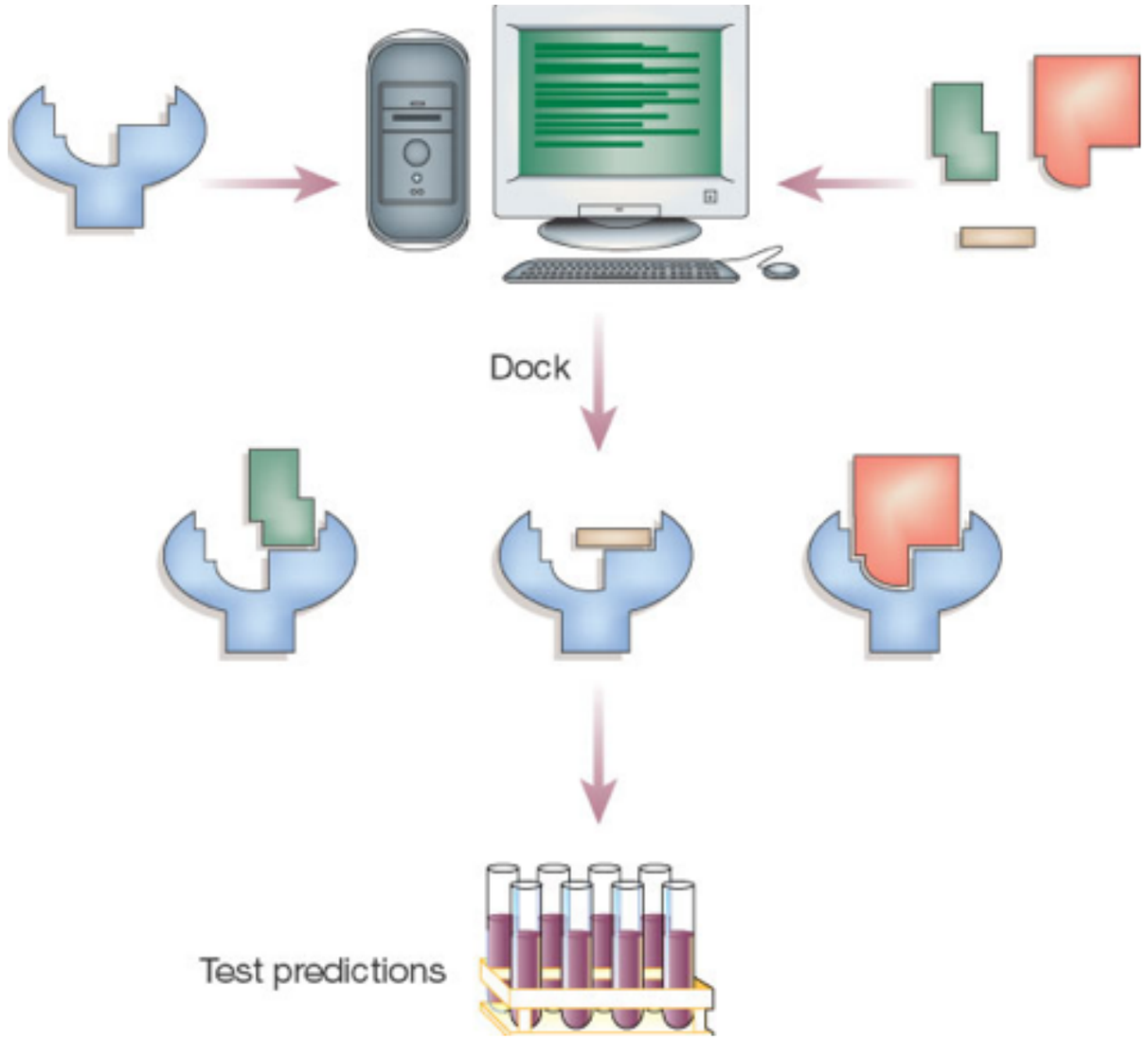
We use physical modeling and statistical mechanics to build predictive models at each of these scales.

How can we compute binding affinities for molecules that have yet to be synthesized or tested?

$$P + L \overset{K_d}{\rightleftharpoons} PL$$



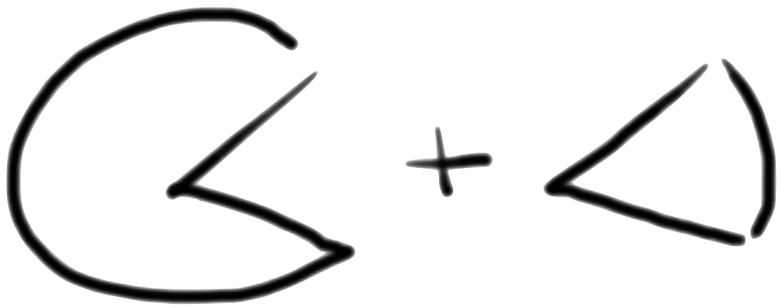
Virtual screening methods are in widespread use in drug discovery efforts today. They must work well, right?



Shoichet BK. Nature 432:862, 2004.

How can we compute binding affinities for molecules that have yet to be synthesized or tested?

$$P + L \overset{K_d}{\rightleftharpoons} PL$$



Virtual screening methods are in widespread use in drug discovery efforts today. They must work well, right?

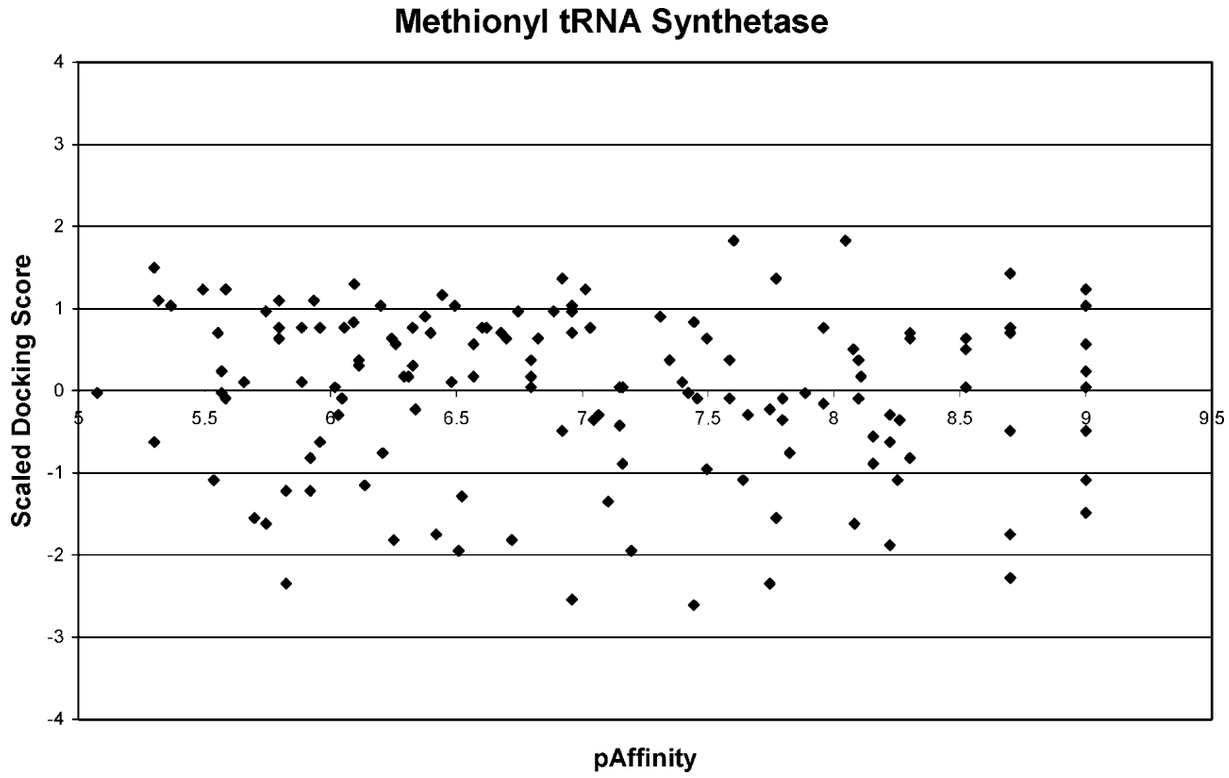
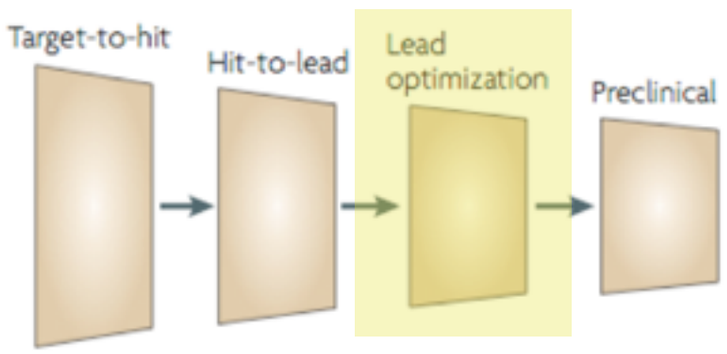


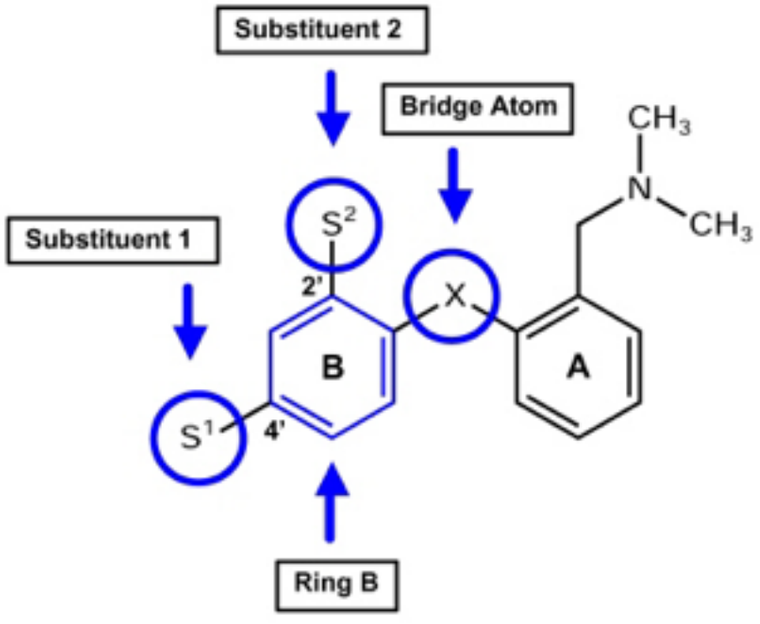
Figure 11. Plot of scaled score vs pAffinity for MRS and PPARδ. While the calculated correlation coefficient for the data shown for MRS is $r = -0.28$, this plot clearly demonstrates that these values are meaningless. No useful correlation exists between the docking score and compound affinity.

"For prediction of compound affinity, none of the docking programs or scoring functions made a useful prediction of ligand binding affinity."

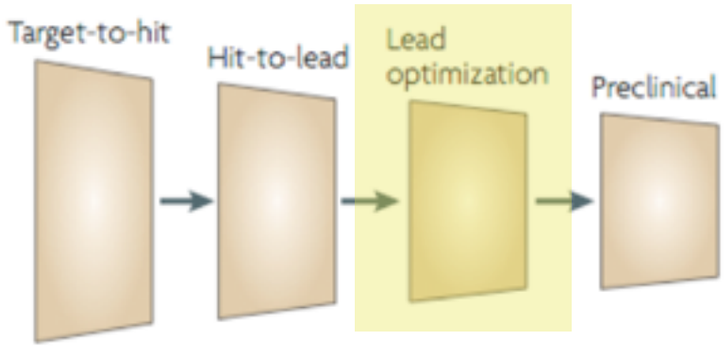
From Warren et al. [GSK], "A Critical Assessment of Docking Programs and Scoring Functions", J. Med. Chem. 49:5912, 2006.

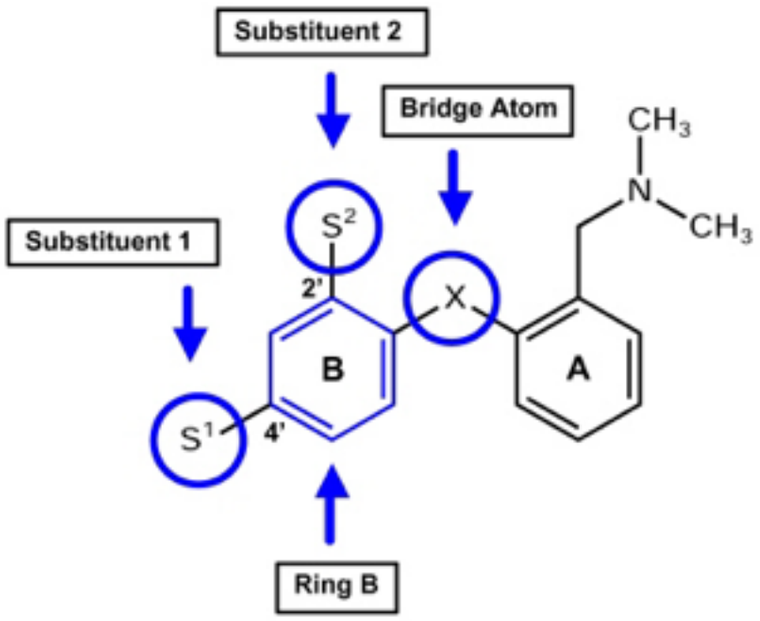
How accurate does a model need to be to aid rational drug design?

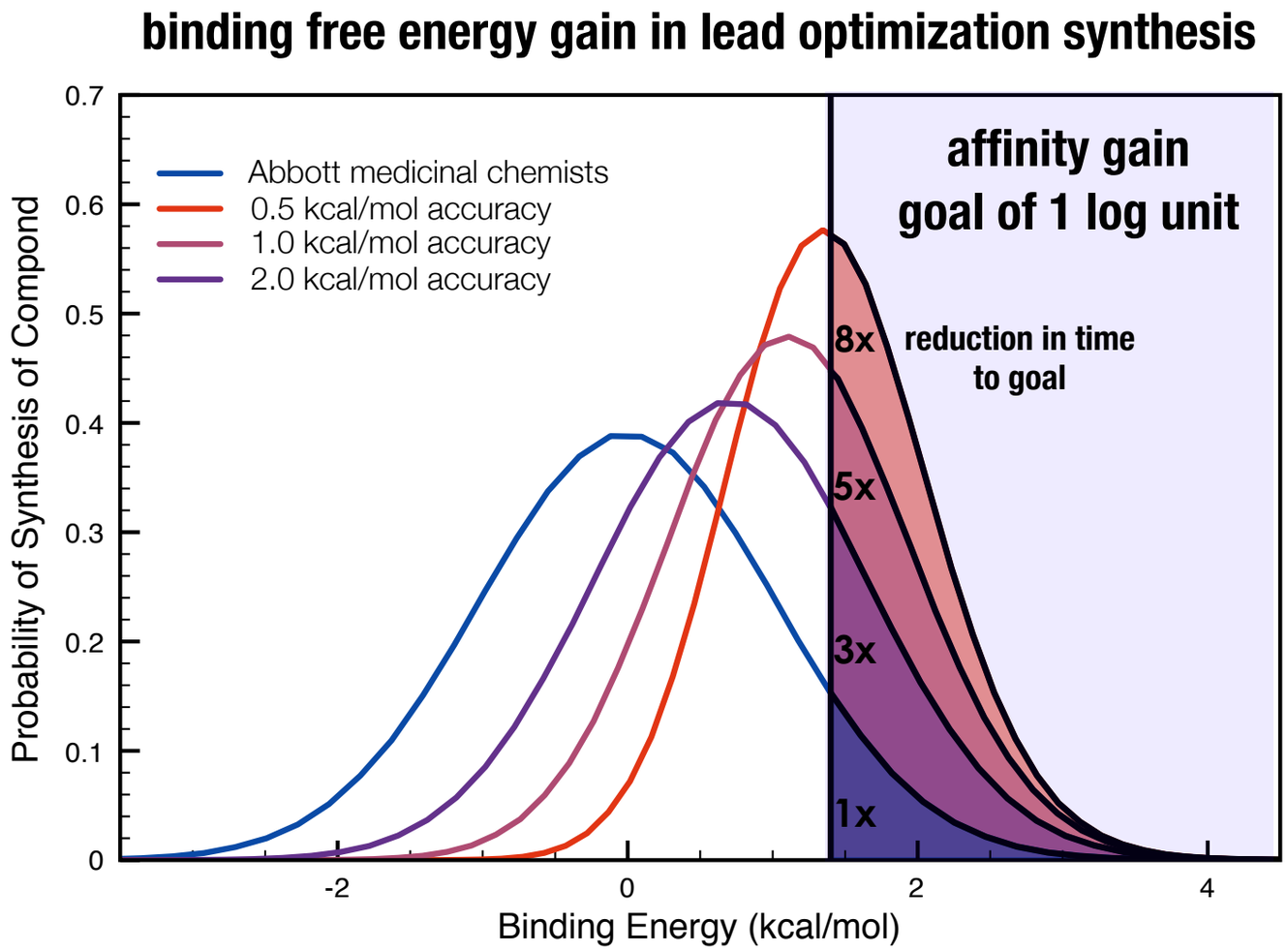




How accurate does a model need to be to aid rational drug design?



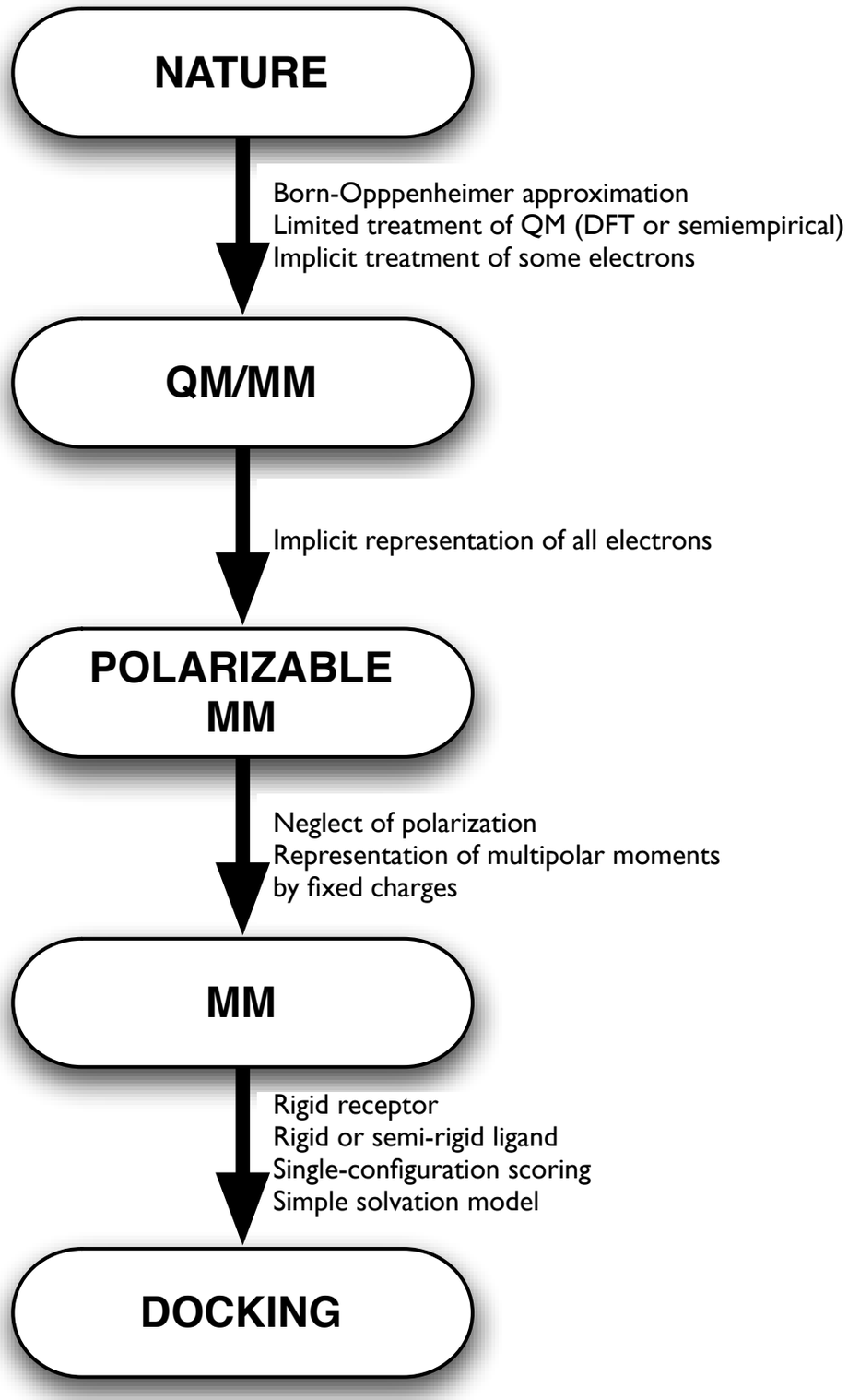




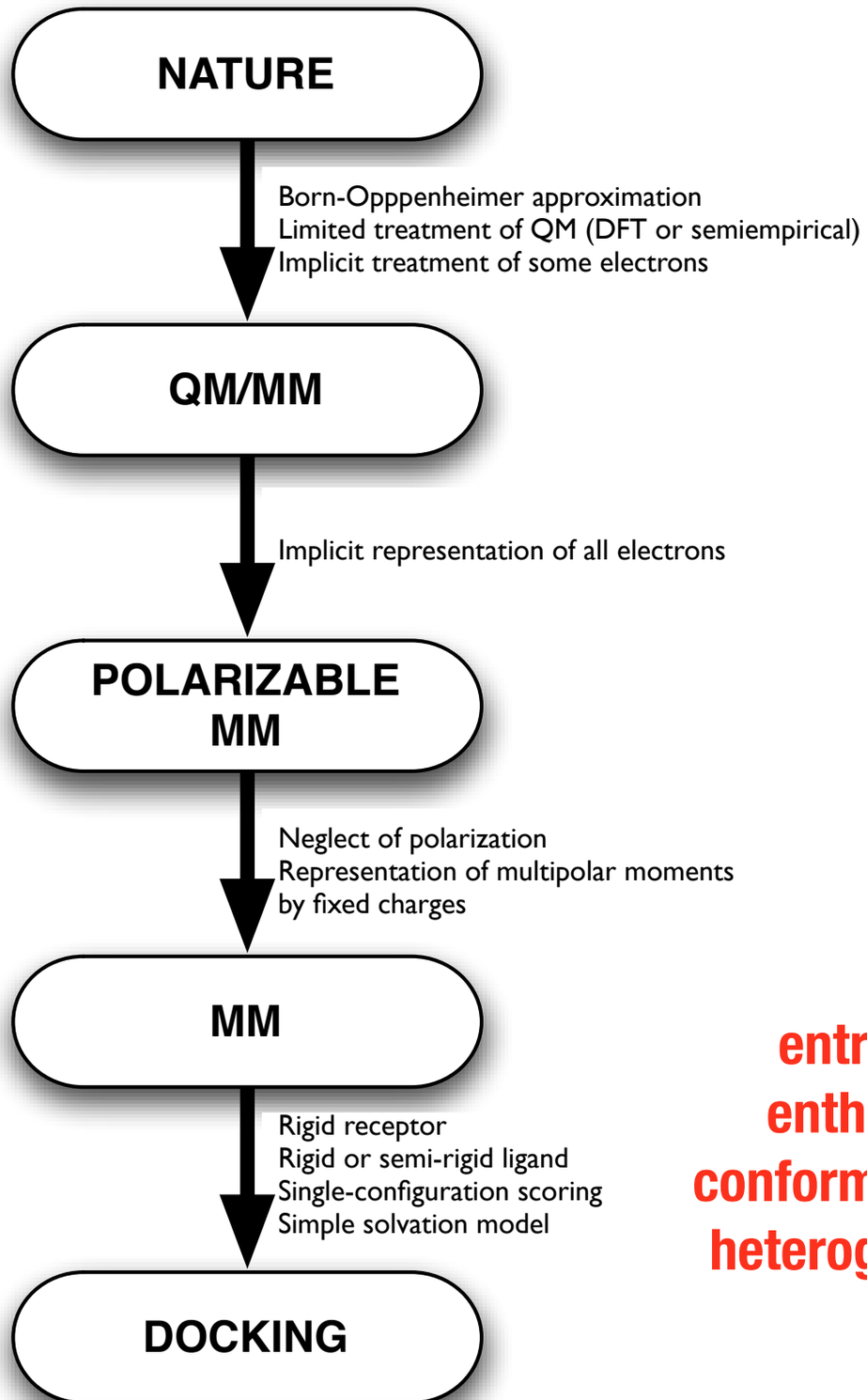
A 2 kcal/mol error in prioritizing lead synthesis would speed lead optimization by 3x but even 10% improvements would be of tremendous benefit

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.

What details are crucial for useful accuracy?
Virtual screening involves a number of approximations

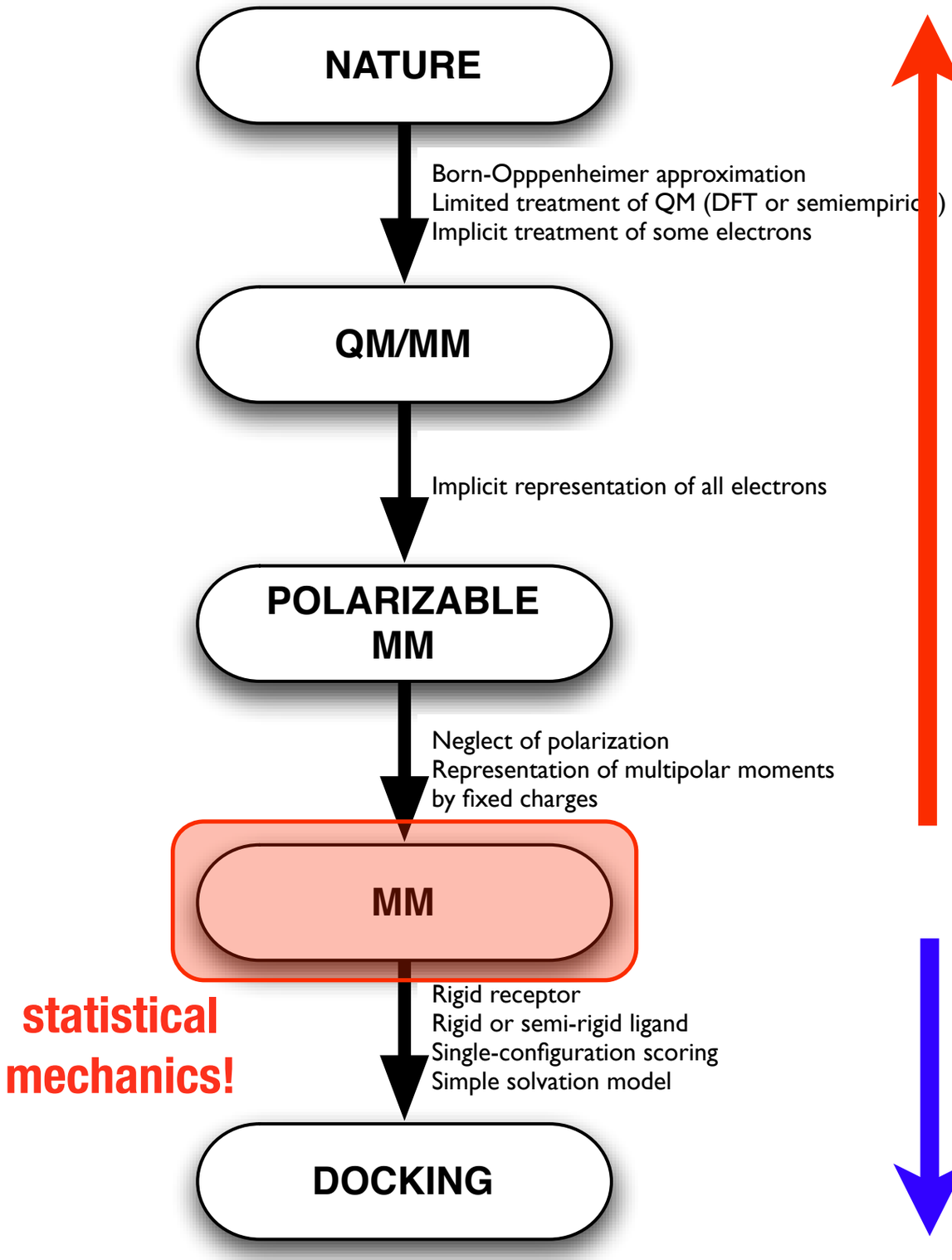


What details are crucial for useful accuracy?
Virtual screening involves a number of approximations



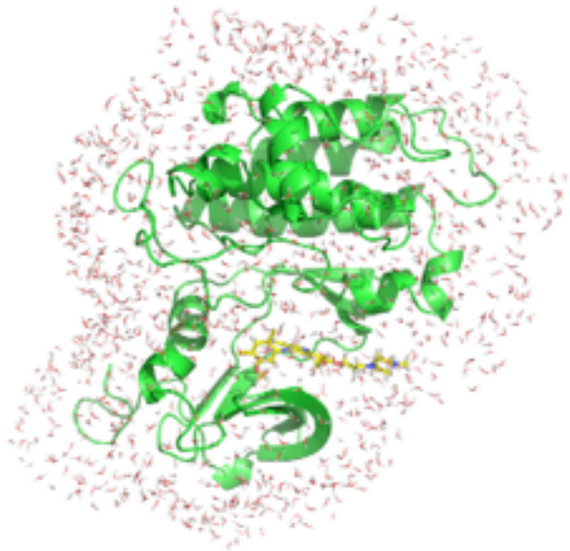
entropy
enthalpy
conformational
heterogeneity

What details are crucial for useful accuracy?
Virtual screening involves a number of approximations



statistical mechanics!

if insufficiently accurate, add detail in perturbative manner



molecular mechanics potential energy function (e.g. AMBER)

$$V(\mathbf{q}) = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2$$

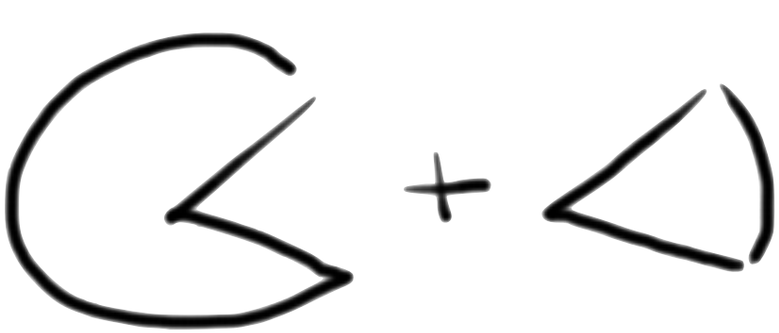
$$+ \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

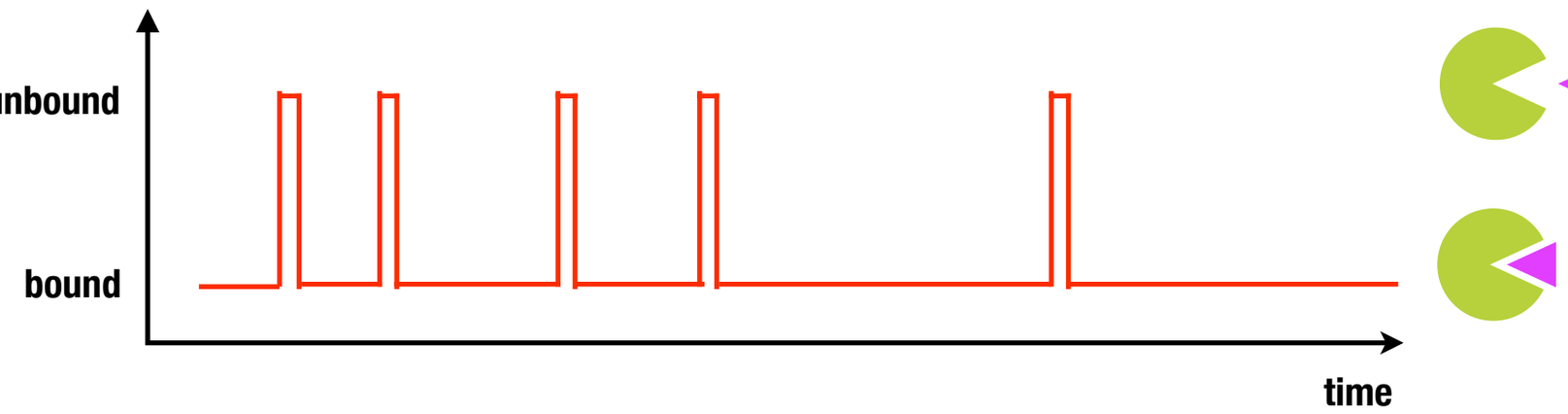
if accurate enough, systematically remove detail

How can we compute binding affinities that include all relevant statistical mechanical effects?

In principle, we could watch many binding/unbinding events to estimate a binding affinity

$$\mathrm{P} + \mathrm{L} \overset{K_d}{\rightleftharpoons} \mathrm{PL}$$





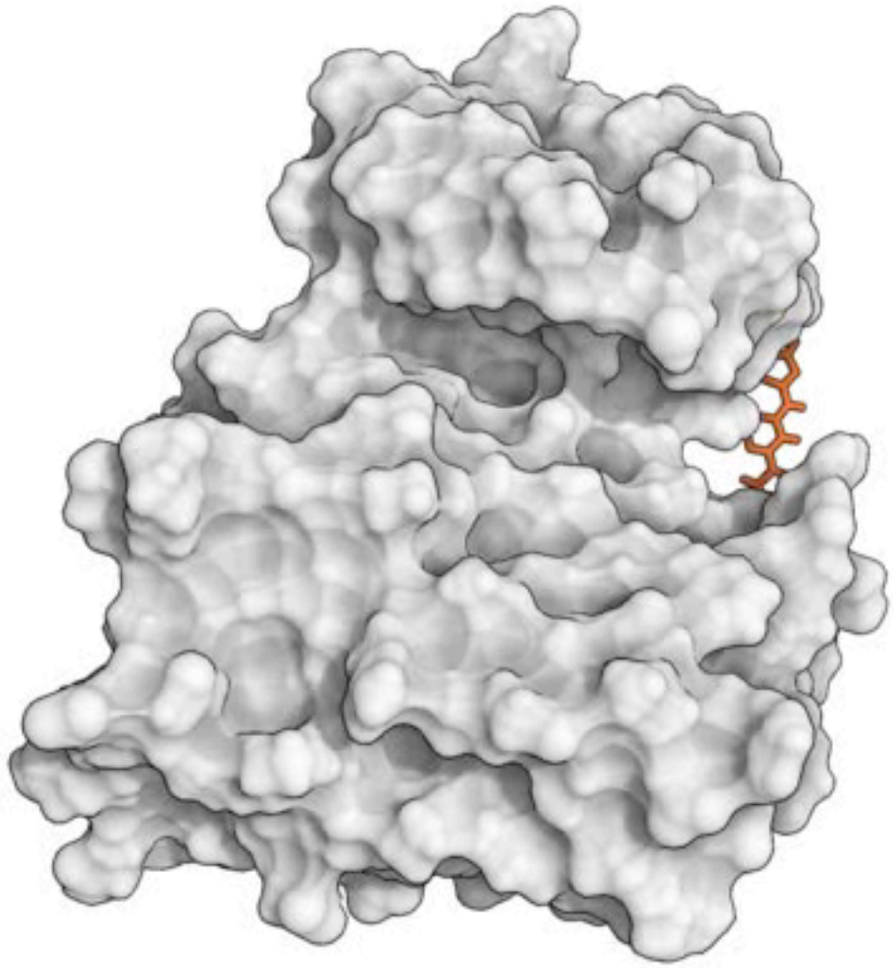
$$K_d \propto \frac{\tau_{\mathrm{unbound}}}{\tau_{\mathrm{bound}}}$$

ANTON
$50M special-purpose supercomputer
(D.E. Shaw Research)

David E. Shaw

ANTON @ DESRES

Src:dasatanib
(4 us simulation)



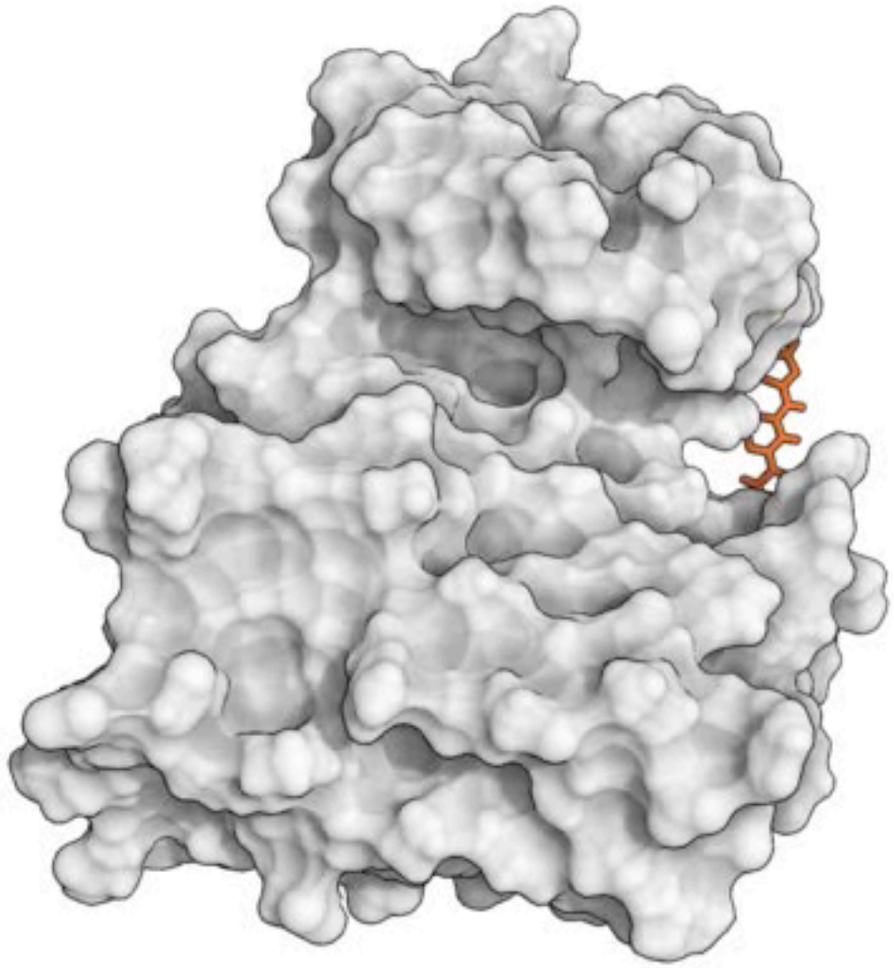
Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. JACS 133:9181, 2011.

ANTON
$50M special-purpose supercomputer
(D.E. Shaw Research)

David E. Shaw

ANTON @ DESRES

Src:dasatanib
(4 us simulation)



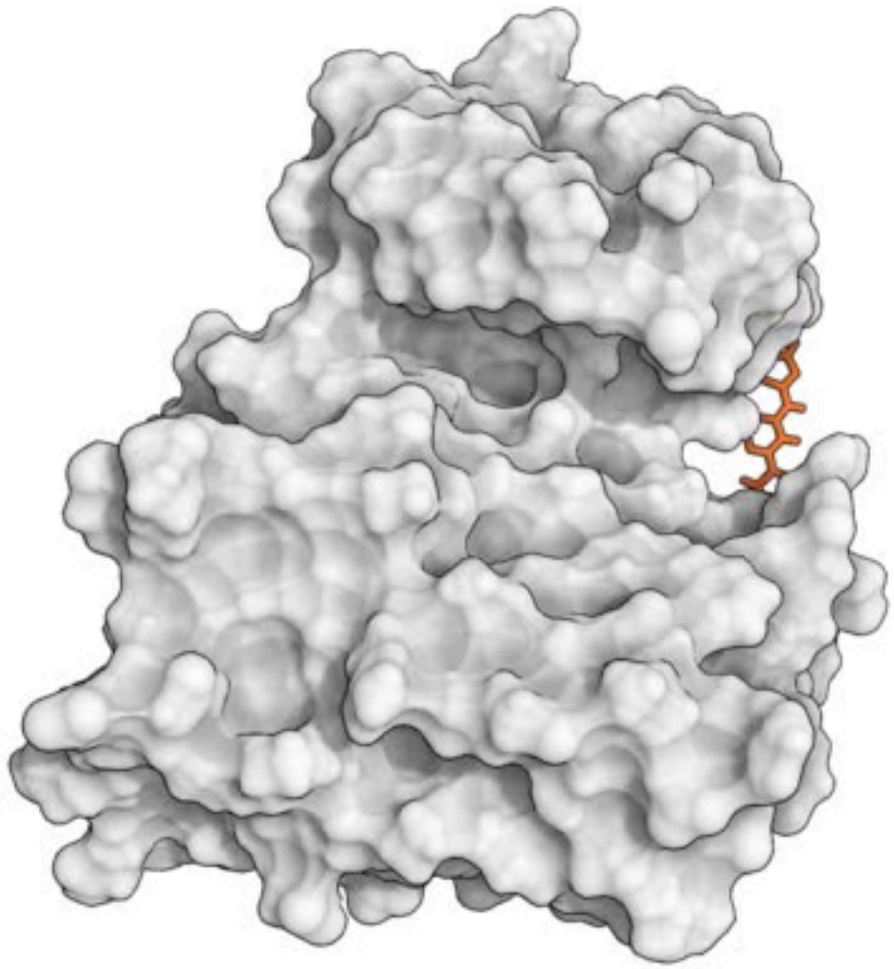
Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. JACS 133:9181, 2011.

ANTON
$50M special-purpose supercomputer
(D.E. Shaw Research)

David E. Shaw

ANTON @ DESRES

Src:dasatanib
(4 us simulation)

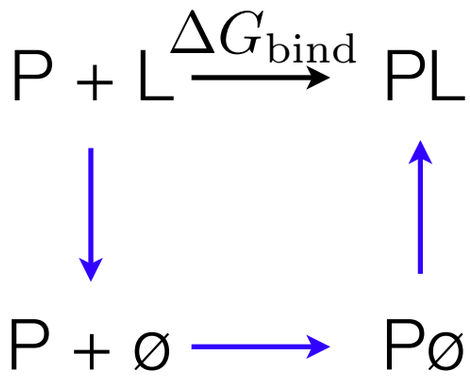


To reliably estimate affinities for molecules with typical off-rates (10^{-4} s^{-1}), trajectories would need to be impractically long (hours)

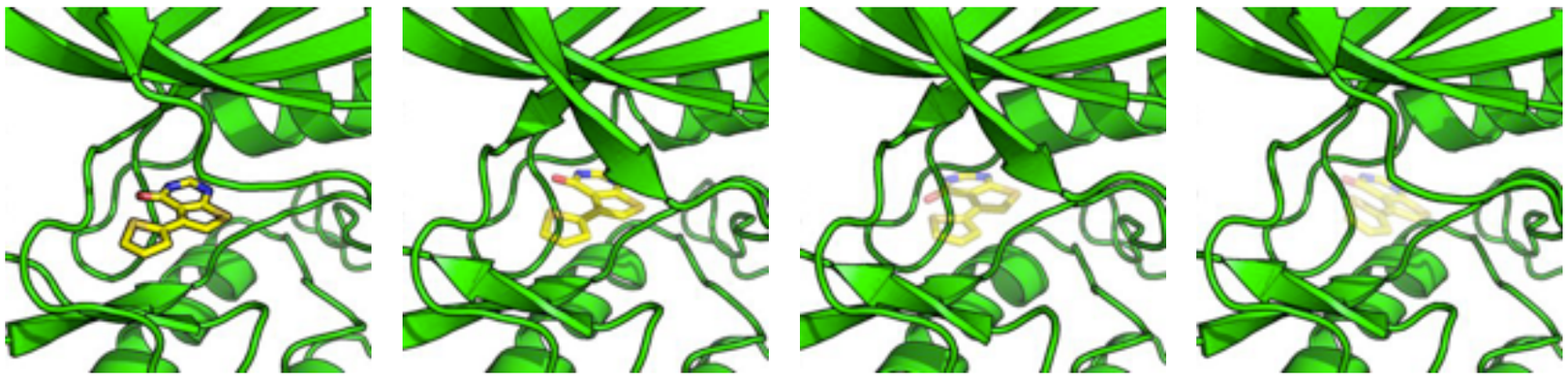
requiring ~10^6 years to simulate.

Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. JACS 133:9181, 2011.

Alchemical free energy calculations provide a rigorous way to efficiently compute binding affinities for a given forcefield



multiple simulations of "alchemical intermediates" are run



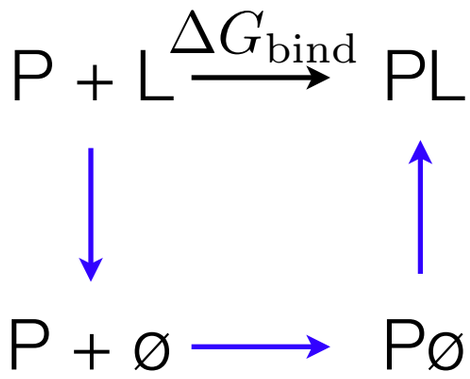
imposition of restraints **discharging** **steric decoupling** **noninteracting**

Requires orders of magnitude less effort than simulating direct association process, but still includes all enthalpic/entropic contributions to binding free energy.

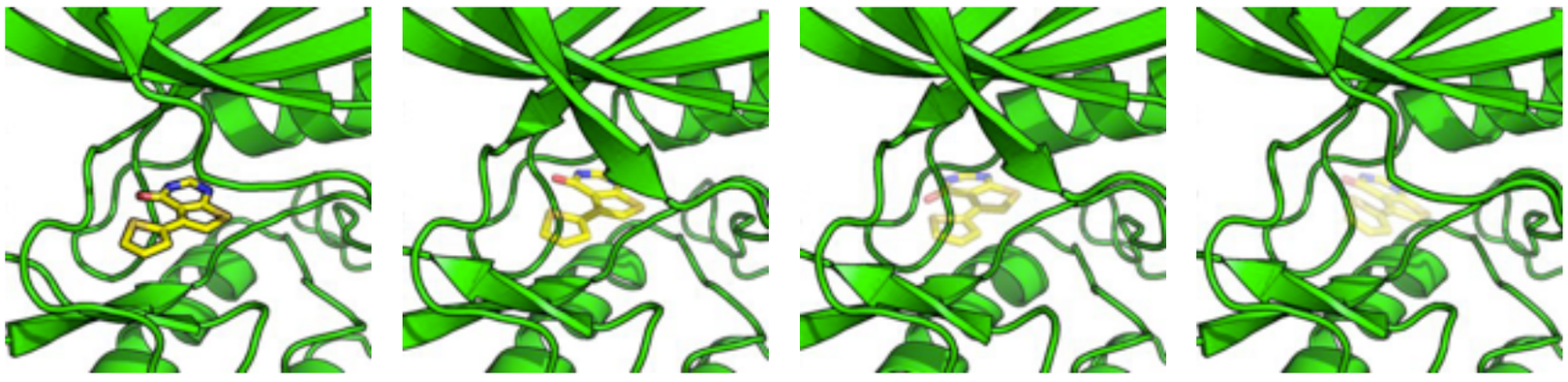
$$\Delta F_{1\to N} = -\beta^{-1} \ln \frac{Z_N}{Z_1} = -\beta^{-1} \ln \frac{Z_2}{Z_1} \cdot \frac{Z_3}{Z_2} \cdots \frac{Z_N}{Z_{N-1}} = \sum_{n=1}^{N-1} \Delta F_{n\to n+1} \qquad Z_n = \int d\mathbf{x}\, e^{-\beta U(\mathbf{x})}$$

Pioneering work from many: McCammon, van Gunsteren, Kollman, Chipot, Roux, Boresch, Pande, Shirts, Swope, Christ, Mobley, Roux, and many more
Recent review: Chodera, Mobley, Shirts, Dixon, Branson, Pande. Curr Opin Struct Biol 21:150, 2011.

Alchemical free energy calculations provide a rigorous way to efficiently compute binding affinities for a given forcefield



multiple simulations of "alchemical intermediates" are run



imposition of restraints **discharging** **steric decoupling** **noninteracting**

Requires orders of magnitude less effort than simulating direct association process, but still includes all enthalpic/entropic contributions to binding free energy.

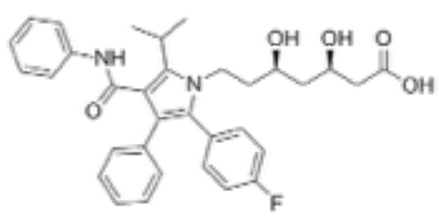
$$\Delta F_{1\to N} = -\beta^{-1} \ln \frac{Z_N}{Z_1} = -\beta^{-1} \ln \frac{Z_2}{Z_1} \cdot \frac{Z_3}{Z_2} \cdots \frac{Z_N}{Z_{N-1}} = \sum_{n=1}^{N-1} \Delta F_{n\to n+1} \qquad Z_n = \int d\mathbf{x}\, e^{-\beta U(\mathbf{x})}$$

Pioneering work from many: McCammon, van Gunsteren, Kollman, Chipot, Roux, Boresch, Pande, Shirts, Swope, Christ, Mobley, Roux, and many more
Recent review: Chodera, Mobley, Shirts, Dixon, Branson, Pande. Curr Opin Struct Biol 21:150, 2011.

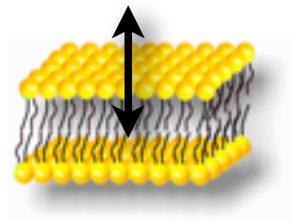
Alchemical free energy calculations can in principle also compute other relevant physical properties

---

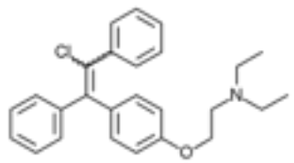
partition coefficients (logP, logD) and permeabilities



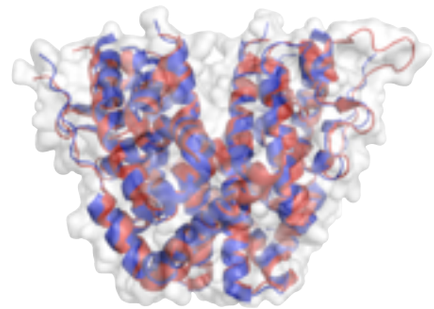
lipitor



selectivity for subtypes or related targets/off-targets



clomifene



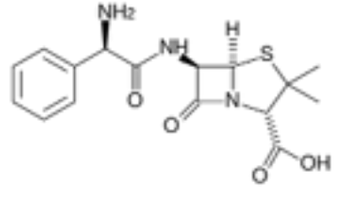
ERα/β

lead optimization of affinity and selectivity

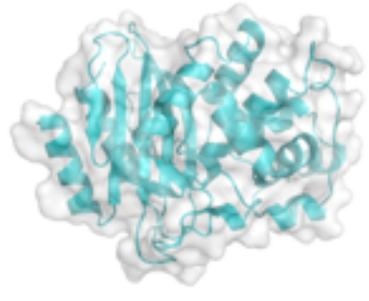
Imatinib

Dasatinib

susceptibility to resistance mutations



ampicillin



β-lactamase

also: solubilities, polymorphs, etc.

with David L. Mobley

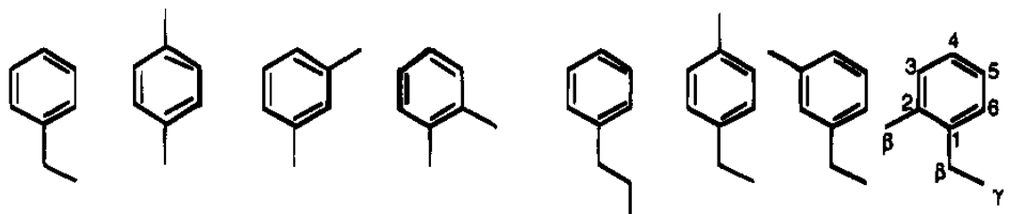
How accurate are alchemical binding free calculations?

Do they meet this 2 kcal/mol threshold for useful accuracy?

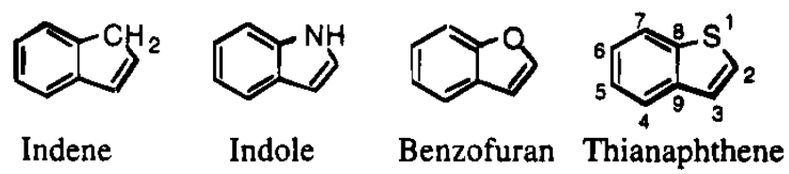
with David L. Mobley

How accurate are binding free energy calculations?
T4 lysozyme L99A as a simple model binding site

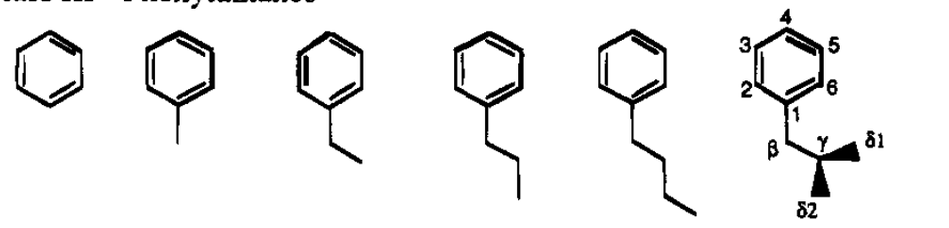
Class I - "Isophobic" Ligands

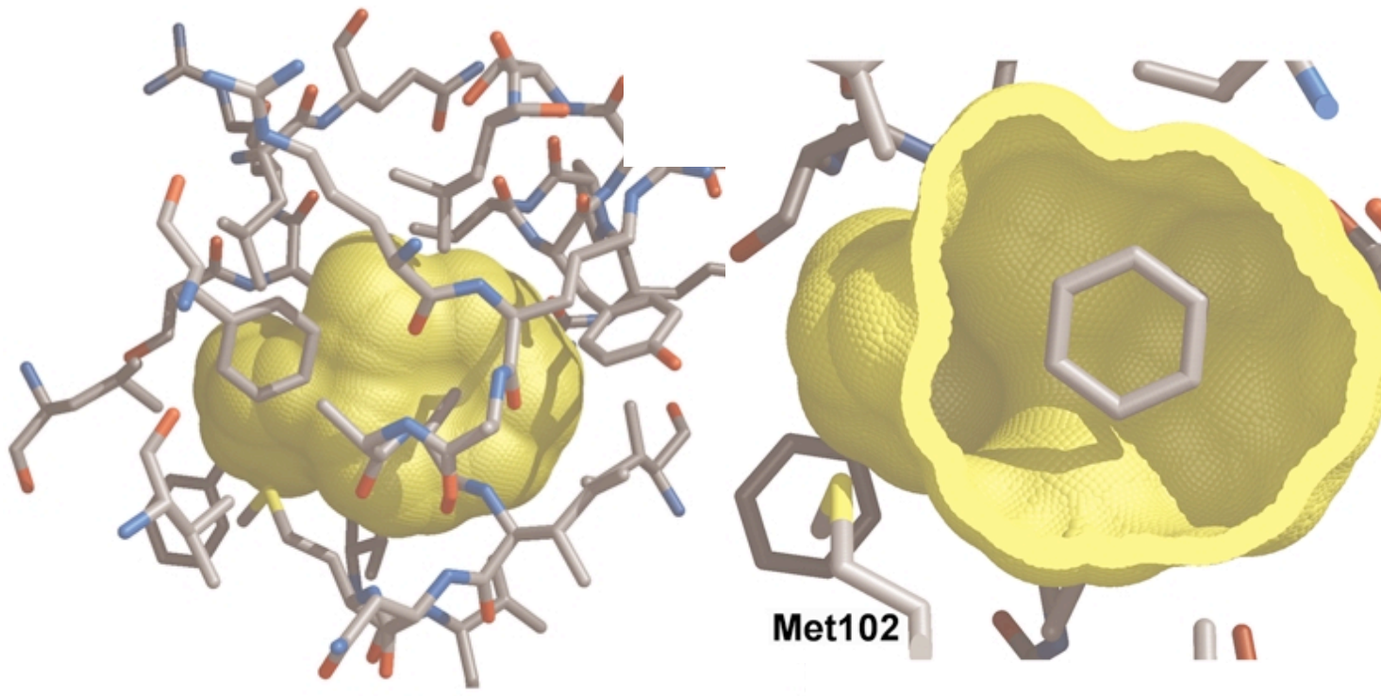


Class II - "Isosteric" Ligands



Class III - Phenylalkanes





Wei, Baase, Weaver, Matthews, and Shoichet. JMB 322:339, 2002.

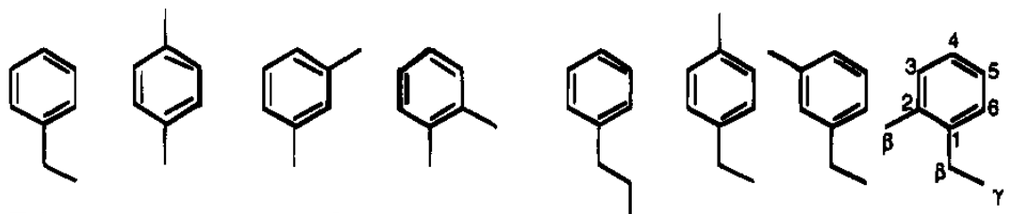
Surprisingly challenging for docking codes to discriminate binders (uM) from non-binders (>>uM)
Polar version of this cavity (L99A/M102Q) even more challenging for docking.

Let's test how well free energy calculations can reproduce measured binding free energies.

with David L. Mobley

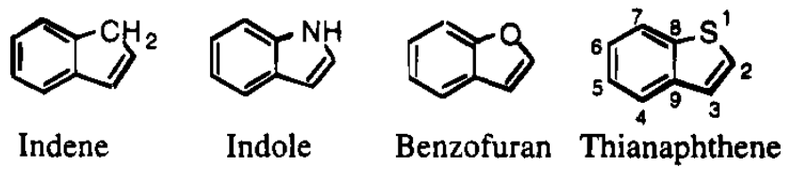
How accurate are binding free energy calculations?
T4 lysozyme L99A as a simple model binding site

Class I - "Isophobic" Ligands



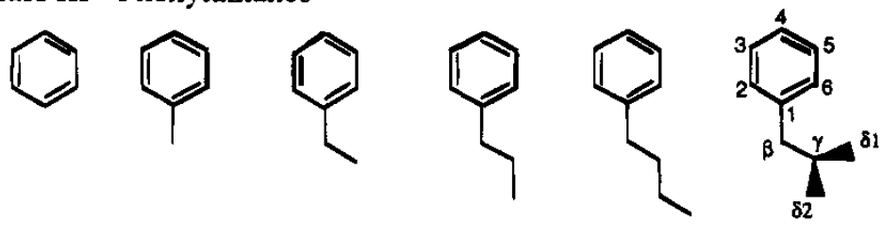
Ethyl benzene | p-, m-, o-Xylene | Propyl benzene | p-, m-, o-Ethyltoluene

Class II - "Isosteric" Ligands

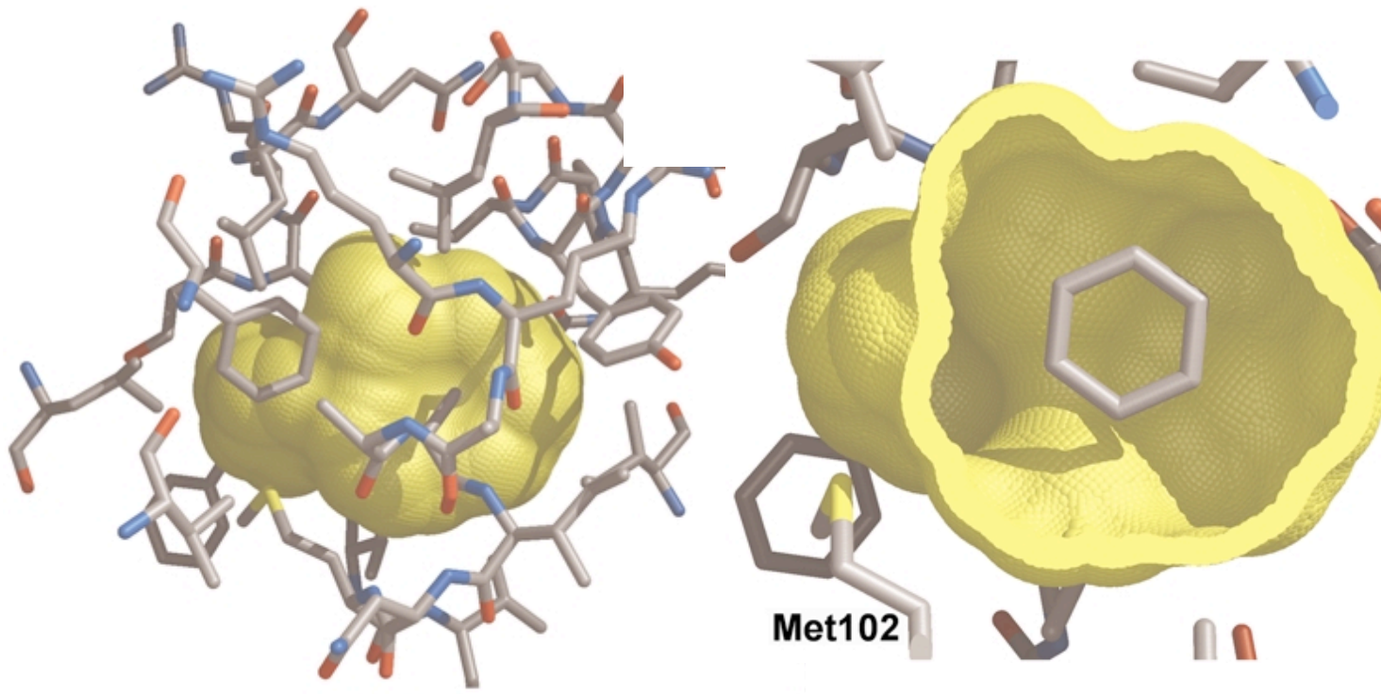


Indene | Indole | Benzofuran | Thianaphthene

Class III - Phenylalkanes



Benzene | Toluene | Ethyl- | Propyl- | Butyl- | iso-Butylbenzene



Wei, Baase, Weaver, Matthews, and Shoichet. JMB 322:339, 2002.

Surprisingly challenging for docking codes to discriminate binders (uM) from non-binders (>>uM)
Polar version of this cavity (L99A/M102Q) even more challenging for docking.

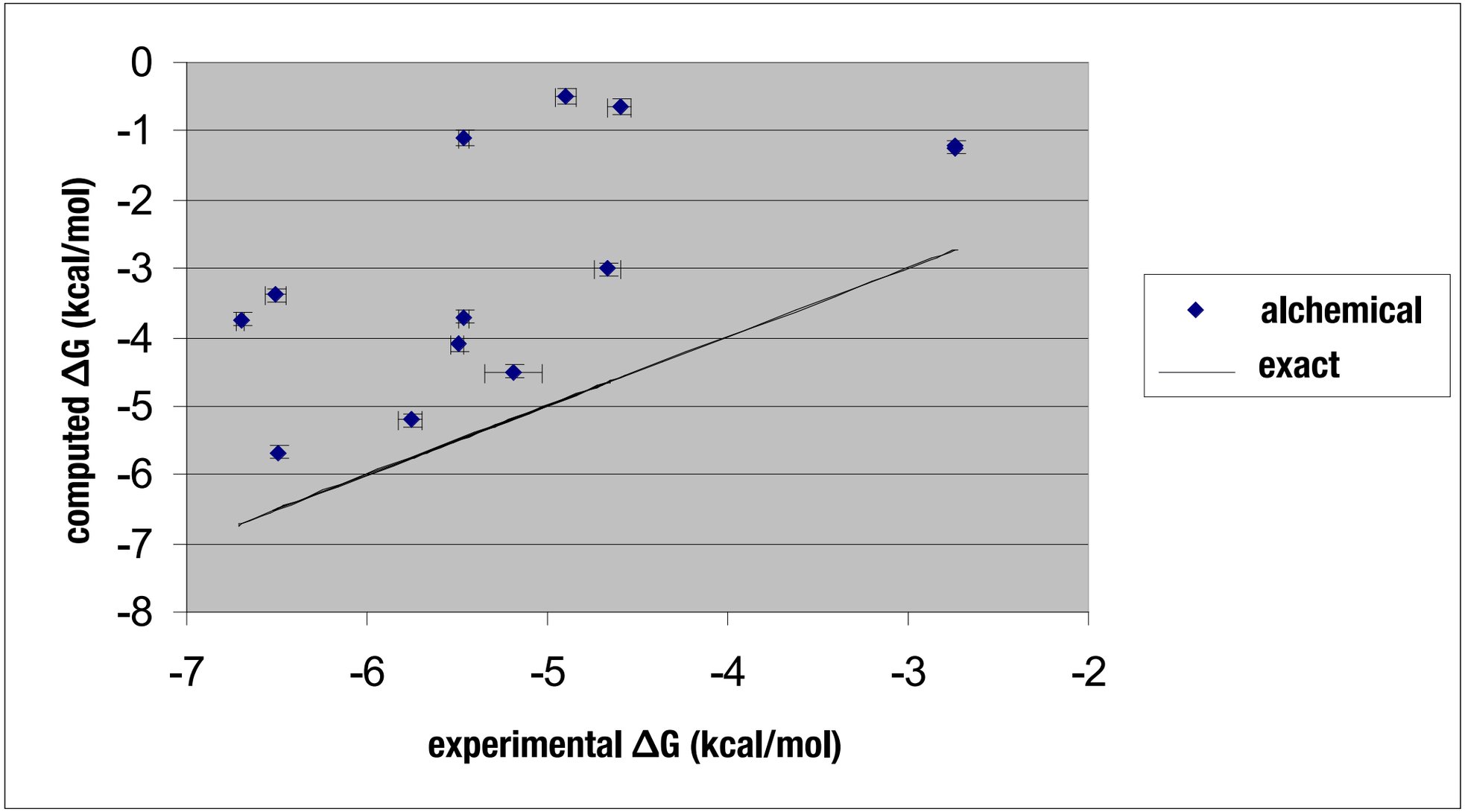
Let's test how well free energy calculations can reproduce measured binding free energies.

"Surely, there's no way we can screw this up."

- Famous last words

“Nothing is foolproof to a sufficiently talented fool”

- Old but pertinent philosophy

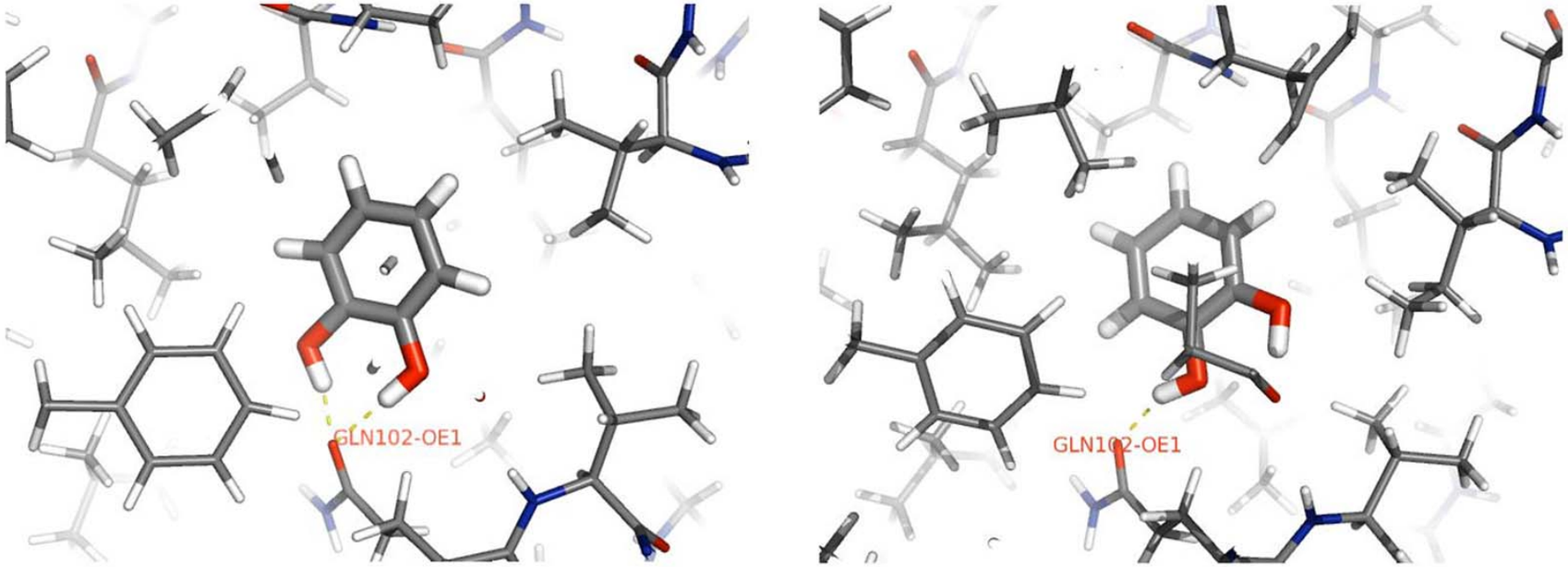


Obviously, we’re missing something.
What’s going on here?

Mobley, Graves, Chodera, McReynolds, Shoichet, Dill. J Mol Biol 371:1118, 2007

Multiple ligand conformations can contribute

Switching between relevant ligand orientations can occur on a timescale of many nanoseconds!



L99A/M102Q polar version of cavity

Difference in affinity between different bound orientations is only ~1 kT

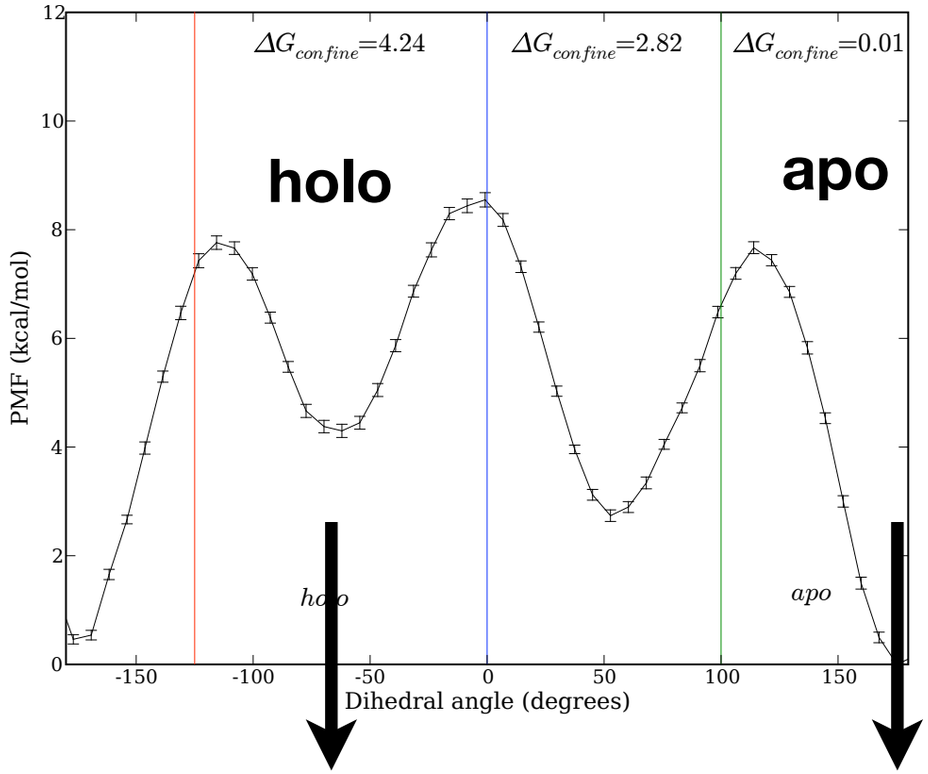
N poses can contribute to the overall binding free energy ~ kT ln N

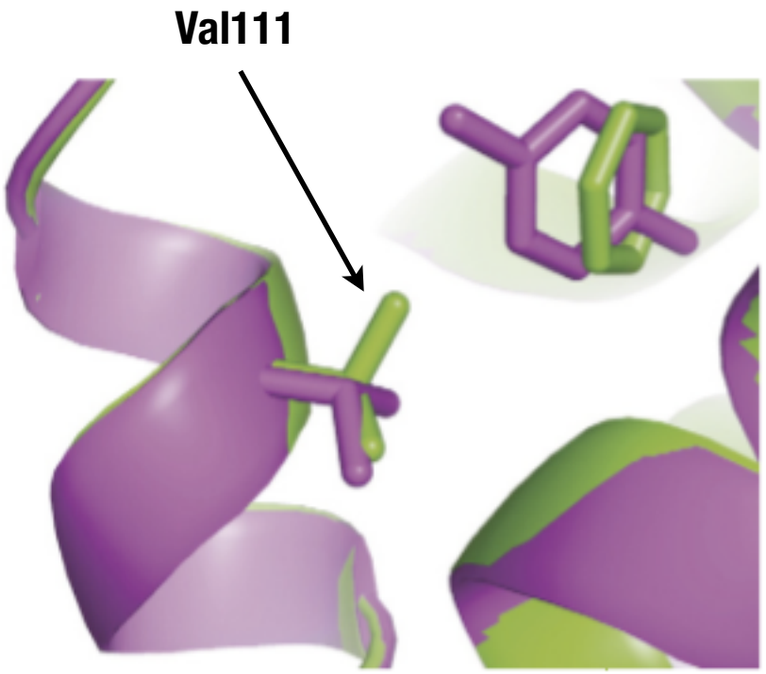
[Also relevant for ligands with pseudosymmetric substituents]

Mobley, Chodera, Dill. JCP 125:084902, 2006.

Multiple protein conformations can contribute

Val111 sidechain χ_1 in apo structure





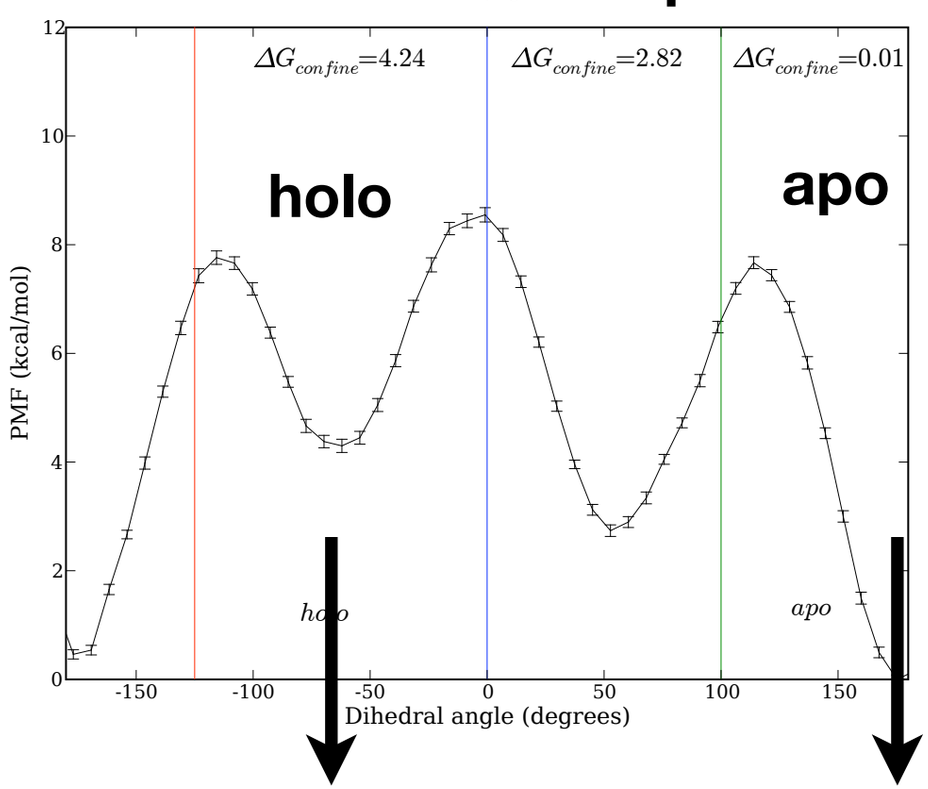
p-xylene vs. benzene bound to T4 lysozyme L99A

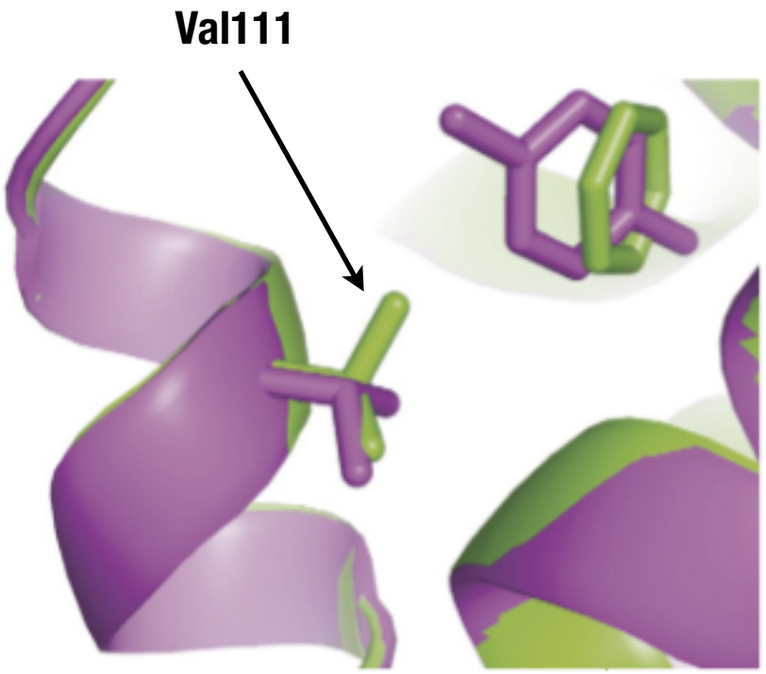
binding free energy	-7.3 kcal/mol	-3.0 kcal/mol	**4.3 kcal/mol difference!**

ΔG_{exp} = -4.7 kcal/mol

Mobley, Chodera, and Dill. JCTC 3:1231, 2007.

Multiple protein conformations can contribute





p-xylene vs. benzene bound to T4 lysozyme L99A

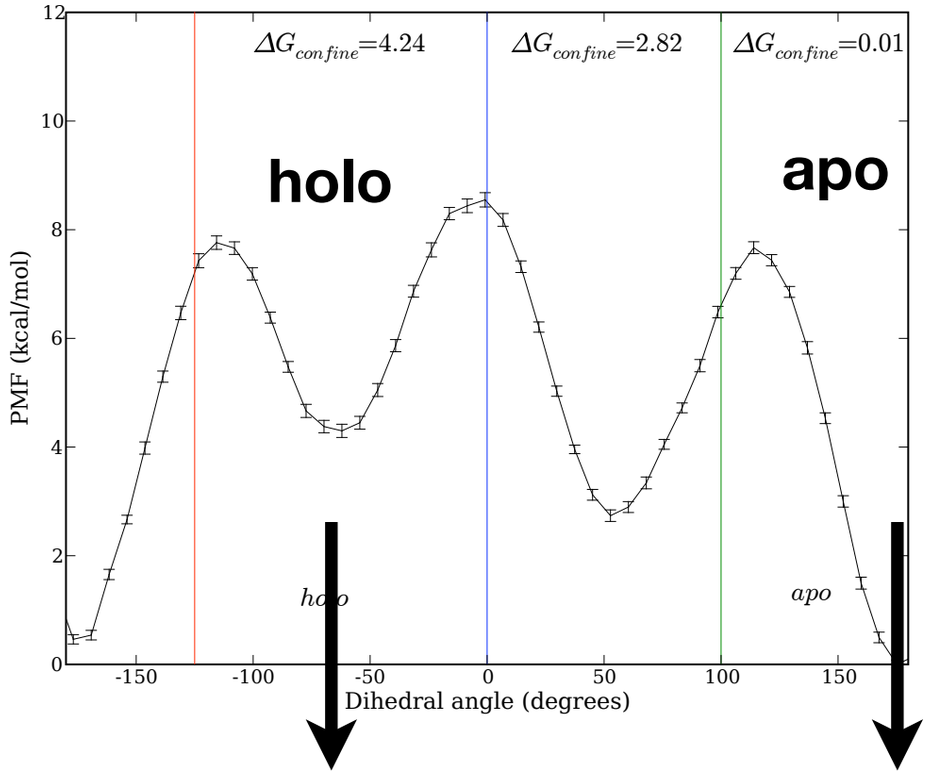
binding free energy	-7.3 kcal/mol	-3.0 kcal/mol	**4.3 kcal/mol difference!**
confinement free energy	4.2 kcal/mol	0.0 kcal/mol	

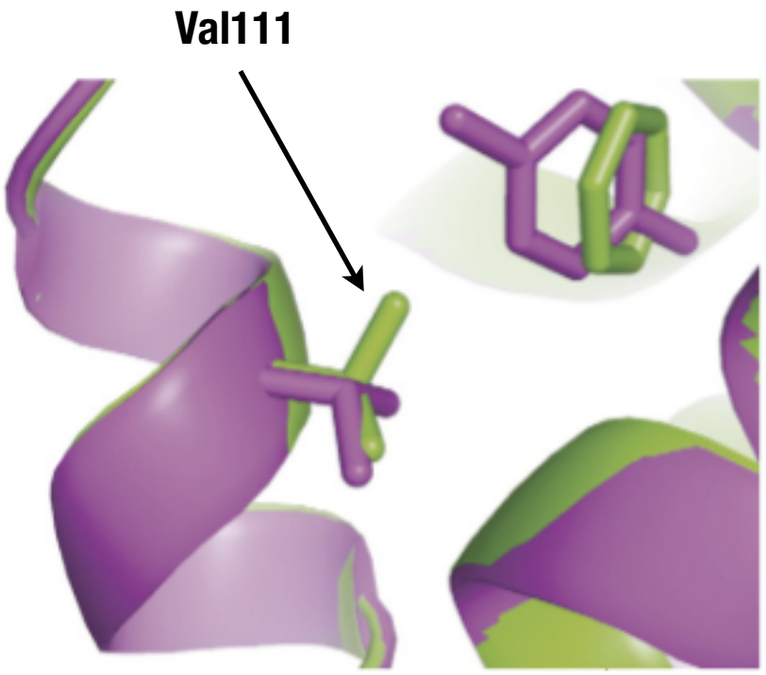
ΔG_{exp} = -4.7 kcal/mol

Mobley, Chodera, and Dill. JCTC 3:1231, 2007.

Multiple protein conformations can contribute

Val111 sidechain χ_1 in apo structure





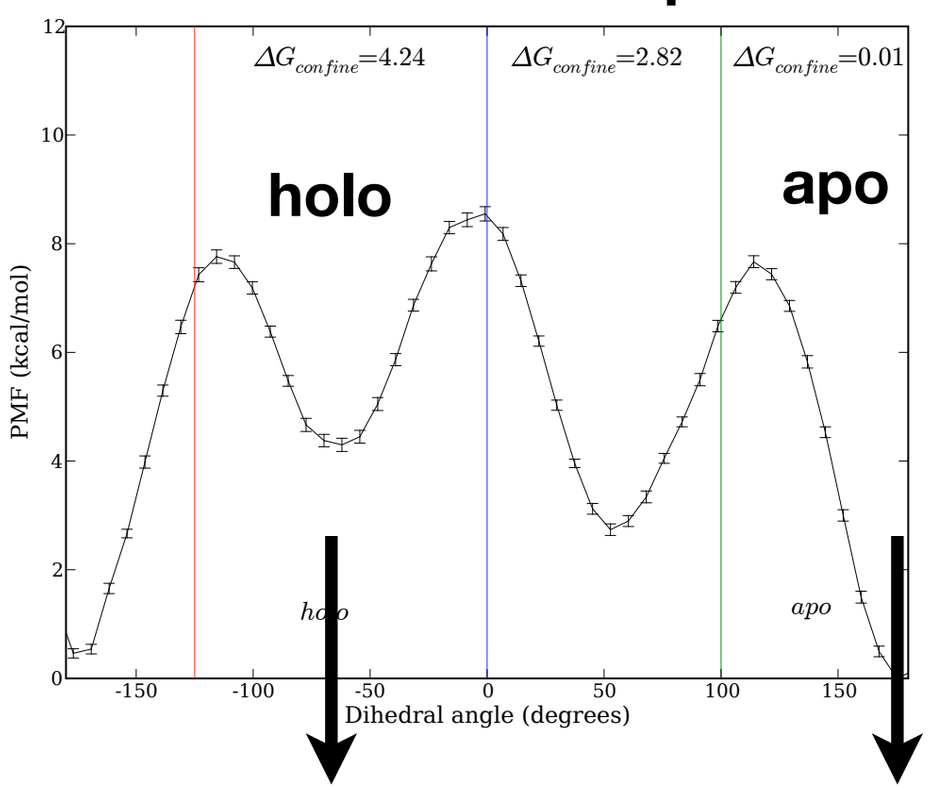
p-xylene vs. benzene bound to T4 lysozyme L99A

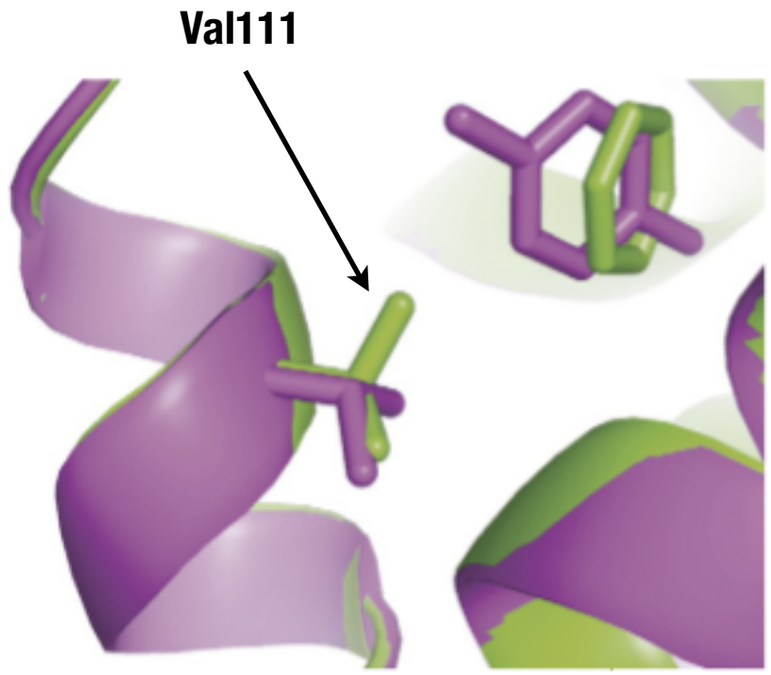
binding free energy	-7.3 kcal/mol	-3.0 kcal/mol	**4.3 kcal/mol difference!**
confinement free energy	4.2 kcal/mol	0.0 kcal/mol	
release free energy following binding	-0.3 kcal/mol	-0.6 kcal/mol	

ΔG_{exp} = -4.7 kcal/mol

Mobley, Chodera, and Dill. JCTC 3:1231, 2007.

Multiple protein conformations can contribute





p-xylene vs. benzene bound to T4 lysozyme L99A

binding free energy	-7.3 kcal/mol	-3.0 kcal/mol	**4.3 kcal/mol difference!**
confinement free energy	4.2 kcal/mol	0.0 kcal/mol	
release free energy following binding	-0.3 kcal/mol	-0.6 kcal/mol	
total binding free energy	-3.4+-0.3	≈ -3.6+-0.3	**ΔG_{exp} = -4.7 kcal/mol**

Mobley, Chodera, and Dill. JCTC 3:1231, 2007.

Numerous improvements were required for T4 lysozyme L99A

- ✔ multiple long-lived ligand orientations — Mobley, Chodera, Dill. JCP 125:084902, 2006.
- ✔ multiple long-lived protein conformations — Mobley, Chodera, and Dill. JCTC 3:1231, 2007.
- ✔ anisotropic dispersion correction — Shirts, Mobley, Chodera, Pande. JPC B 111:13052, 2007.
- ✔ optimal use of all data in analysis — Shirts and Chodera. JCP 129:124105, 2008.
- ✔ binding site restraints to reduce simulation times — Mobley, Chodera, Dill. JCP 125:084902, 2006.
- ✔ improved ligand charge models — Mobley, Dumont, Chodera, Dill. 111:2242, 2007.

These issues are very general, and algorithmic improvements that address them should universally improve accuracy for protein-ligand systems.

Resulting RMS error: 1.89±0.04 kcal/mol **[originally 3.51±0.04 kcal/mol]**

But it's easy to fool ourselves when working with a known dataset.
How well do we do on data we've never seen?

Mobley, Graves, Chodera, McReynolds, Shoichet, Dill. J Mol Biol 371:1118, 2007

How accurately can we predict unknown binding affinities?

Brian Shoichet, UCSF

Blinded test of prediction power with new molecules:

Ligand	Prediction[1]	ΔG^{o}_{calc}[2] (kcal/mol)
1,2-dichlorobenzene	Binder	-5.66 ± 0.15
n-methylaniline	Binder	-5.37 ± 0.11
1-methylpyrrole	Binder	-4.32 ± 0.08
1,2-benzenedithiol	Binder	-2.79 ± 0.13
thieno-[2,3-c]pyridine	Nonbinder	-2.56 ± 0.07

Mobley, Graves, Chodera, McReynolds, Shoichet, Dill. J Mol Biol 371:1118, 2007

How accurately can we predict unknown binding affinities?

Brian Shoichet, UCSF

Blinded test of prediction power with new molecules:

Ligand	Prediction[1]	ΔG^o_{calc} [2] (kcal/mol)	Experiment	$\Delta G^o_{expt.}$ (kcal/mol)
1,2-dichlorobenzene	Binder	-5.66 ± 0.15	Binder	-6.37
n-methylaniline	Binder	-5.37 ± 0.11	Binder	-4.70
1-methylpyrrole	Binder	-4.32 ± 0.08	Binder	-4.44
1,2-benzenedithiol	Binder	-2.79 ± 0.13	Binder	N.D.
thieno-[2,3-c]pyridine	Nonbinder	-2.56 ± 0.07	Nonbinder	N.D.

All binding predictions confirmed!

RMS error: 0.6 kcal/mol

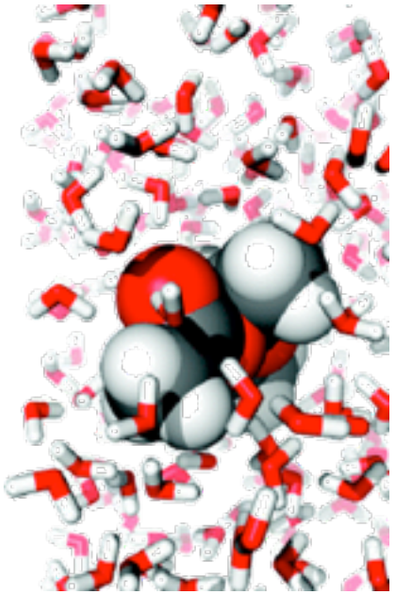
(but we could only convince them to do N=3 ITC measurements)

Mobley, Graves, Chodera, McReynolds, Shoichet, Dill. J Mol Biol 371:1118, 2007

Alchemical free energy methods can work reliably in simple systems but complex systems remain challenging

---

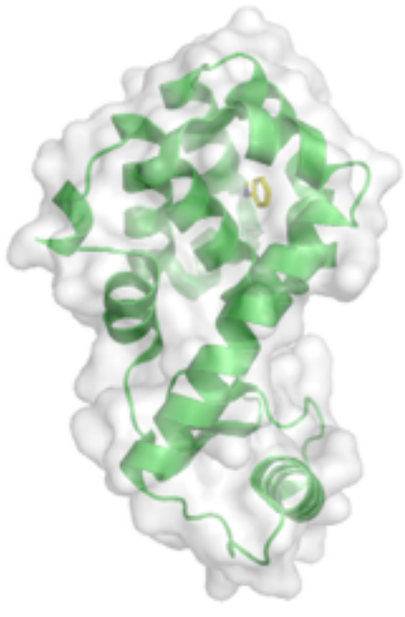
model systems ⟷ pharmaceutically relevant



hydration free energies of small neutral molecules

1.04±0.03 kcal/mol (N=44)
Mobley, Dumont, **Chodera**, Dill. JPC B, 2007
1.23±0.01 kcal/mol (N=502)
Mobley, Bayly, Cooper, Dill. JPC B 2009.

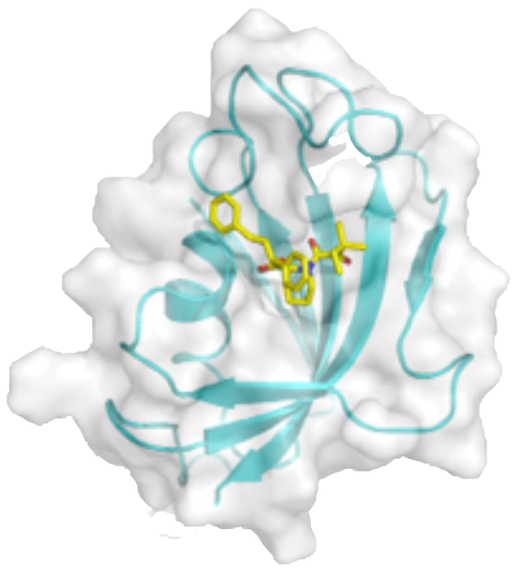
1.33±0.05 kcal/mol (N=17)
Nicholls, Mobley, Guthrie, **Chodera**, Pande. J Med Chem 2008.



small apolar ligands T4 lysozyme L99A

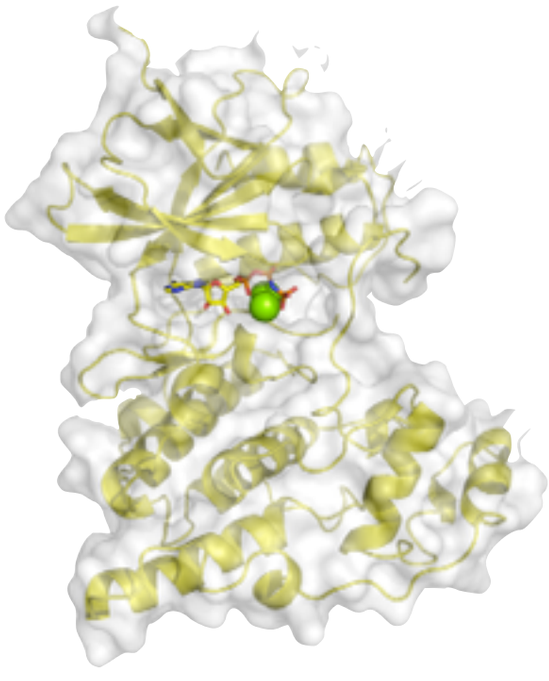
1.89±0.04 kcal/mol (N=13)
Mobley, Graves, **Chodera**, McReynolds, Shoichet Dill. JMB 2007

0.6±0.2 kcal/mol (N=3)
Mobley, Graves, **Chodera**, McReynolds, Shoichet Dill. JMB 2007



polar ligands FKBP12

1.42 kcal/mol (N=9)
0.94 kcal/mol (N=7)
(Shirts et al., in preparation)



JNK3 kinase

Anecdotal literature reports of success (publication bias?)

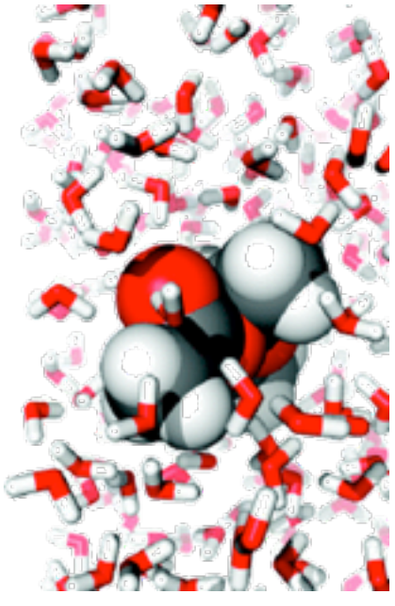
Calculations are notoriously unreliable. (e.g. SAMPL challenges)

retrospective RMS error [sample size]
prospective RMS error [sample size]

(not to scale)

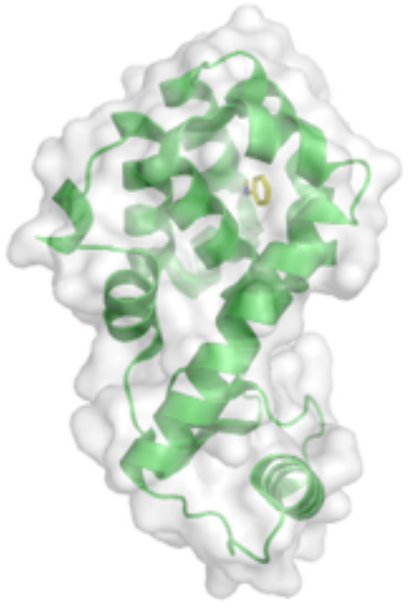
Alchemical free energy methods work reliably in simple systems but complex systems have remained challenging

model systems ⟷ pharmaceutically relevant



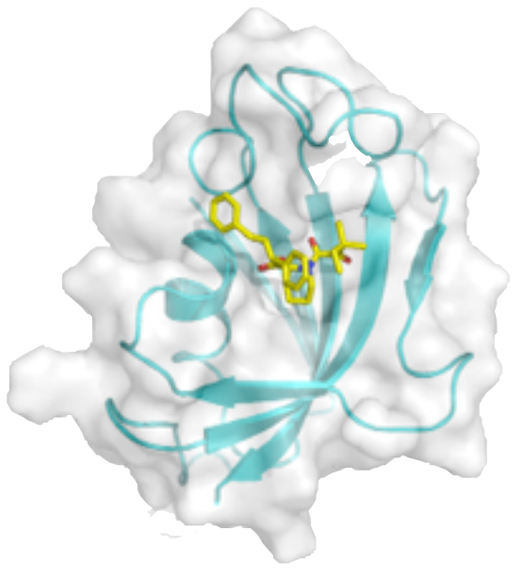
hydration free energies
of small neutral molecules

solvent only
small, neutral molecules
fixed protonation states



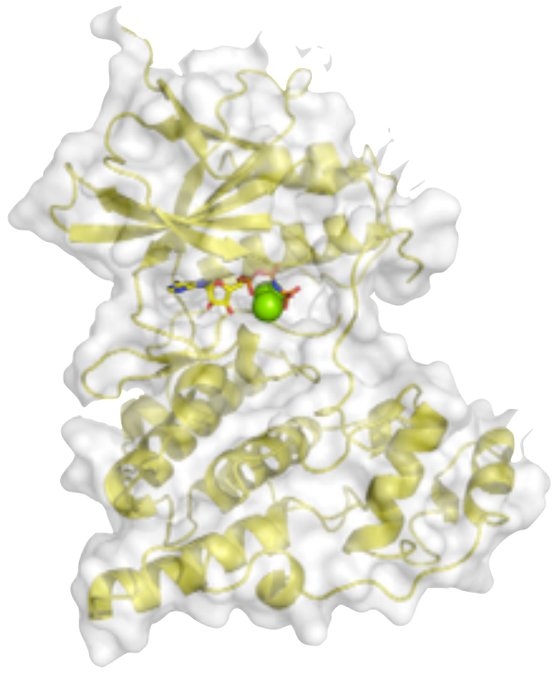
small apolar ligands
T4 lysozyme L99A

small, rigid protein
small, neutral ligands
fixed protonation states
multiple sidechain orientations
multiple ligand binding modes



polar ligands
FKBP12

small, rigid protein
fixed protonation states
larger drug-like ligands
with few rotatable bonds

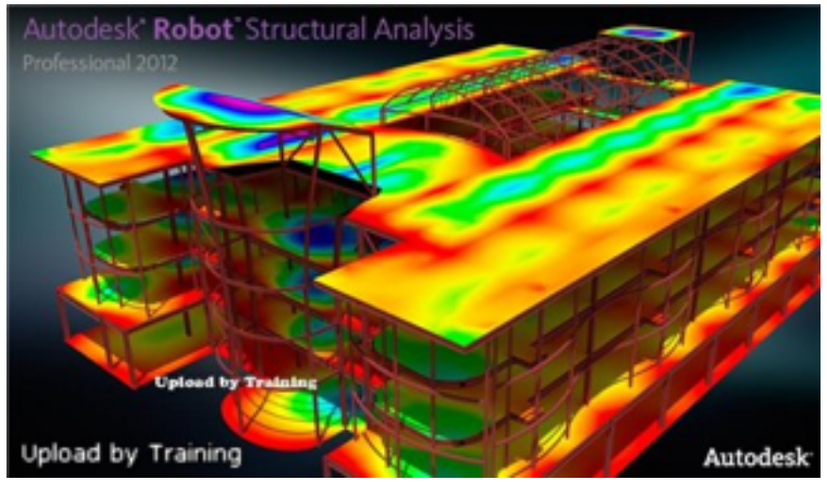


JNK3 kinase

large protein, multiple conformations
large drug-like ligands, rotatable bonds
multiple protonation states? tautomers?
phosphorylation and activation
peptide substrate?
$MgCl_2$ salt effects?

easy
hard

(not to scale)


Autodesk® Robot™ Structural Analysis
Professional 2012
Upload by Training
Upload by Training
Autodesk

Structural engineering wasn’t always so successful, either

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

Computational predictions fail for one of three reasons

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

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

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

We need to figure out why failures occur and how we can improve our algorithms to be more robust for prediction.

Speeding up the cycle of learning from failure can accelerate progress toward rational ligand design

---

computational predictions

experimental confirmation

“Fail fast, fail cheap”

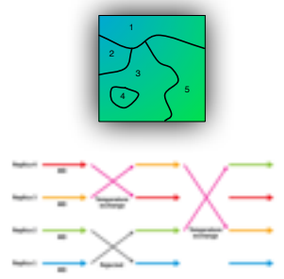
Make rigorous calculations of affinity fast and accurate

GPU acceleration

$$\pi(x)K(x,y) = \pi(y)K(y,x)$$

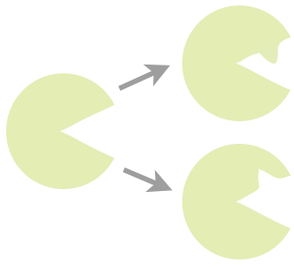
$$AH \leftrightharpoons A^- + H^+$$

modular MCMC for sampling and chemical effects



enhanced sampling algorithms

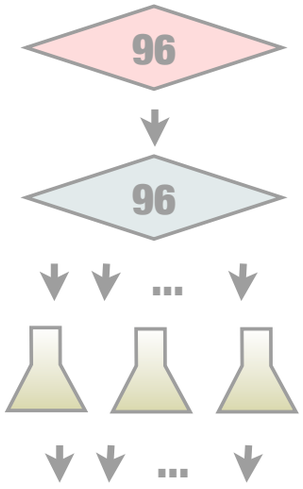
Test and improve models quickly and cheaply by inverting the drug discovery problem



mutate proteins instead of ligands

¢

buy inexpensive ligands



high-throughput experiments

How can we speed up the calculations?

$50K
5 TFLOP
doubling every 18 mo

many CPU-weeks/calculation

Doesn't fit neatly in a synthetic chemist's timeframe to wait weeks for an answer.

How can we speed up the calculations?

$50K
5 TFLOP
doubling every 18 mo

many CPU-weeks/calculation

How can we speed up the calculations?

$50K
5 TFLOP
doubling every 18 mo

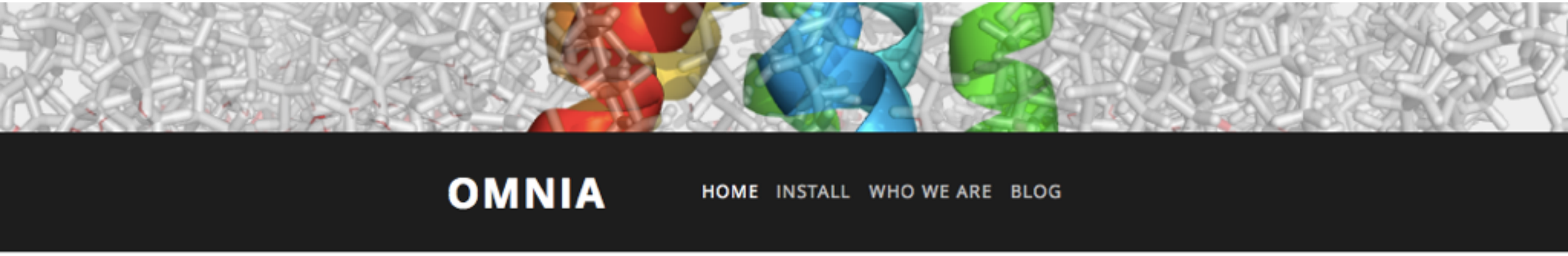
$500
5 TFLOP
doubling every 12 mo

many CPU-weeks/calculation

overnight on a workstation?

Can we exploit new GPU technologies to reach practical computation times?

YANK: An open-source, community-oriented platform for GPU-accelerated free energy calculations



http://www.omnia.md

Simulate. Analyze. Explore. High performance, high usability toolkits for predictive biomolecular simulation.

OpenMM speedup (GTX Titan) over 12-core Xeon X5650 CPU for DHFR

method	natoms	gromacs CPU	OpenMM GPU	speedup
GB/SA	2,489	2.54 ns/day	287 ns/day	**113 x**
RF	23,558	18.8 ns/day	163 ns/day	**8.7 x**
PME	23,558	6.96 ns/day	104 ns/day	**15 x**

http://simtk.org/home/openmm
gromacs benchmarks from http://biowulf.nih.gov/apps/gromacs-gpu.html

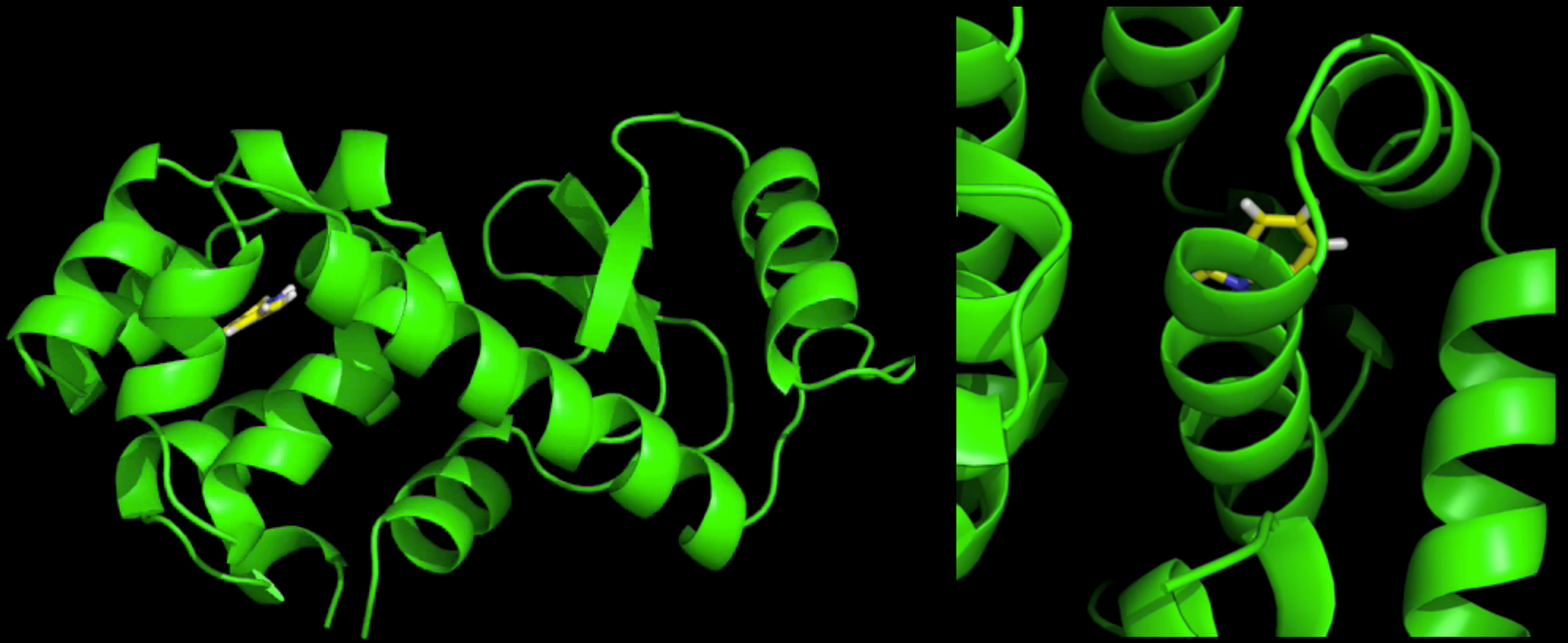
NVIDIA GTX-Titan ($1000)

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

Replica-exchange algorithms facilitate sampling of multiple binding modes

---

solid fully interacting
transparent noninteracting



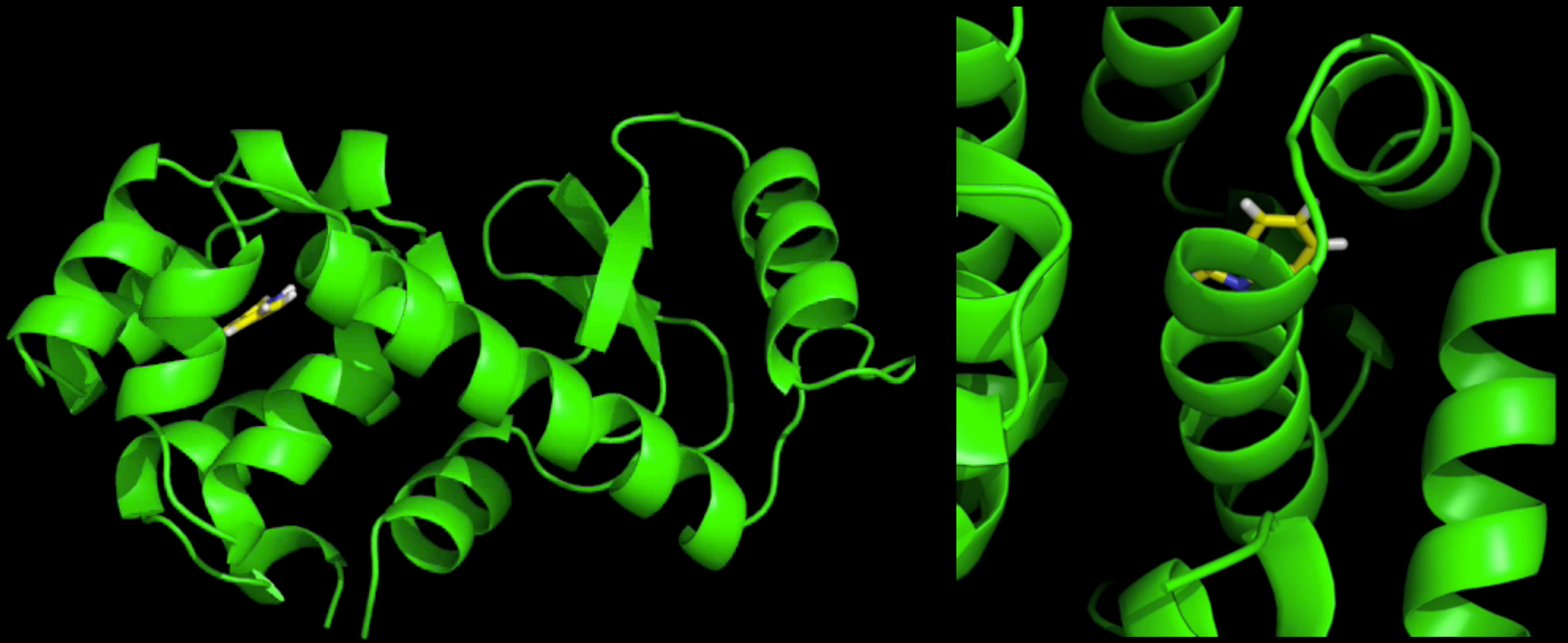
indole binding to T4 lysozyme L99A
12 h on 2 NVIDIA Tesla M2090 GPUs
Hamiltonian exchange with Gibbs sampling

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

Replica-exchange algorithms facilitate sampling of multiple binding modes

---

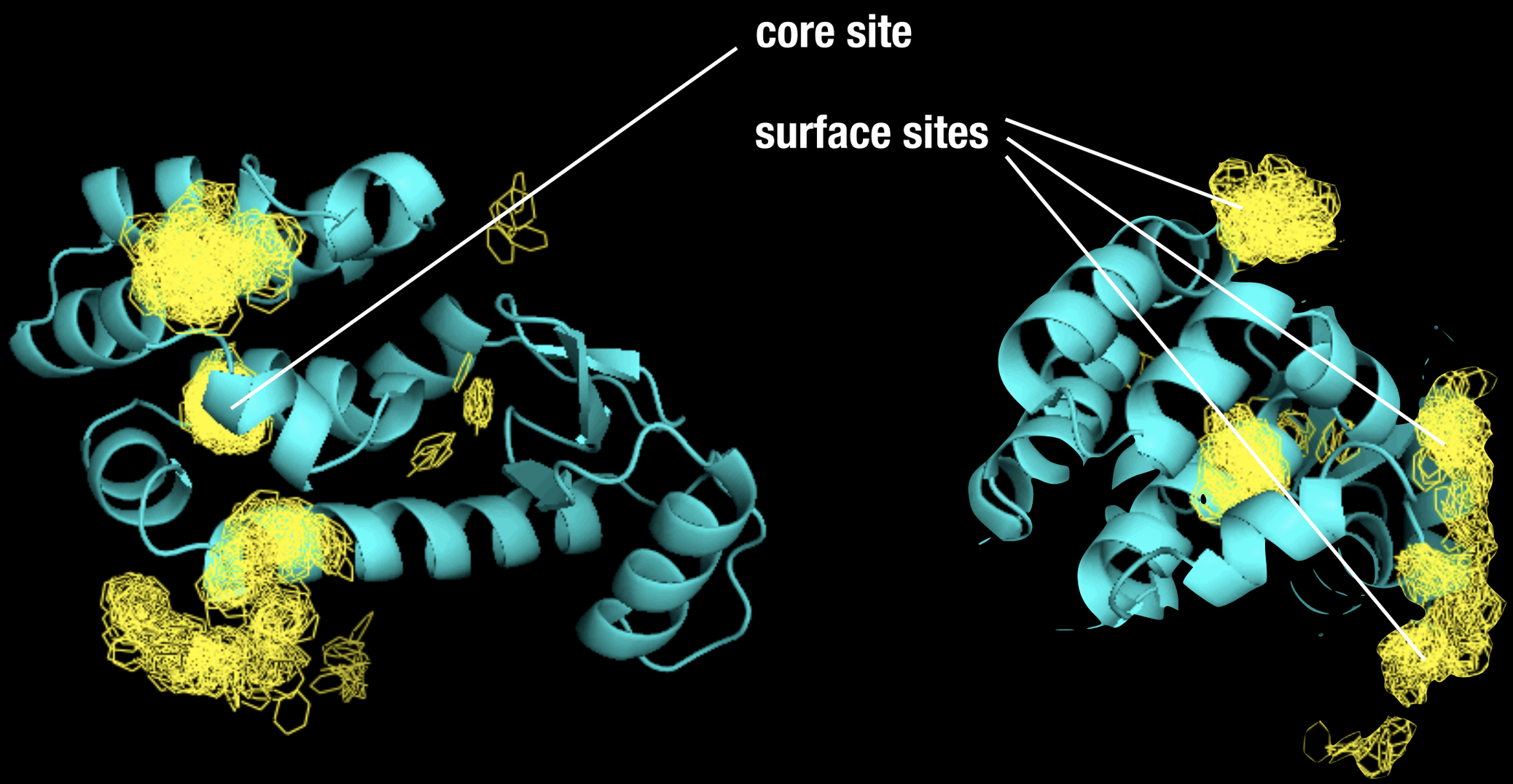
solid fully interacting
transparent noninteracting



indole binding to T4 lysozyme L99A
12 h on 2 NVIDIA Tesla M2090 GPUs
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

Additional and unknown binding sites can be identified, and their individual affinities estimated by the addition of MC moves

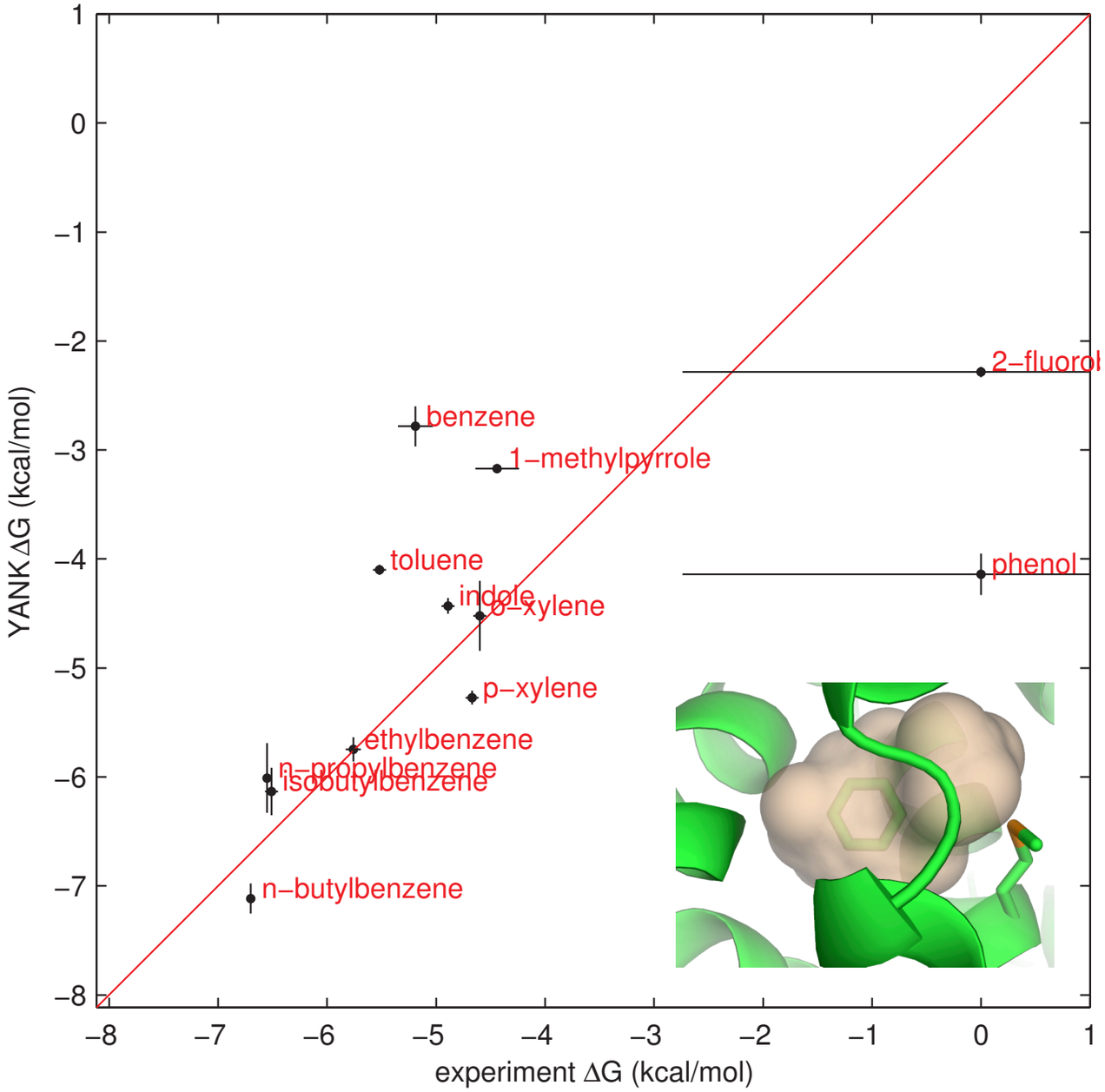


benzene bound to T4 lysozyme L99A
AMBER96 + OBC GBSA

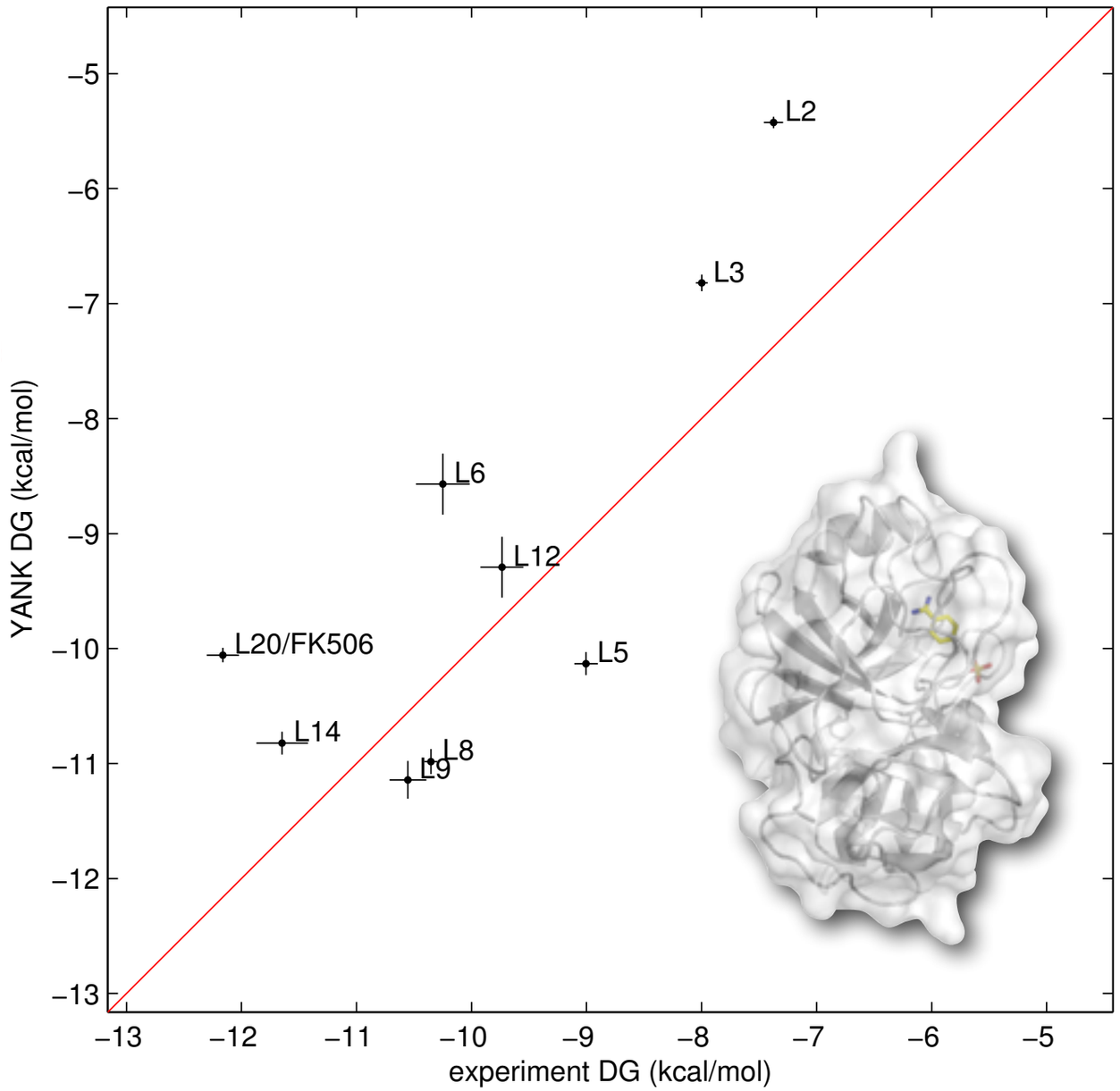
Chodera and Shirts. JCP 135:194110, 2011
Wang, Chodera, Yang, and Shirts. JCAMD 27:989, 2013.
http://github.org/choderalab/yank

Initial results using implicit models of solvent are promising: Could have a role in rapid affinity prediction

T4 lysozyme L99A



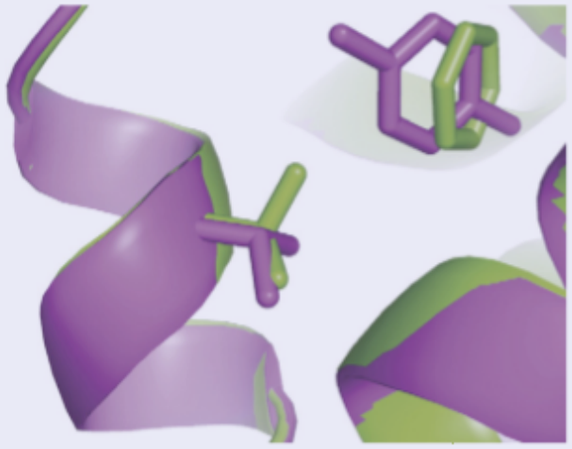
FKBP12



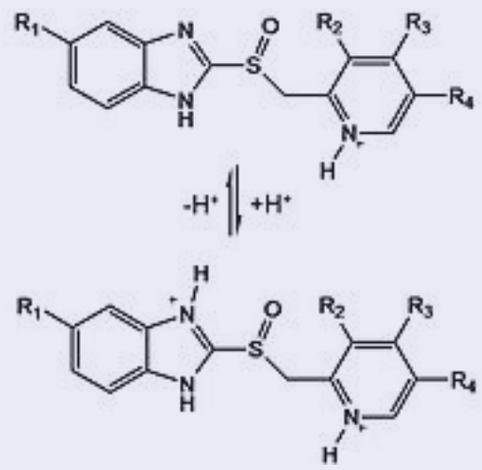
AMBER ff96 + OBC GBSA (no cutoff) + GAFF/AM1-BCC
12 h on 2 GPUs

http://github.org/choderalab/yank

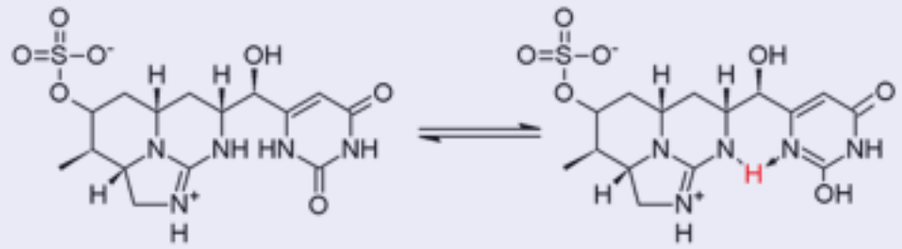
New Monte Carlo techniques open up new possibilities for treating physical effects and facilitating design



sidechain rotamer sampling



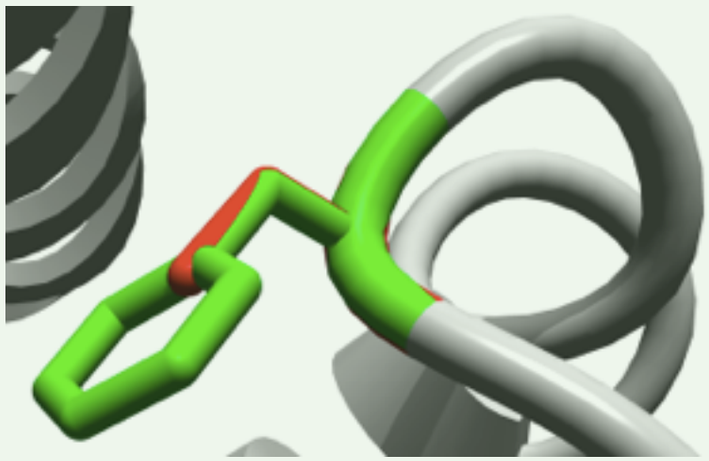
dynamic protonation states (protein and ligand)



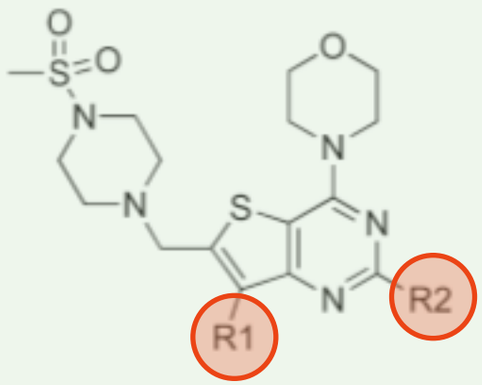
dynamic tautomerization (ligand)

ENHANCING CHEMICAL ACCURACY

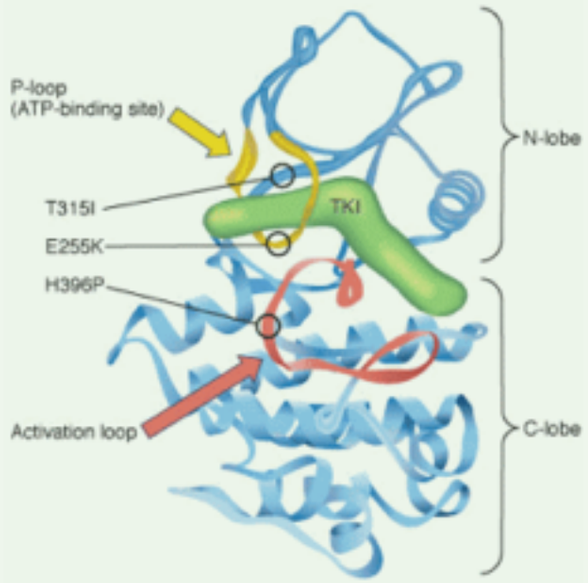
FACILITATING DESIGN



sampling over target mutations



sampling over substituents for ligand design



ligand design considering potential resistance mutations

Speeding up the cycle of learning from failure can accelerate progress toward rational ligand design

---

computational predictions

experimental confirmation

"Fail fast, fail cheap"

Make rigorous calculations of affinity fast and accurate

GPU acceleration

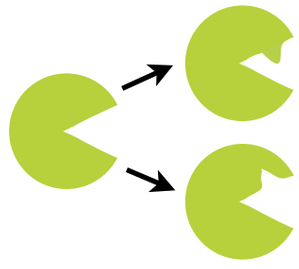
$$\pi(x)K(x,y) = \pi(y)K(y,x)$$

$$AH \leftrightharpoons A^- + H^+$$

modular MCMC for sampling and chemical effects

enhanced sampling algorithms

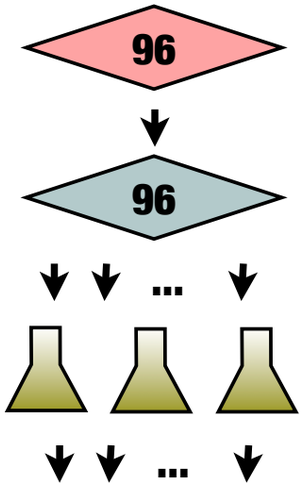
Test and improve models quickly and cheaply by inverting the drug discovery problem



mutate proteins instead of ligands

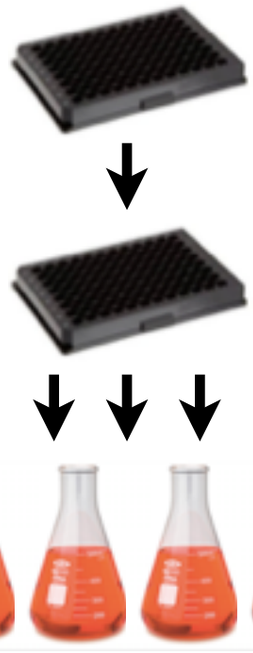
¢

buy inexpensive ligands



high-throughput experiments

Inverting the drug discovery problem:
A faster, cheaper way to build (and break) computational models



quick change mutagenesis

~ 1-2 weeks, >70% efficiency

expression assessment in 1 ml culture

~ 2-3 weeks, select mutants that express >25 ug/ml

expression and purification in ~2L culture

~ 4/week, >25 mg

+

purchase known ligands
several ligands/receptor

isothermal titration calorimetry
~40/day (automated); <0.5 mg/experiment

surface plasmon resonance
6 x 6 parallel experiments; ~10 ug protein

fluorescence binding assays
96 assays/plate; ~1 ug protein

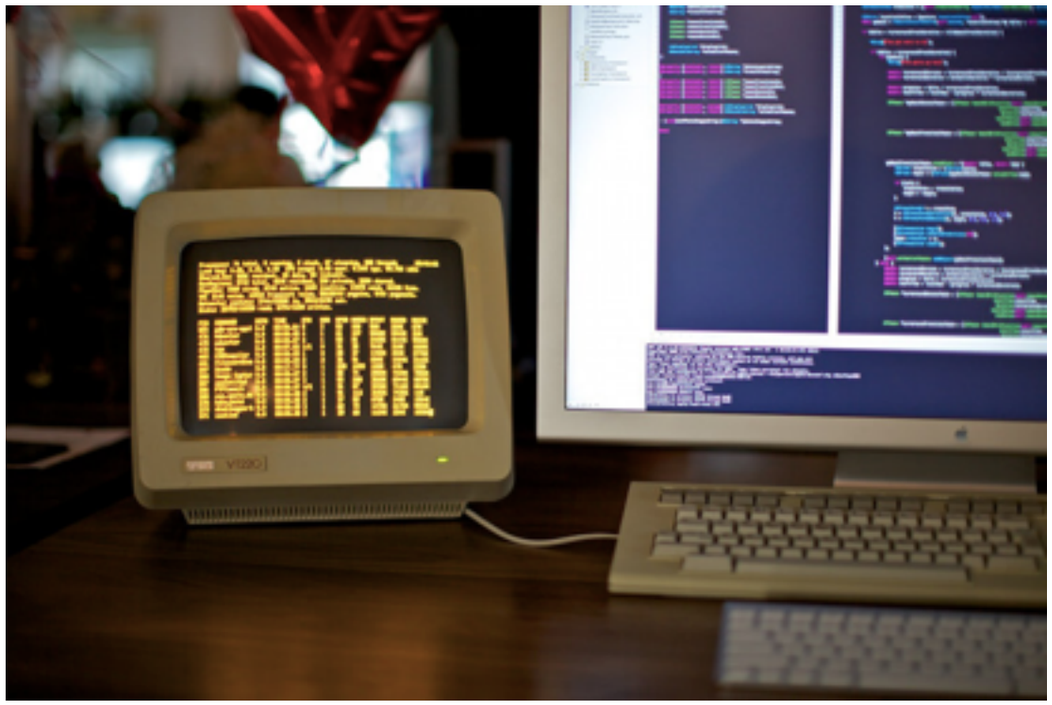
How can we make wetlab experiments look more like problems we know how to solve efficiently?

messy
laborious
inconsistent
skill-dependent
9 am - 5 pm

How can we make wetlab experiments look more like problems we know how to solve efficiently?

messy
laborious
inconsistent
skill-dependent
9 am - 5 pm

?



precise
structured
consistent
reproducible
round-the-clock

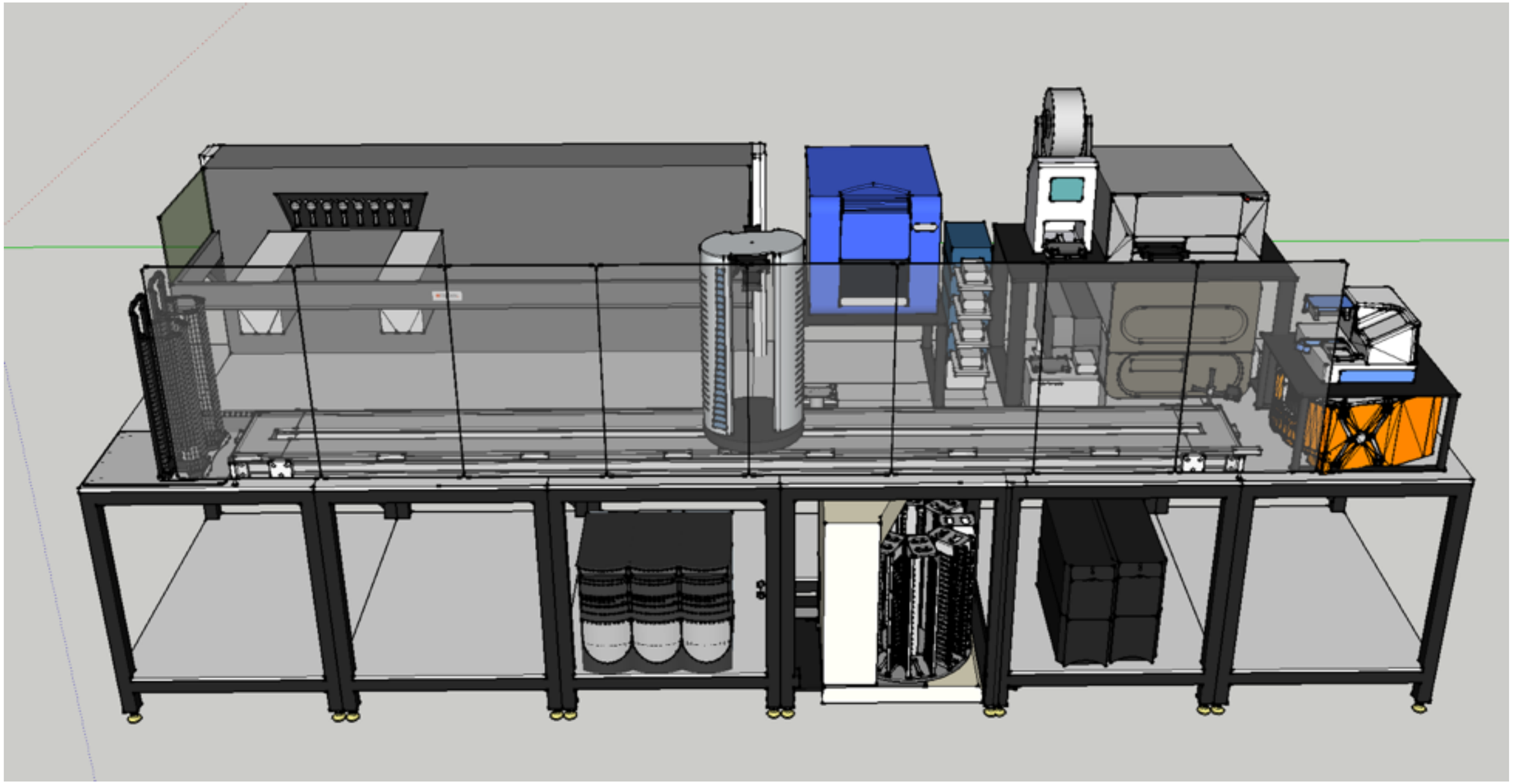
AUTOMATE. EVERYTHING.

Automated platform for bacterial cloning, mutagenesis, expression, purification, and binding affinity measurement with 24/7 operational capacity

AUTOMATE. EVERYTHING.

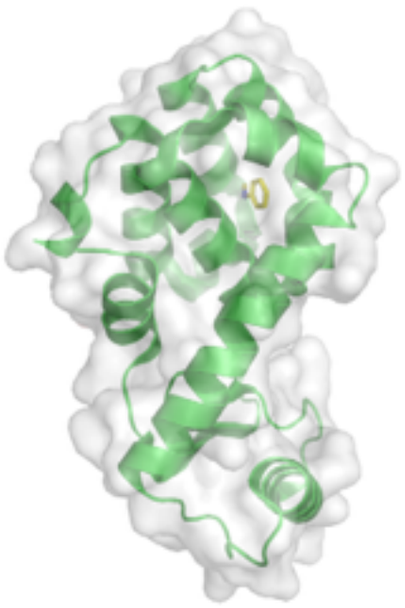
Automated platform for bacterial cloning, mutagenesis, expression, purification, and binding affinity measurement with 24/7 operational capacity

AUTOMATE. EVERYTHING.



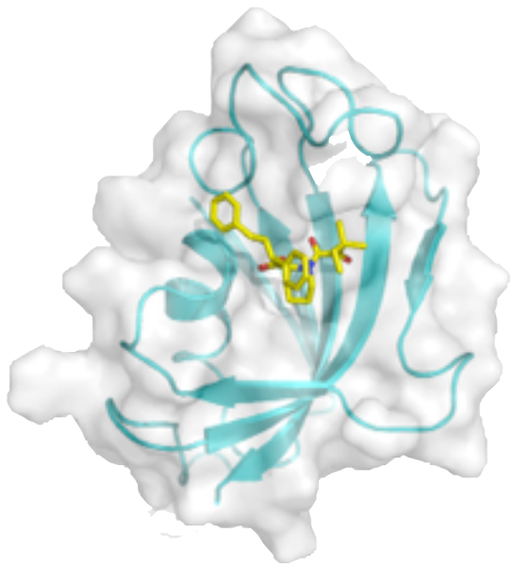
Automated platform for bacterial cloning, mutagenesis, expression, purification, and binding affinity measurement with 24/7 operational capacity

We're building a benchmark set covering a range of complexity, isolating individual challenges to modeling



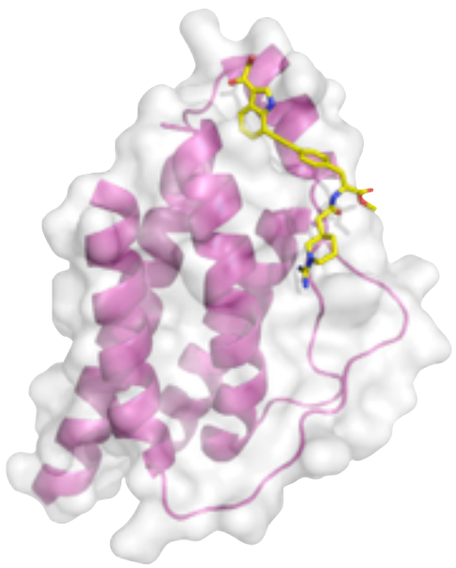
small apolar ligands
T4 lysozyme L99A

small, rigid protein
small, neutral ligands
fixed protonation states
multiple sidechain orientations
multiple ligand binding modes



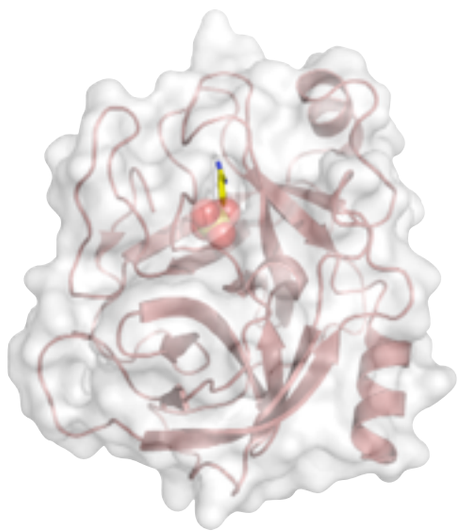
polar ligands
FKBP12

small, rigid protein
fixed protonation states
larger natural product-like
ligands with rotatable bonds



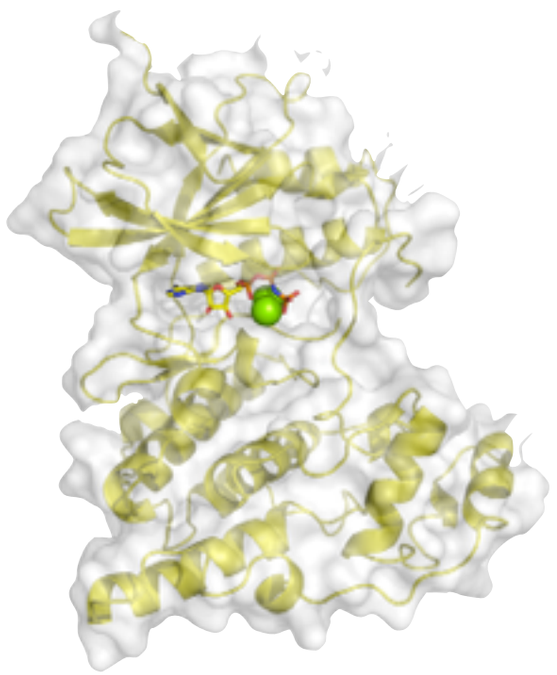
IL-2

small protein
fixed protonation states
some allostery and
binding site plasticity



trypsin proteases

small, rigid protein
small ligands
charged ligands
potential for protonation state changes



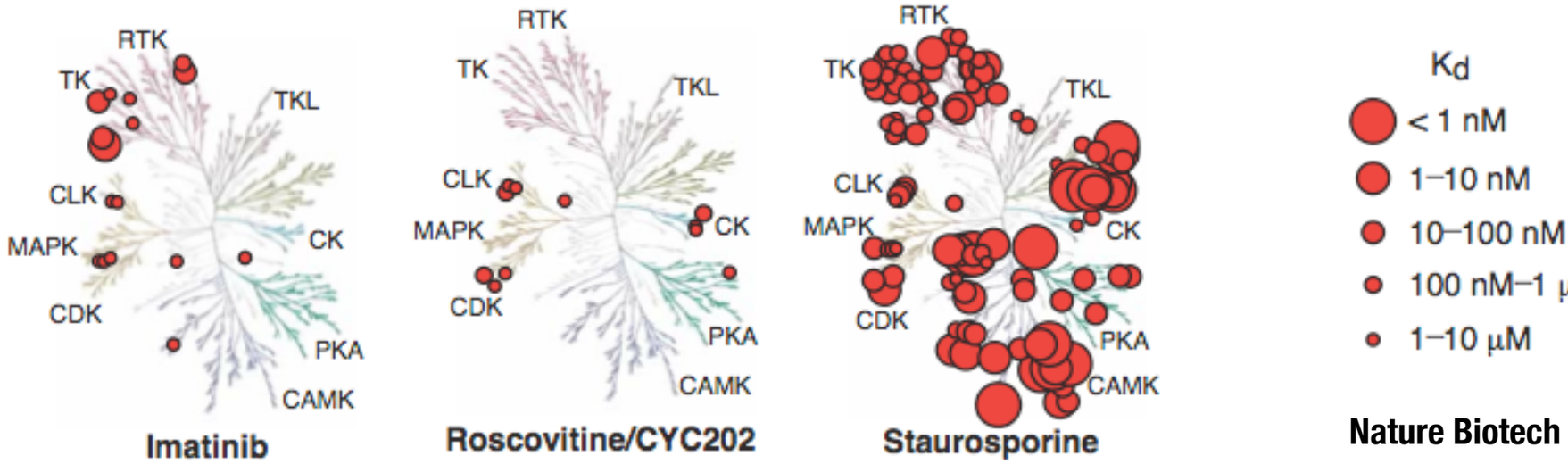
JNK3 kinase

large protein, multiple conformations
large drug-like ligands, rotatable bonds
multiple protonation states? tautomers?
phosphorylation and activation
peptide substrate?
$MgCl_2$ salt effects?

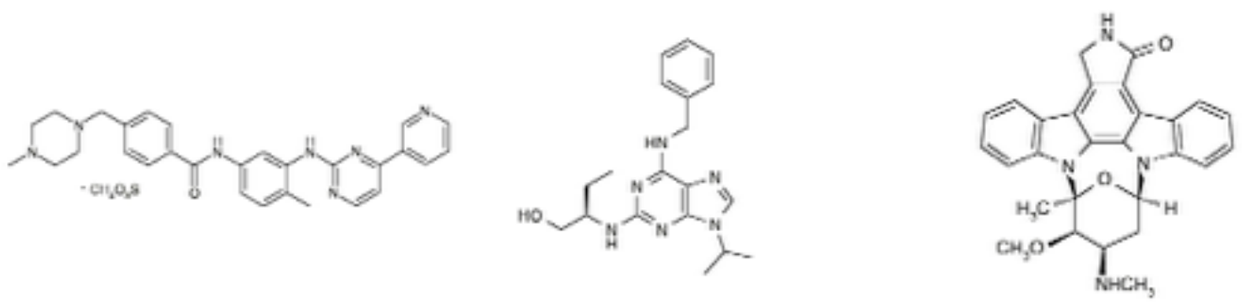
easy
hard

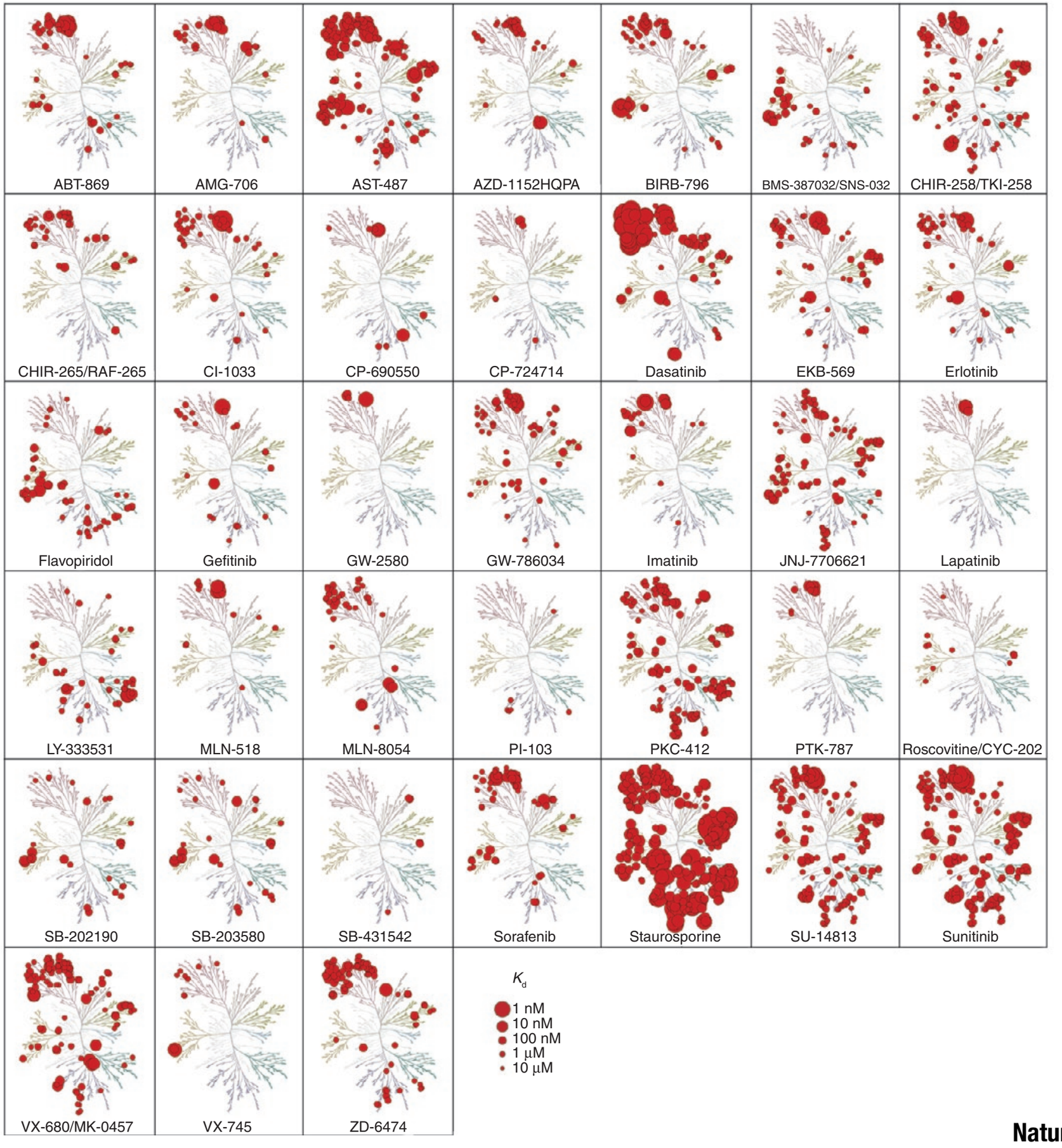
many other good model systems to choose from: DHFR, cytochrome C peroxidase, AmpC, adenylate kinase, etc.

How can we quantitatively understand (and later design) the selectivity of inhibitors for kinases?



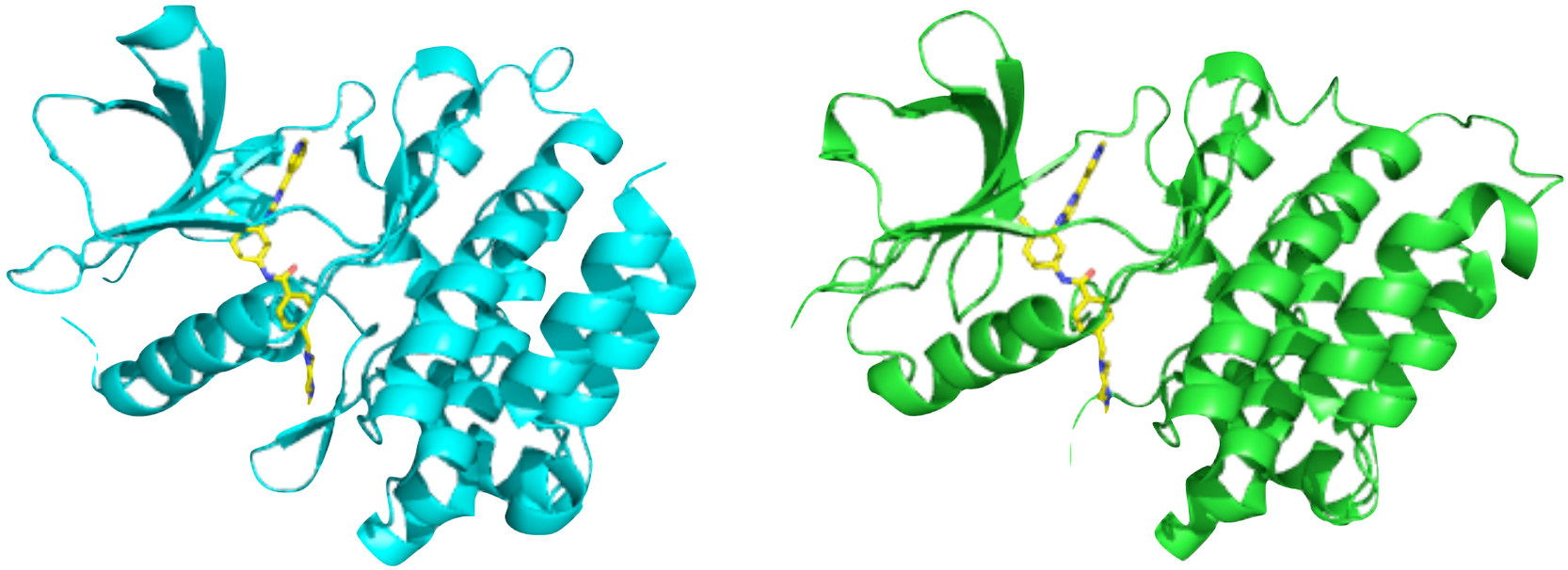
Nature Biotech 23:329, 2005





Nature Biotech 26:127, 2008

Differences in stabilities of inactive states may be responsible for origin of some kinase inhibitor selectivity



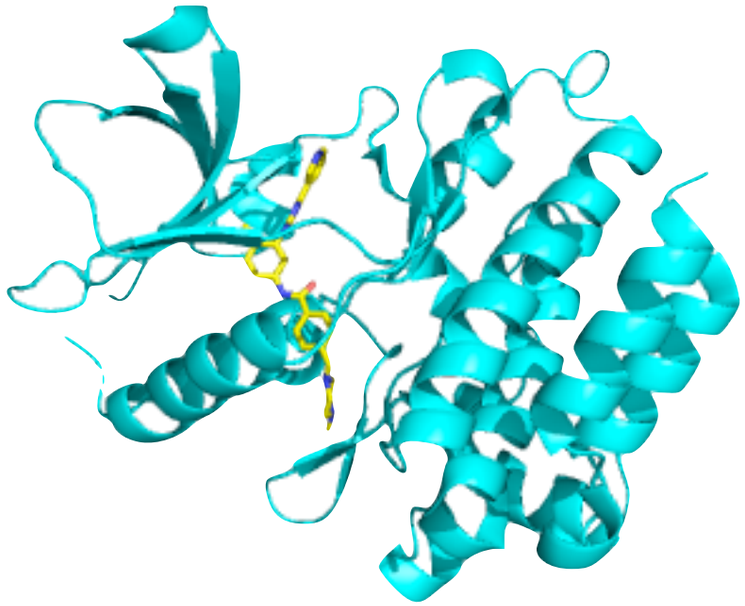
imatinib bound to Abl kinase [2HYY]

imatinib bound to Src kinase [2OIQ]

$$\Delta\Delta G = 4.6 \ \mathrm{kcal/mol}$$ **(favoring Abl binding)**

*** essentially same binding mode in X-ray structure!**
*** essentially same interactions**
*** calculations suggest no difference in binding free energy for this conformation**
[Aleksandrov and Simonson , J Biol Chem 285:13807, 2010]

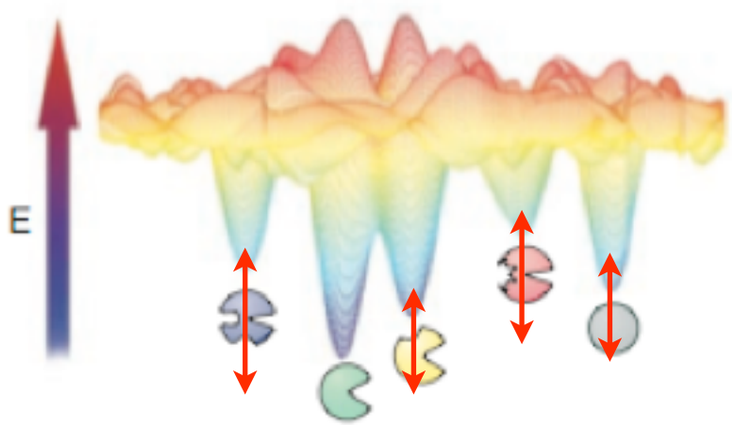
Differences in stabilities of inactive states may be responsible for origin of some kinase inhibitor selectivity



imatinib bound to Abl kinase [2HYY]

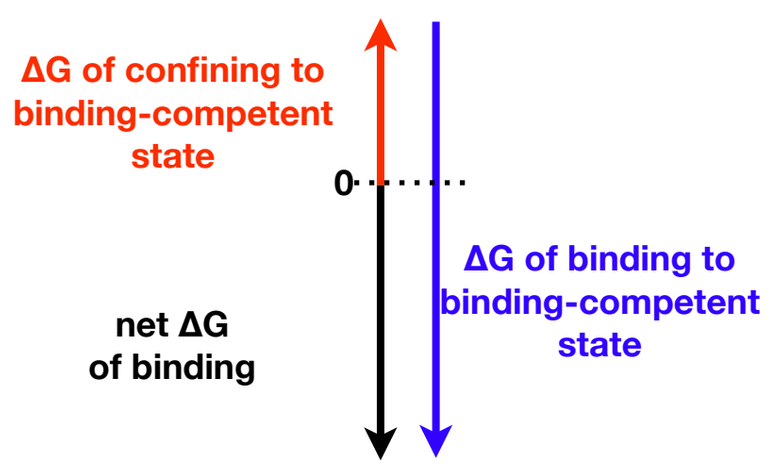
imatinib bound to Src kinase [2OIQ]

$$\Delta\Delta G = 4.6 \ \text{kcal/mol}$$ **(favoring Abl binding)**



inactive (high free energy states)

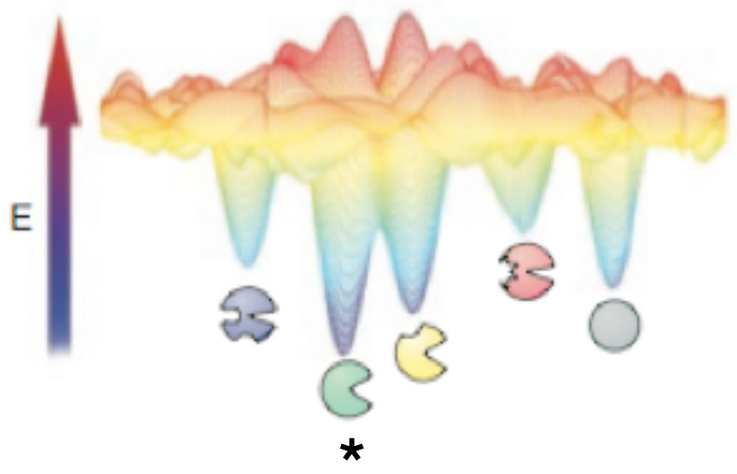
active (low free energy state)



Enumeration of metastable states will be crucial to successful design of selective kinase inhibitors.

Identifying these conformational substates structurally and energetically is often difficult

* allow substrate access
* expose binding surface
* expose allosteric regulatory site
* release product
* alter catalytic activity
* transduce signals



invisible / high free energy states

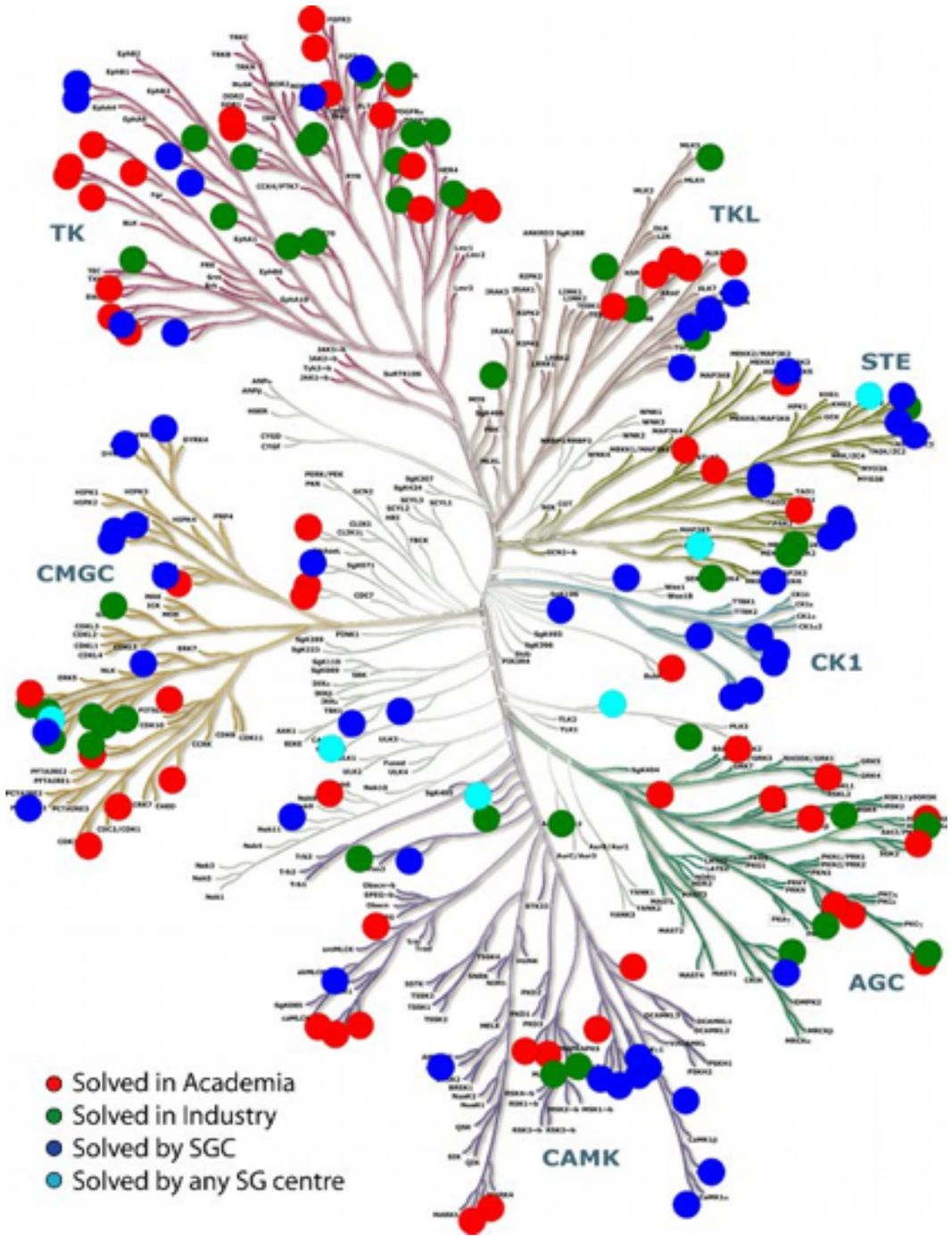
visible / low free energy state

Structural biology approaches can only “see” these states if fortuitously trapped by ligands
Simulations get “stuck” in these conformations for long times

Lee and Craik. Science 324:213, 2009

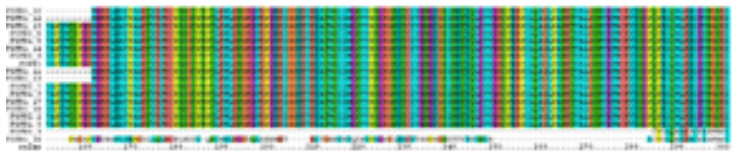
Structural data on human kinases exists, but is incomplete

Human kinases with available structural data



http://www.sgc.ox.ac.uk/research/kinases/

sequences of targets/off-targets



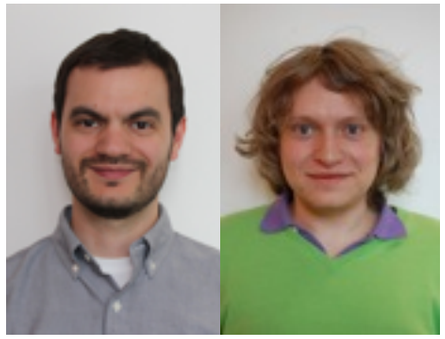
+

UniProt

RCSB PDB
PROTEIN DATA BANK

template structures from Uniprot

Daniel Parton
Postdoc



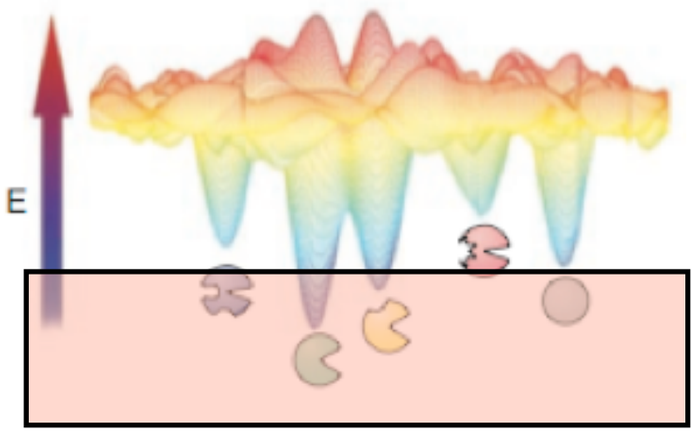
Patrick Grinaway
PBSB student

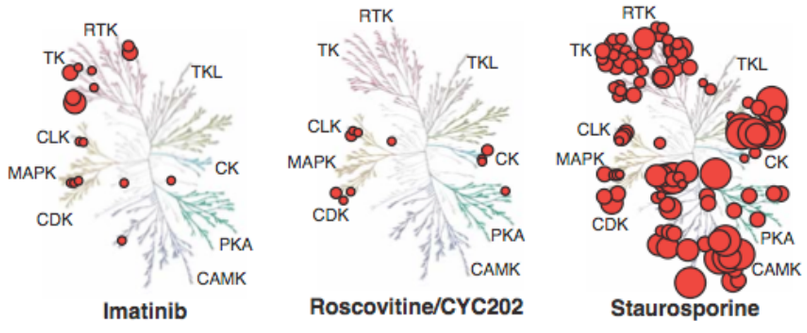
structural models of many conformations

Kyle Beauchamp
Postdoc

Jan-Hendrik Prinz
Postdoc

conformational ensembles with relative energetics





multiple binding free energy calculations to compute affinities and selectivities

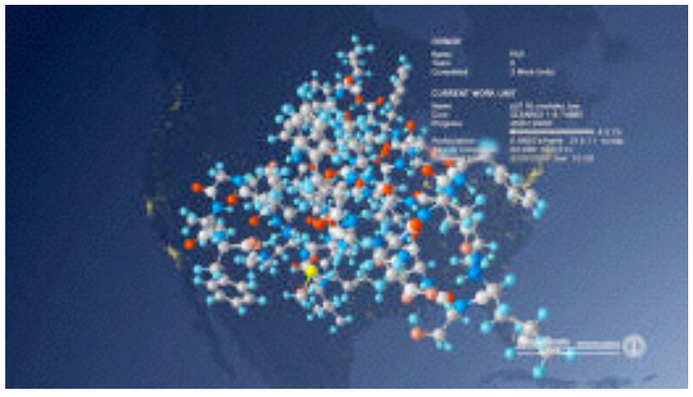
Folding@Home gives us access to enormous computational resources for probing biomolecular dynamics

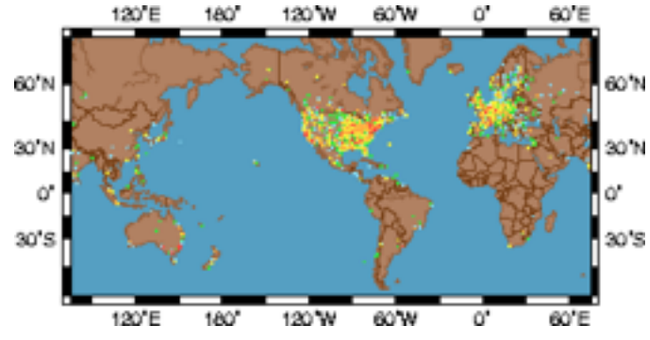
Folding@home DISTRIBUTED COMPUTING

Vijay S. Pande
Stanford University

OS Type	Native TFLOPS*	x86 TFLOPS*	Active CPUs	Active Cores	Total CPUs
Windows	359	359	82503	182921	5565251
Mac OS X	25	25	9100	82678	177010
Linux	38	38	12122	34440	785708
ATI GPU	1105	2332	7785	7785	391018
NVIDIA GPU	1960	4136	10372	10372	335374
NVIDIA Fermi GPU	18934	39951	55852	244235	492581
Total	22421	46841	177734	562431	7746942

Table last updated at Mon, 19 Jan 2015 23:02:19
http://fah-web.stanford.edu/cgi-bin/main.py?qtype=osstats2





Over 22 PFLOP/s of aggregate computational power!

http://folding.stanford.edu

Folding@Home enables whole-kinome simulation

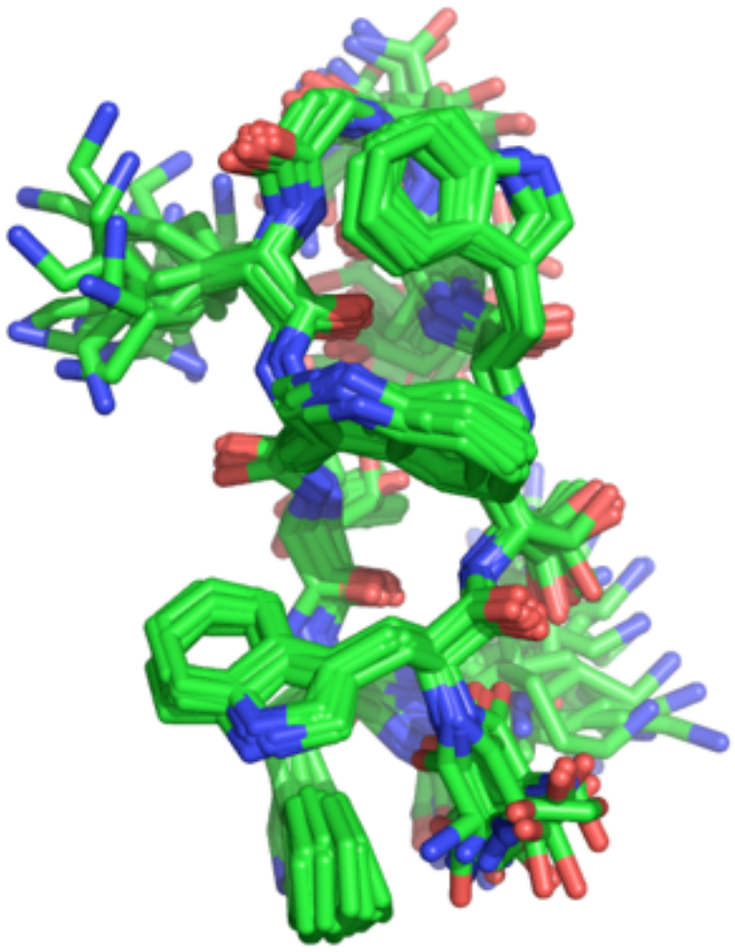
518 **human protein kinases**
excluding splice and disease variants

x **3,507** **kinase catalytic domain structures**
in UniProt

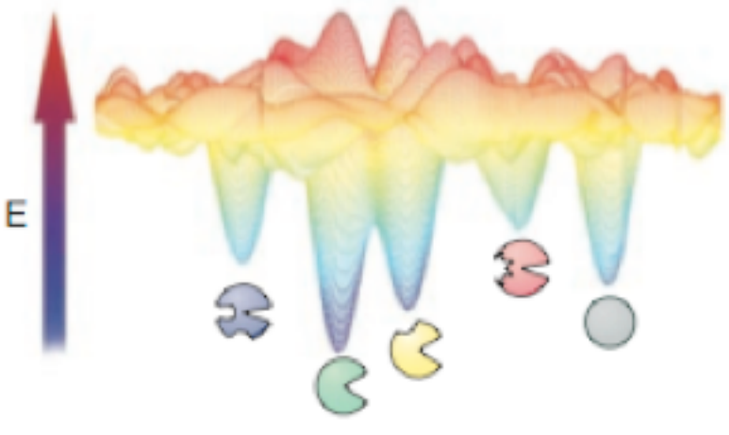
= **1,816,626** **kinase models will be built and refined**
on new MSKCC compute resources housed at SDSC

~ **18,166,260** **kinase simulations on Folding@Home**
over one year

NMR model of trpzip2 at 288 K



trpzip2
[PDB:1LE1]



Many distinct metastable states can be identified at T ~ T_m

backbone
Trp sidechains

"invisible" states (<1%)

8.1 kT	8.2 kT	8.8 kT	10.0 kT		
7.3 kT	7.3 kT	7.5 kT	7.7 kT	7.7 kT	7.9 kT
6.7 kT	6.8 kT	6.9 kT	6.9 kT	7.2 kT	7.3 kT
6.5 kT	6.5 kT	6.5 kT	6.6 kT	6.7 kT	6.7 kT
6.2 kT	6.2 kT	6.3 kT	6.3 kT	6.4 kT	6.4 kT
4.8 kT	5.6 kT	5.8 kT	5.9 kT	5.9 kT	6.0 kT

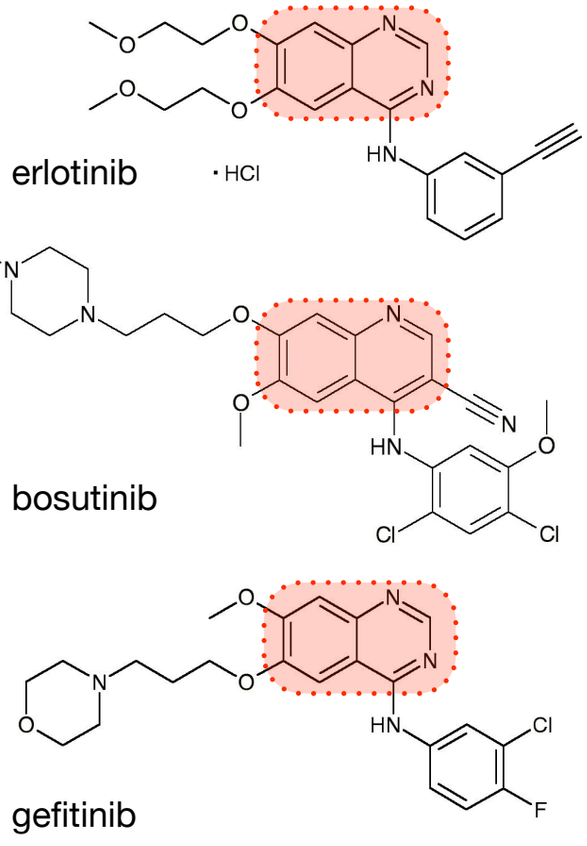
"ground states"

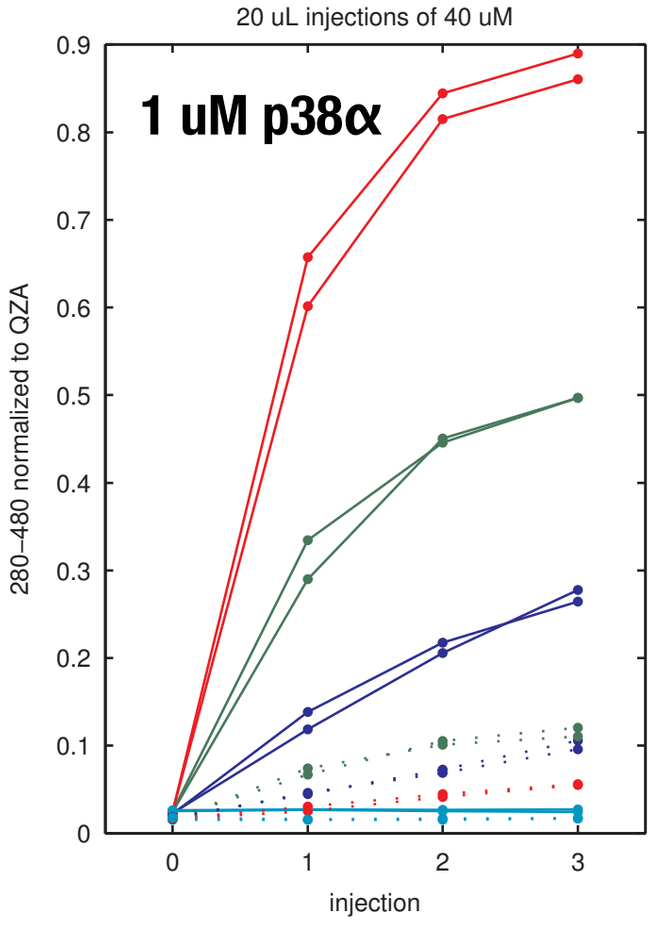
0.0 kT	0.2 kT	2.5 kT	3.0 kT	3.9 kT	4.7 kT

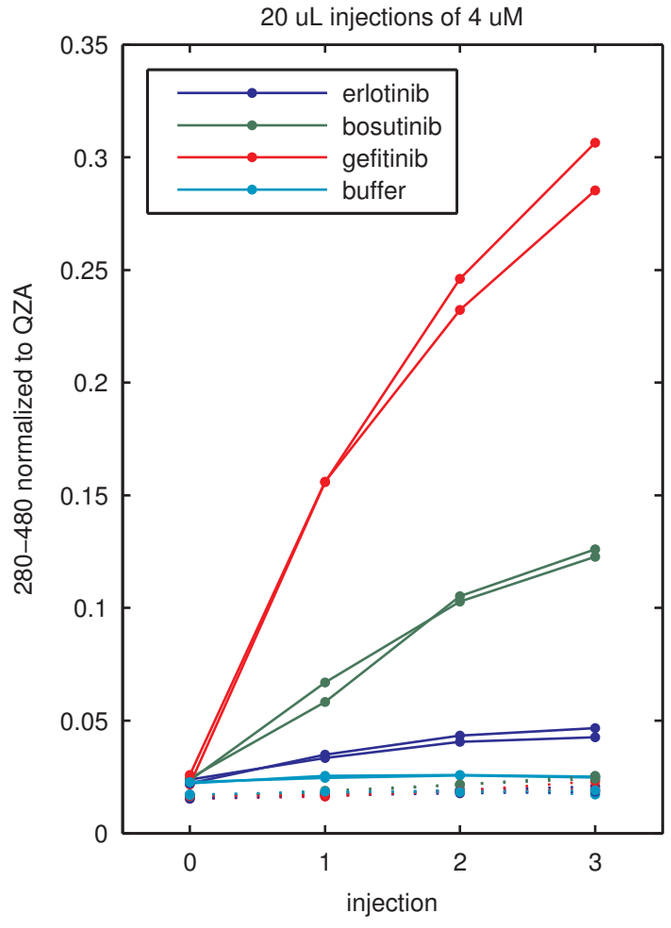
Chodera, Singhal, Swope, Pande, Dill. JCP 126:155101, 2007.

High-throughput fluorescence assays can measure binding affinities of a panel of ligands to kinase mutants

fluorescent inhibitors





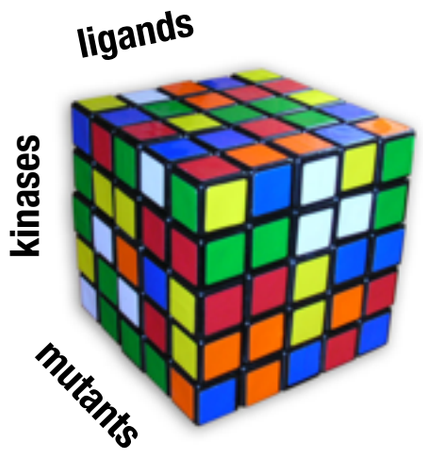


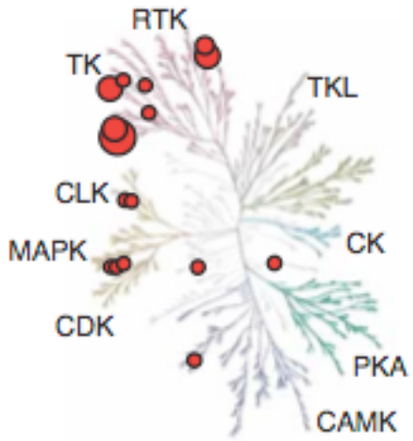
excite 280 nm
[Trp FRET ~350 nm]
measure 480 nm

with Nick Levinson, Boxer lab, Stanford

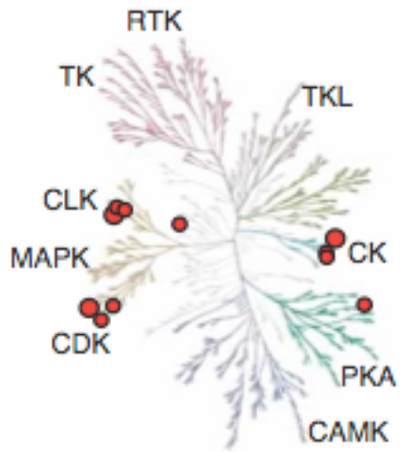
What determines selectivity of inhibitors for kinases?

High-throughput fluorescence measurements and free energy calculations can address physical determinants of kinase inhibitor selectivity:

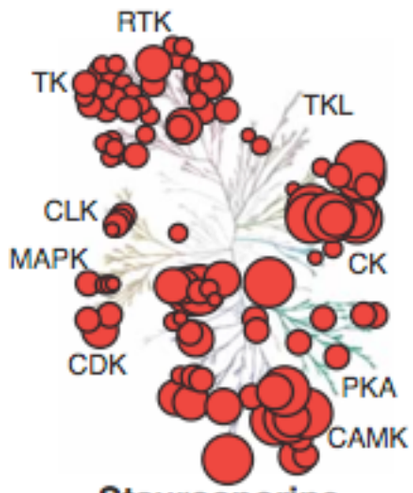




Imatinib



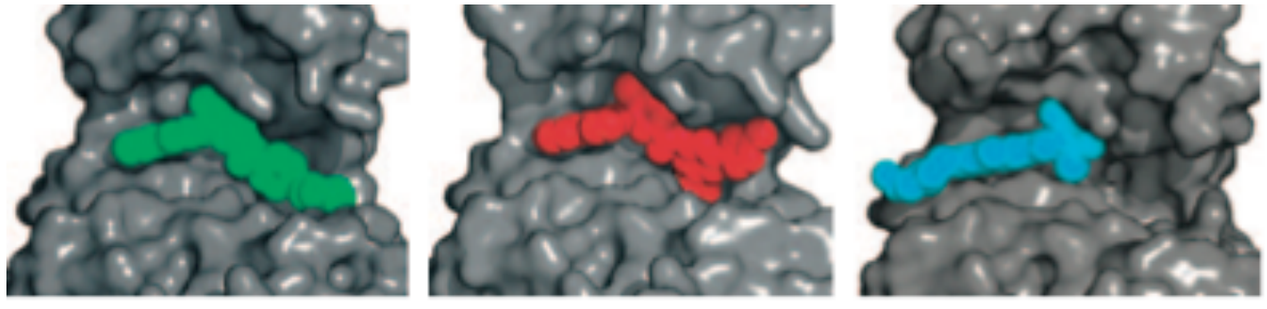
Roscovitine/CYC202



Staurosporine

Nature Biotech 23:329, 2005

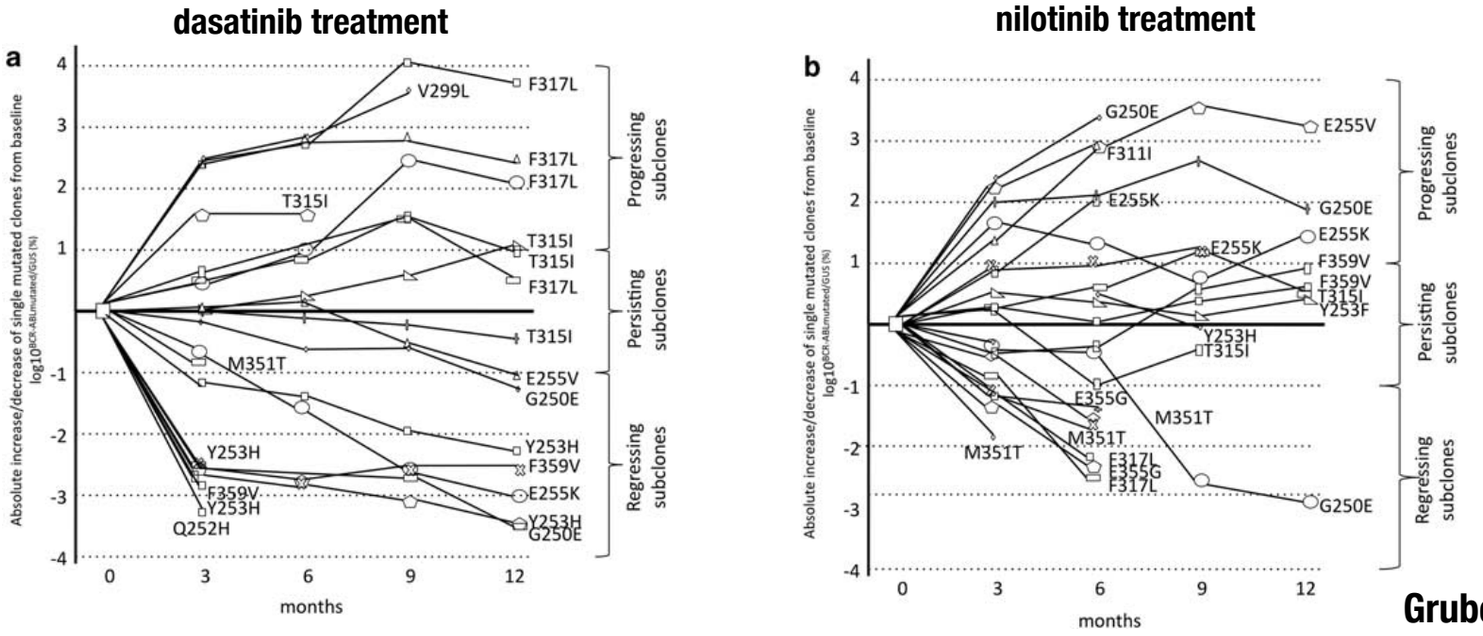
* Are particular ligand scaffolds privileged with specificity?
* Are particular binding modes better for specificity?
* Are certain kinases inherently more promiscuous?



Weisberg et al.
Nature Rev. Cancer 7:345, 2007.

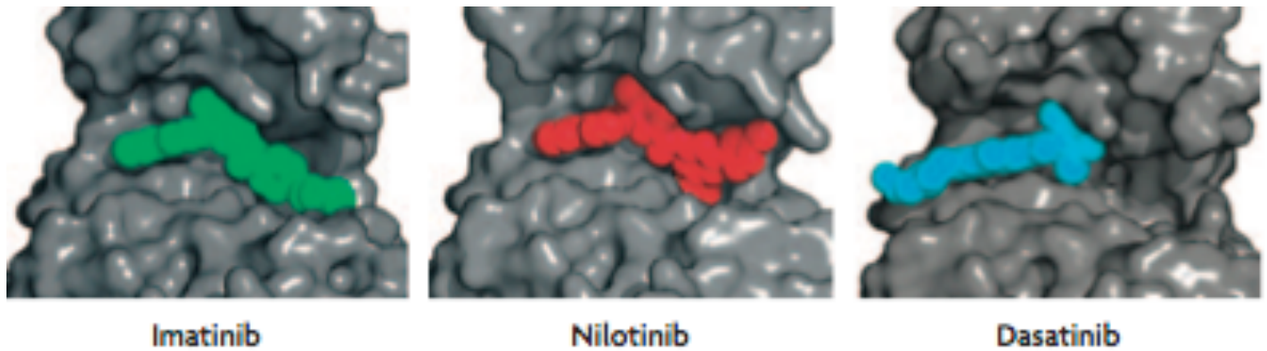
Can we develop a physical model of resistance mutations?

Treatment of CML with imatinib often induces resistance, predominantly E255K, T315I
Second-line drugs elicit further resistance:



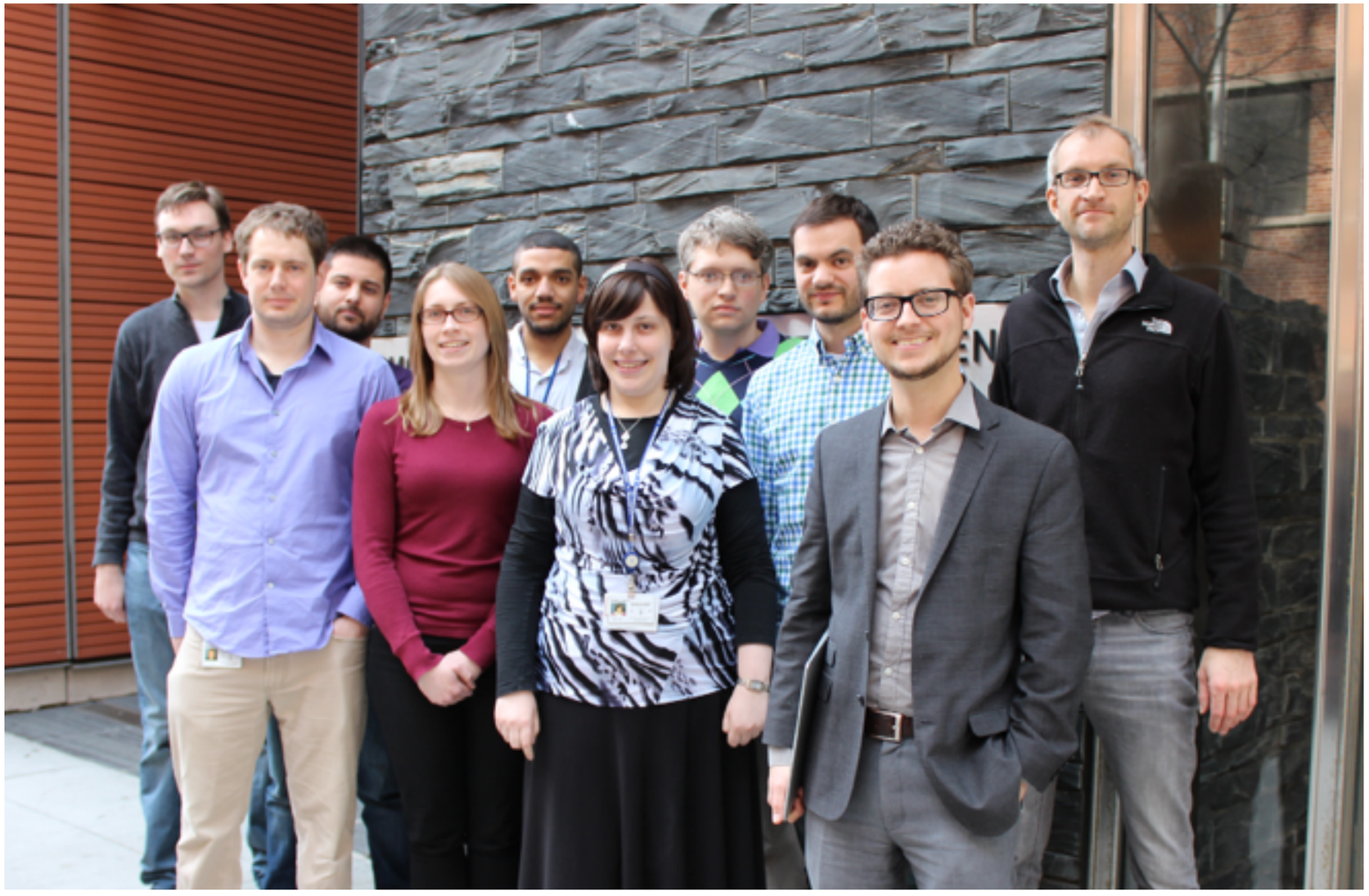
Gruber et al. Leukemia 26:172, 2012.

We can hypothesize and test a simple physical mechanism of resistance:
Resistance mutations reduce inhibitor binding affinity but retain ATP affinity (a surrogate for activity)
*** Are certain inhibitors or binding modes less likely to elicit resistance?**
*** Can we incorporate likelihood of eliciting resistance mutations into rational ligand design?**



Weisberg et al.
Nature Rev. Cancer 7:345, 2007.

The Chodera Lab @ MSKCC



Bas Rustenburg
PBSB

Guillermo Perez-Hernandez
visitor

Kadeem Ho Sang
admin

Danny Parton
postdoc

Jan-Hendrik Prinz
postdoc

Patrick Grinaway
PBSB

Kyle Beauchamp
postdoc

Sonya Hanson
postdoc

Chaya Stern
TPCB

Code and data available at http://www.choderalab.org
Start Folding at http://folding.stanford.edu

Collaborators

Stanford
Vijay Pande
Sergio Bacallado
NIH SimBios

IBM Almaden
Bill Swope
Jed Pitera

University of Chicago
Nina Singhal Hinrichs

UC Irvine
David Mobley

CCNY
Marilyn Gunner

Stony Brook
Ken Dill
Markus Seeliger

UCSF
Brian Shoichet
David Sivak

University of Virginia
Michael Shirts

Duke
David Minh

Freie Universität Berlin
Frank Noé
Bettina Keller
Jan-Hendrik Prinz
Antonia S. J. S. Mey

Rutgers
Zhiqiang Tan

UC Berkeley
Phillip Elms (BioRad)
Susan Marqusee
Carlos Bustamante
Christian Kaiser
Gheorghe Christol
Suri Vaikuntanathan

LBNL
Gavin Crooks

Vanderbilt
Joel Tellinghuisen

Hessian Informatics
Kim Branson

Vertex Pharmaceuticals
Richard Dixon

Code and data available at http://www.choderalab.org
Start folding at http://folding.stanford.edu