YANK: Free energy calculations made useful.

Kim Branson
Vertex Pharmaceuticals

John D. Chodera
QB3-Berkeley Fellow

Vijay Pande
Stanford University

Free energy: The view from pharma...

- Pharmaceutical research would love to have a fast accurate and reliable free energy tool.

- Usually the old engineering joke is fast accurate reliable, cheap, pick two...

- For free energy you get to pick one. If you are lucky.

- They've heard it all before. They live in the 'real world' apparently.

Ocker:
Software for discovery of inhibitors of protein-protein

---

ocker (slang)
noun

1. An oafish uncultured Australian man.
   Form: Ocker (often)

adj

1. Typical of an ocker; boorish; uncultured.

Etymology: 1970s: a form of the name Oscar, the name of a TV character.

Other: a name for a computational drug design program that was initially thought to be a clever play on DOCK, but seems less amusing now.

www.simtk.org/ocker

Yank: Ligand binding free energy calculations in implicit solvent

Project Overview

Description Yank is a new code for estimating free energies of ligand binding using free energy perturbation and implicit solvent models, utilizing the OpenMM GPU-accelerated dynamics toolkit. While not as accurate as explicit solvent simulations, the myriad of advantages offered by implicit solvent (including constant pH treatments) and speed advantages offered by GPU acceleration are expected to provide significant utility for medicinal chemists.

Purpose/Synopsis Provide a computational tool for rapidly estimating protein-ligand binding free energies using implicit solvent models

Audience Molecular modelers interested in virtual screening methods for identification of small molecule ligands of biomolecules

Long Term Goals and Related Uses Provide a computational tool for rapidly estimating protein-ligand binding free energies using implicit solvent models

Project Lead


John Chodera
Contact


Vijay Pande
Contact


Kim Branson
Contact


Imran Haque
Contact

SimTK, the Simulation Toolkit, is a part of the Simbios project funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 GM072970. Information on the National Centers for Biomedical Computing can be obtained here.

The drug development cycle

- **Identify the therapeutic target**
- **Find from (biopiracy, patents, screening [uHTS, virtual], academia) a lead compound.**
- **Develop lead compound to have appropriate potency for the target, selectivity, and pharmacokinetic and metabolic profiles**
- **Ideally its a short synthesis with no stereo centres, and has a single crystal allomorph.**
- **Formulate and hand it over to the crack team of cheerleaders and bribery experts we call pharmaceutical sales.... ***

* except at vertex, because we are ethical..

So where would we use free energy ?

potency	selectivity / DPMK	formulation

- Of the hundreds of derivatives I could make, which one should I make? Ranking congeneric series is very hard
- Should I put a stereocenter in? will this give me selectivity?
- Can I lock in a tautomeric state? Will this help me? Is this a bio-isostere?
- How can I make a change to avoid a metabolic liability and not loose potency?
- What are the most stable crystal forms for this molecule?
- What is the solubility of the molecule from each crystal form?

This free energy you speak of, is it fast ?

- No. Not currently.
- Yeah we tried that ages ago, took a long time and didn’t work
- “I’m never writing AMBER inputs again.”
- But I can just do the experiment and have the ‘right answer”.
- Takes too long. I get paid to make molecules!

In short: We need to turn around an accurate free energy calculation in a day.

The aim of YANK:
Fast, accurate binding free energies in a day.

- At Vertex we deploy tools in two forms. “EZ” such as “EZDock” or “EZSim”

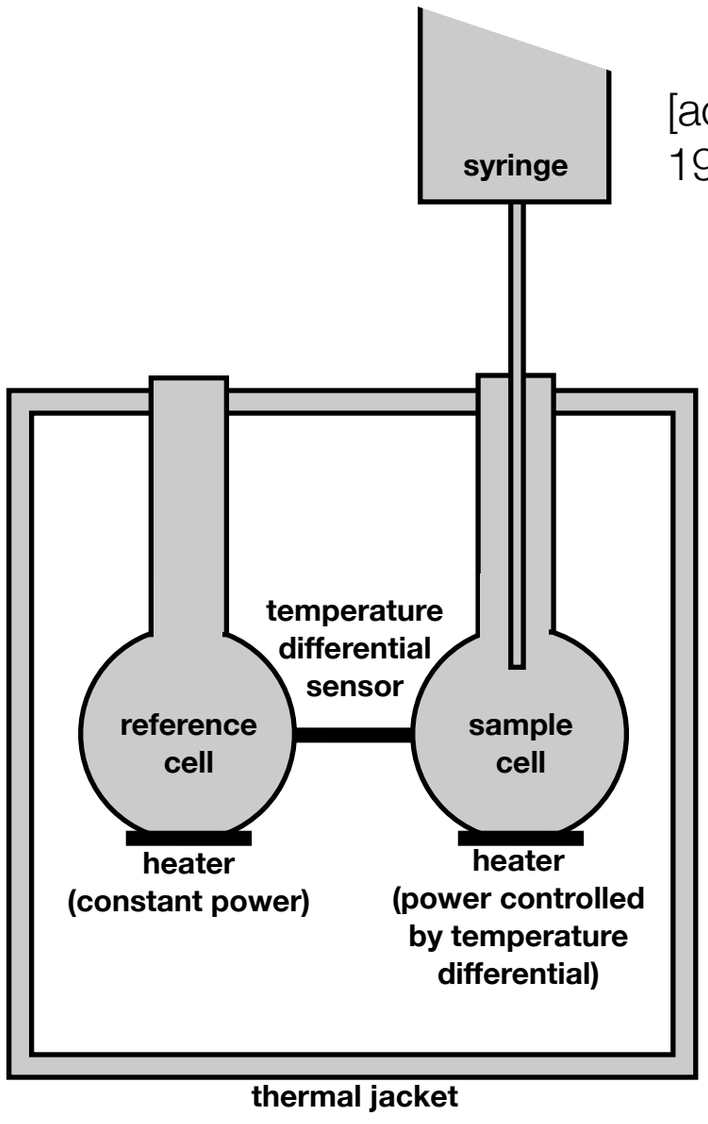
- This allows chemist to perform their own calculations. “Professional” modelers use more detailed tools for project specific support.

- If we can do a compound a day on a GPU, we can easily scale this up by using more GPU’s/compound.

- Having an actual energy function that is fast and works would allow alot of useful modeling that we currently don’t do.. (Computational fragment screening, maybe making docking work better than ligand based methods)

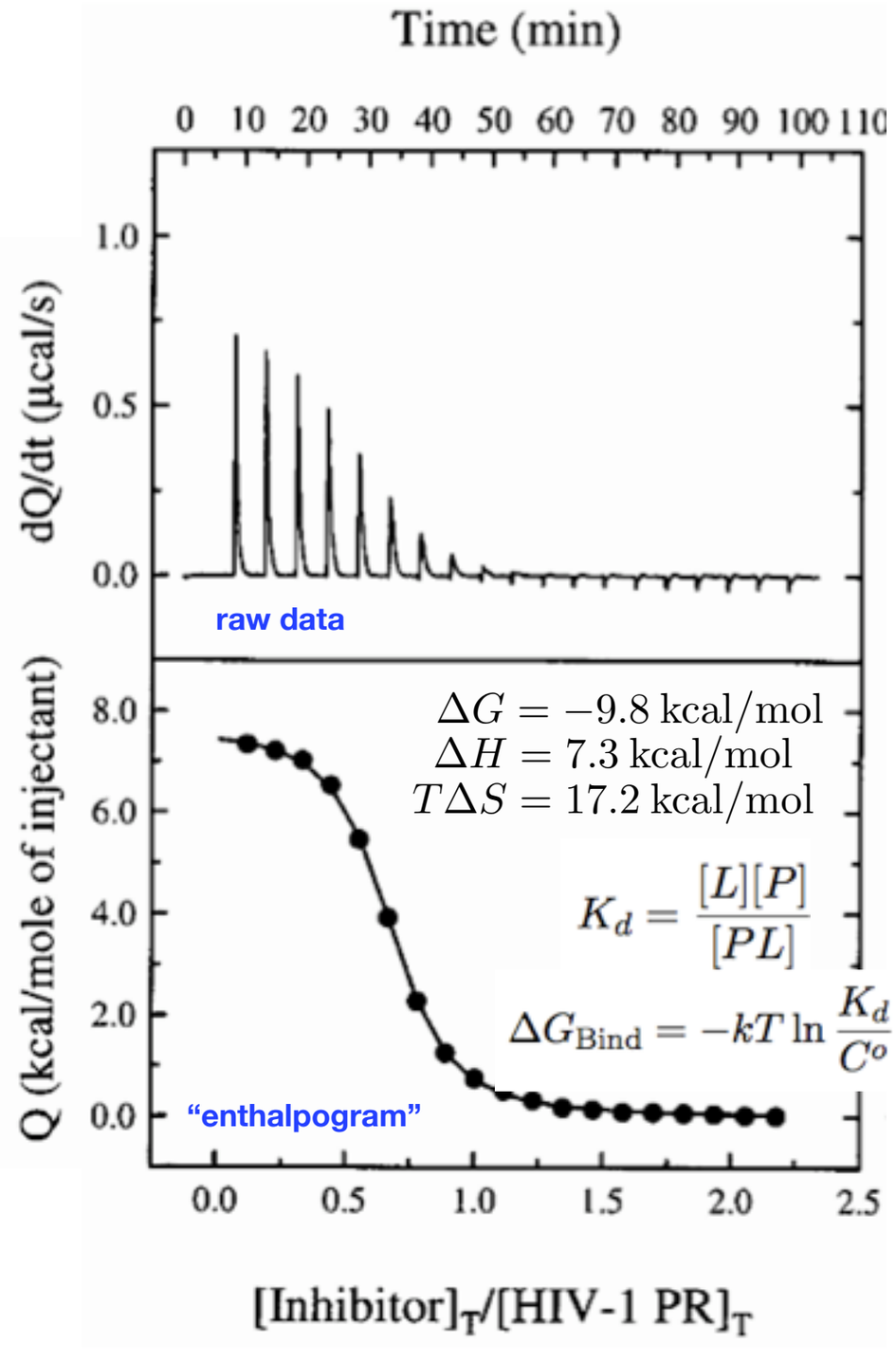
- Chemists might start with the most soluble lead, rather than the most potent one if they were confident they could easily build in potency. (They always use the most potent one, because its hard to say to your manager, “we’ve got great ADME/Tox we just need to work on the potency....”)

Binding affinities can be (in)directly measured by experiment



[acetyl pepstatin] = 200 uM
19 injections @ 10 uL/injection

[HIV-1 PR] = 20 uM in 1.4 mL
(0.6 mg/cell)



Simultaneously provides estimates of both ΔG and ΔH (and hence TΔS).

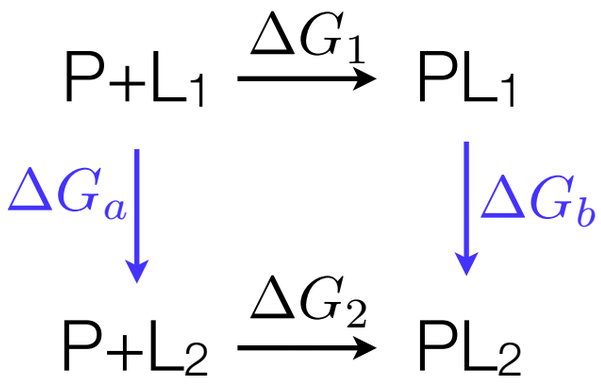
(Note that some reactions have no measurable change in heat, and are not measurable by ITC.)

Velazquez-Campoy A, Kiso Y, and Friere E. Arch. Biochem. Biophys. 390:169, 2001.

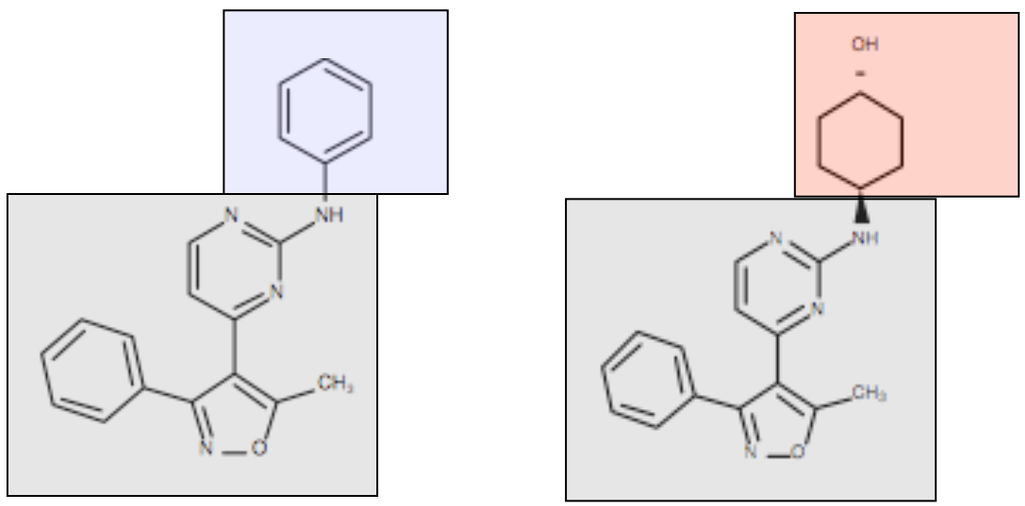
Binding affinities can be directly computed through equivalent alchemical thermodynamic cycles

---

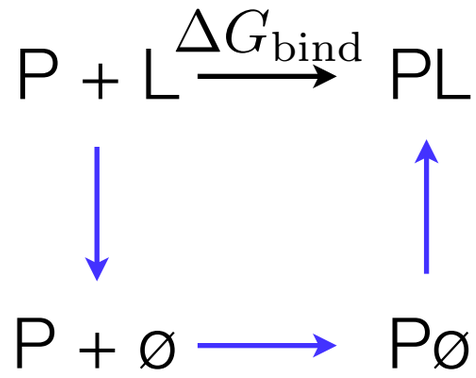
Relative free energies of binding

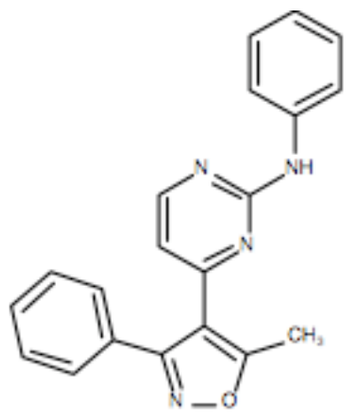


$$\Delta\Delta G = \Delta G_a - \Delta G_b$$



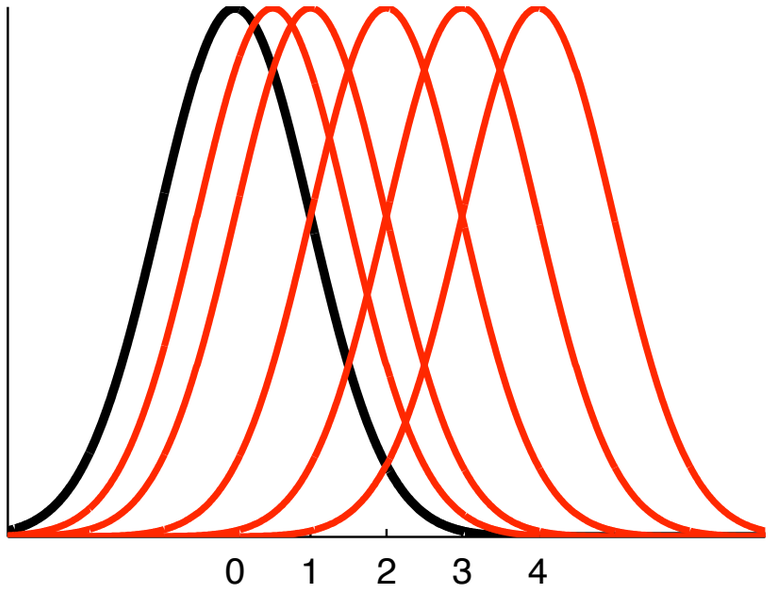
Absolute free energies of binding

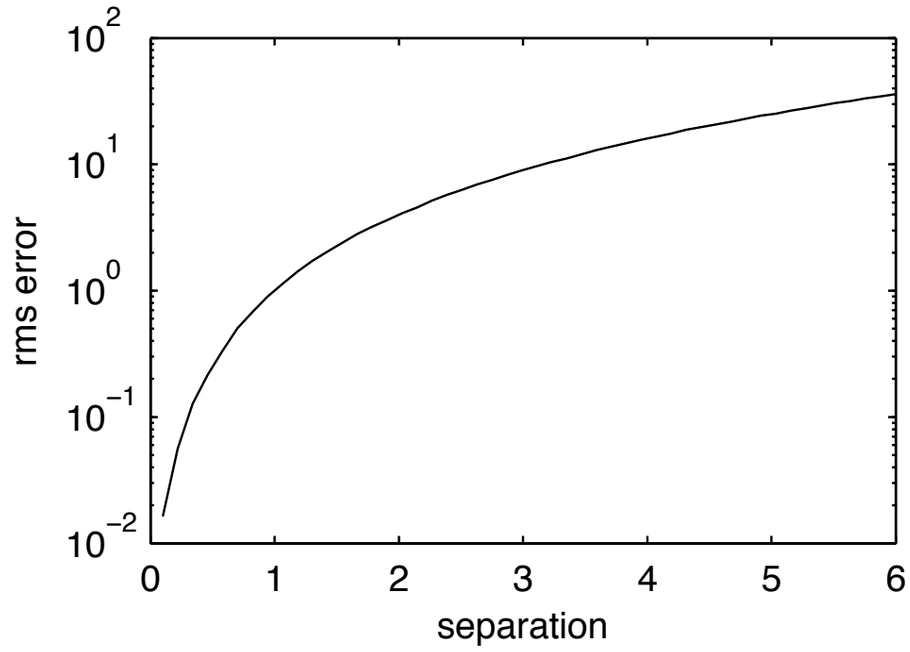




Alchemical intermediates can facilitate convergence

Error increases rapidly with diminishing phase space overlap





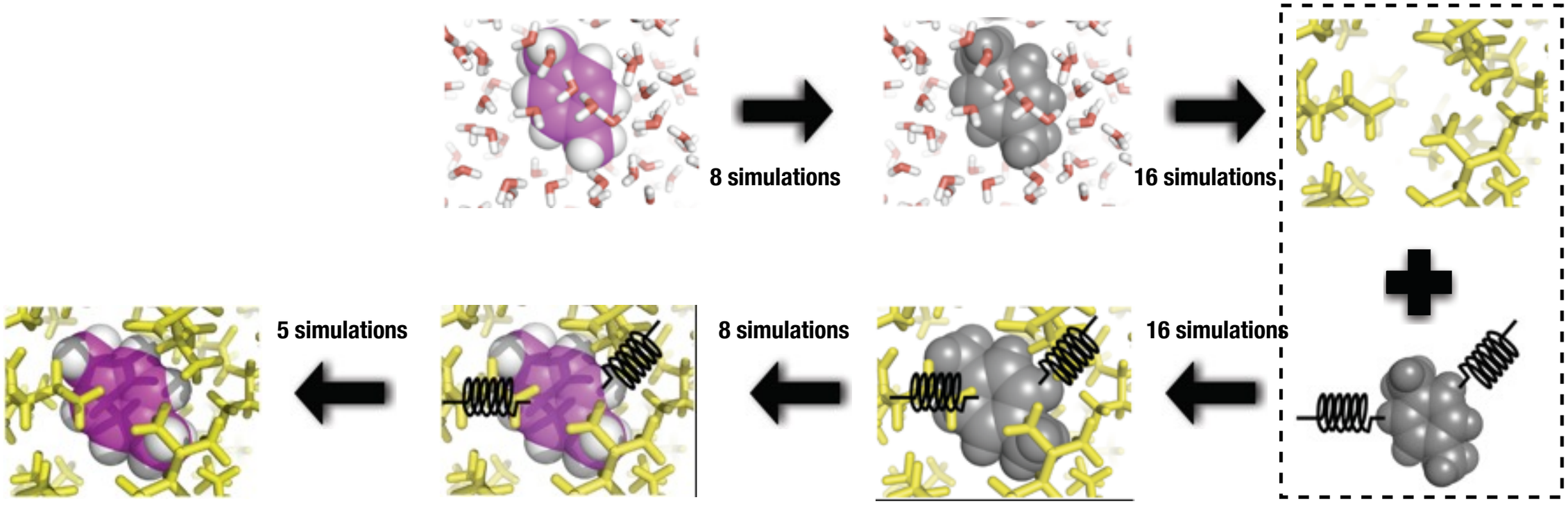
Instead, introduce intermediate states to ensure a contiguous chain of good overlap

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

D. Wu and D. A. Kofke. "Phase-space overlap measures. I. Fail-safe bias detection...", *J. Chem. Phys.* **123**: 054103 (2005).

Absolute alchemical free energy calculations involve simulations at multiple thermodynamic states

Alchemical transformation progresses through a number of intermediates

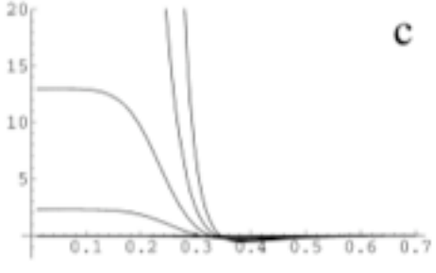


Free energy differences for each step are estimated from equilibrium simulations of intermediates

electrostatic annihilation $q_i \longrightarrow \lambda_{elec} \cdot q_i$ (for ligand charges)

Lennard-Jones decoupling $U_{ij}^{LJ}(r^{ij};\lambda) = \lambda^n 4\epsilon_{ij}\left(\frac{1}{[\alpha_{LJ}(1-\lambda)^2+(r_{ij}/\sigma_{ij})^6]^2} - \frac{1}{\alpha_{LJ}(1-\lambda)^2+(r_{ij}/\sigma_{ij})^6}\right)$

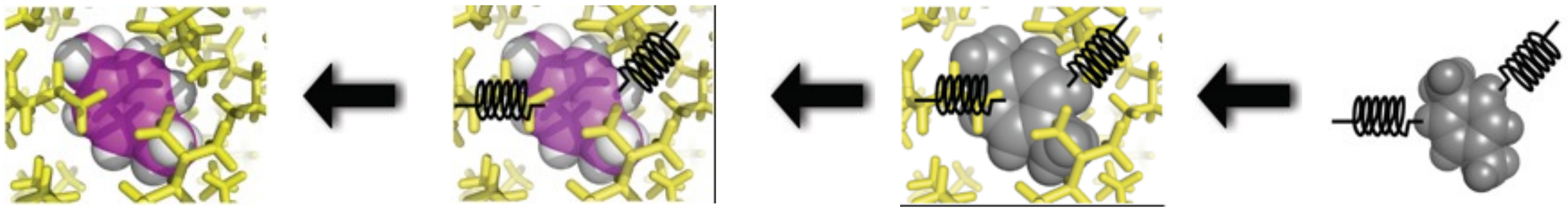
(for ligand-environment interactions)



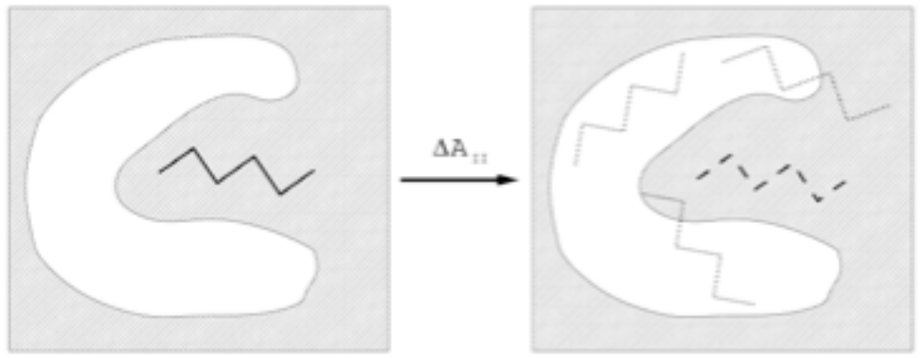
Each simulation must sample from equilibrium!

Graphics from David Mobley

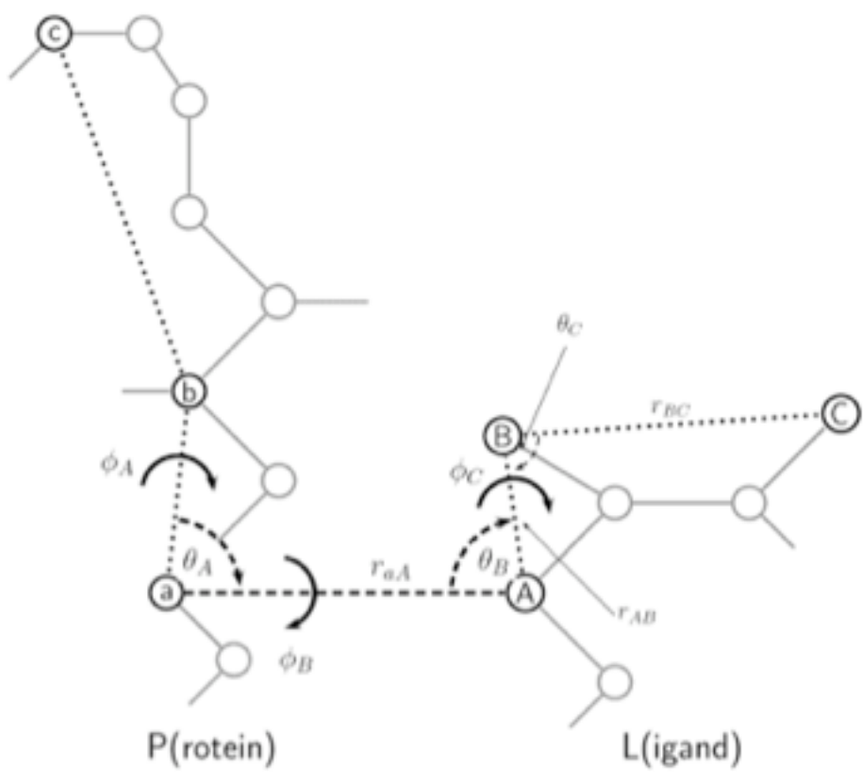
Restraints are used to aid convergence



Without restraining ligand in binding pocket, would need to sample entire simulation box at each discharging/decoupling intermediate



Choice of atoms to restrain is arbitrary in principle, minor practical differences among choices



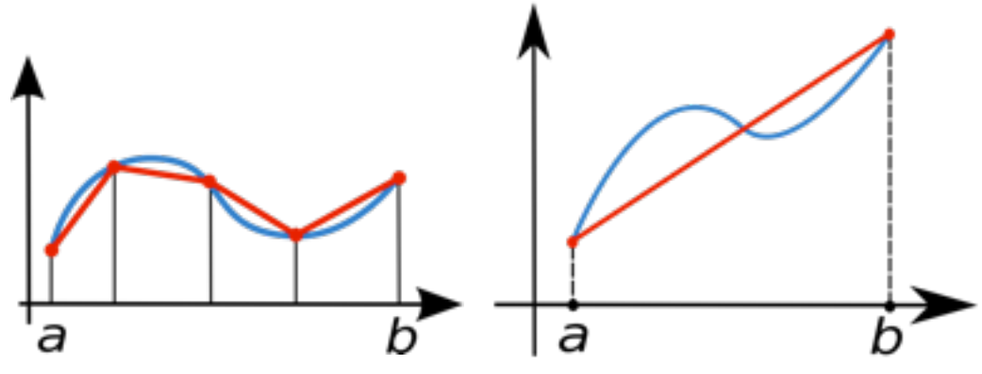
Absolute Binding Free Energies: A Quantitative Approach for Their Calculation
Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M.
J. Phys. Chem. B.; **(Article); 2003**; 107(35); 9535-9551. DOI: 10.1021/jp0217839

Free energy differences can be estimated in several ways

TI (thermodynamic integration)

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

quadrature error difficult to quantify

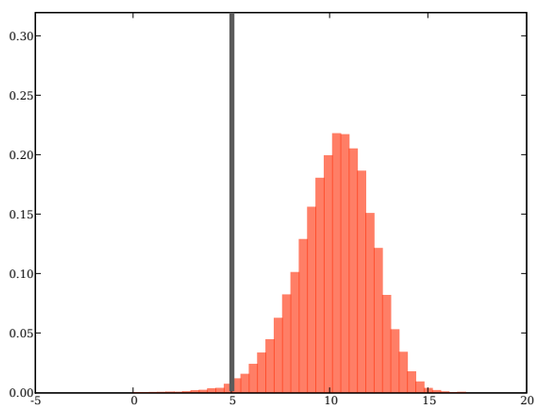


EXP (exponential reweighting)

$$\Delta F = -\beta^{-1} \ln \left\langle e^{-\beta(U_2 - U_1)} \right\rangle_{\lambda_1} = +\beta^{-1} \ln \left\langle e^{-\beta(U_1 - U_2)} \right\rangle_{\lambda_2}$$

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

suffers from large bias and variance

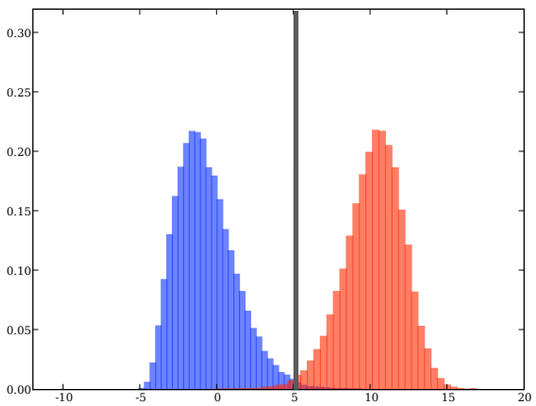


BAR (Bennett acceptance ratio)

$$\Delta F = -\beta^{-1} \ln \frac{\langle f(U_2 - U_1) \rangle_{\lambda_1}}{\langle f(U_1 - U_2) \exp[-\beta(U_2 - U_1)] \rangle_{\lambda_2}}$$

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

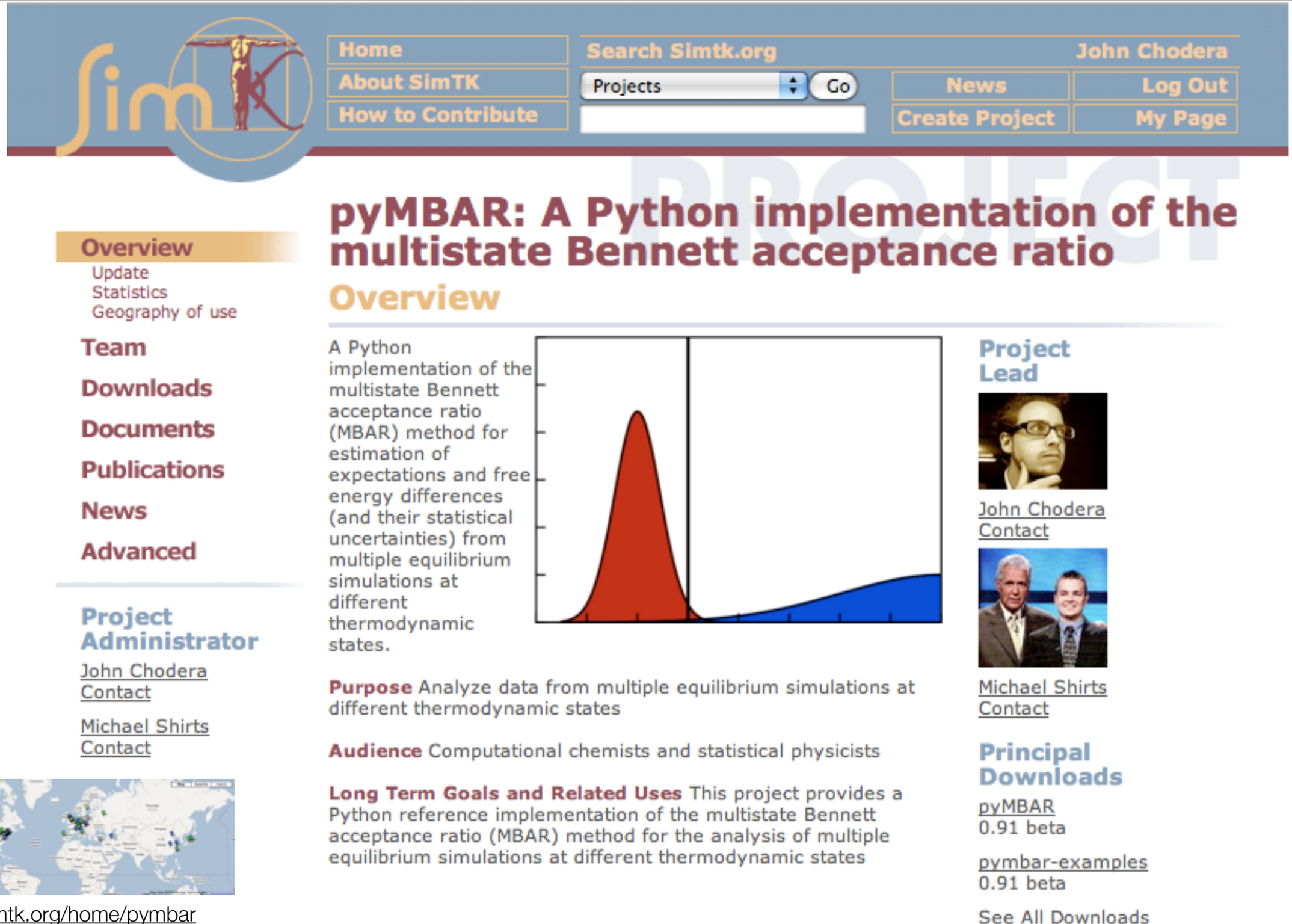
only applicable to two states



[+ Michael R. Shirts]

Python implementation of MBAR available now at

http://simtk.org/home/pymbar

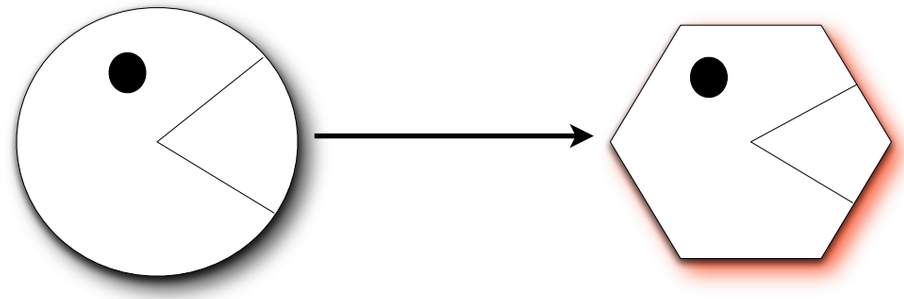


https://simtk.org/home/pymbar

Checklist of potential concerns in binding calculations

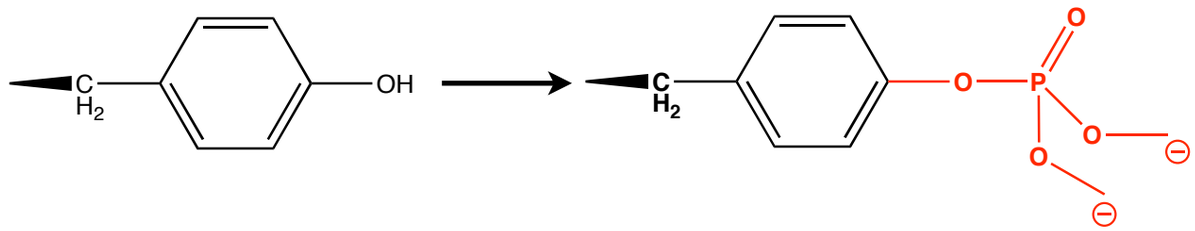
Protein conformation

- Which conformation is most likely?
- Conformational change upon binding
- Multiple conformations contributing to binding



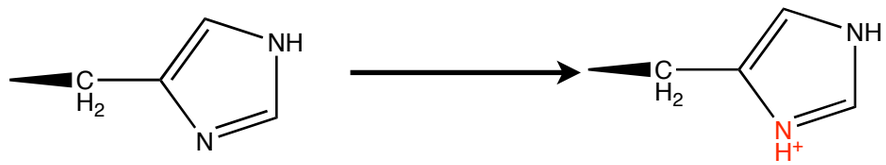
Post-translational modifications

- Phosphorylation, glycosylation, acylation, alkylation



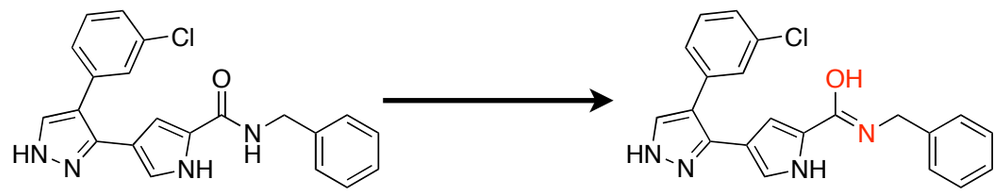
Protein protonation state

- Appropriate choice of protonation state
- Change in protonation state upon binding
- Mixture of protonation states relevant to binding



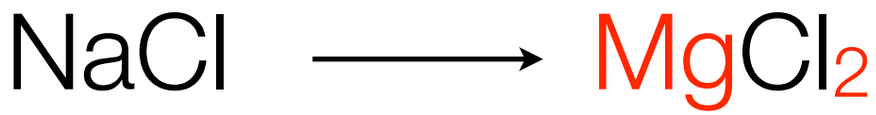
Ligand protonation/tautomeric state

- Appropriate choice of protonation/tautomeric state
- Change in protonation/tautomeric state upon binding
- Mixture of protonation/tautomeric states relevant to binding



Salt environment

- Salt required for function
- Appropriate salt parameters
- Other cosalts, cosolvents, and chelators



Checklist of potential concerns in binding calculations

Ligand parameter assignment

Anecdotal reports of Antechamber issues

ANTECHAMBER & GAFF

http://amber.scripps.edu/antechamber/

Protein forcefield choice

parm96 deprecated; parm03 unvalidated for free energies

ffAMBER

ffamber96, ffamber99sb, ffamber03

http://chemistry.csulb.edu/ffamber/

modified amino acid parameters

Don't have time to rederive appropriate

Only found parameters for parm99

AMBER parameter database

http://www.pharmacy.manchester.ac.uk/bryce/amber

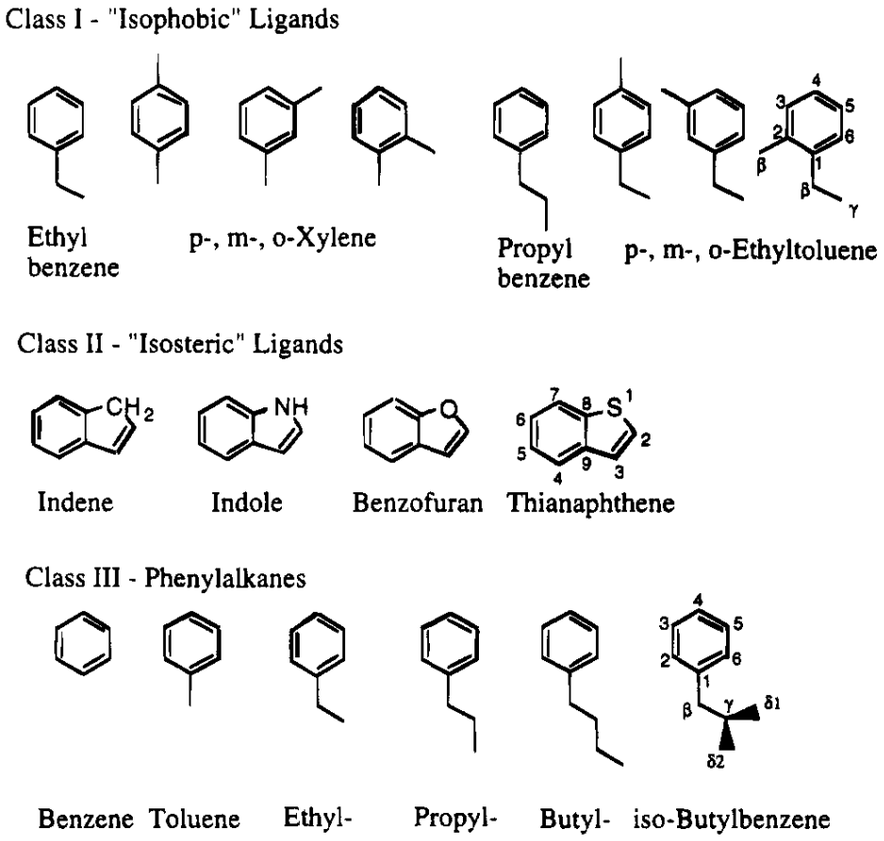
cofactors or other peptides bound?

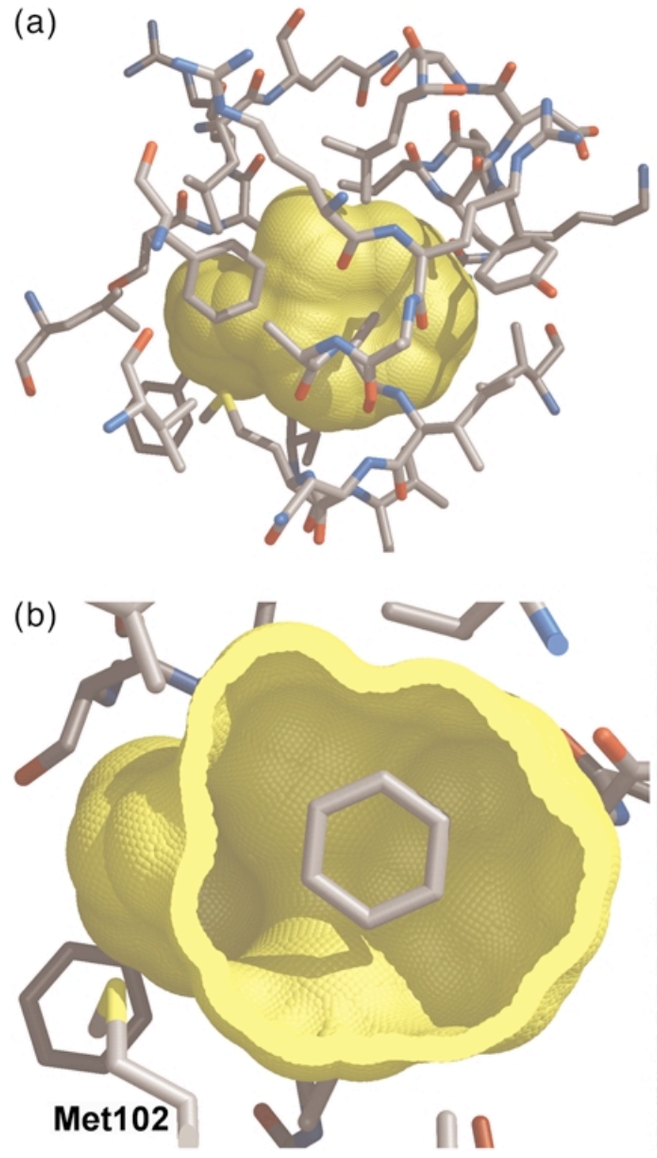
Simulation timescales

Can we converge estimates for even a single conformation state?

Image by Leo Reynolds

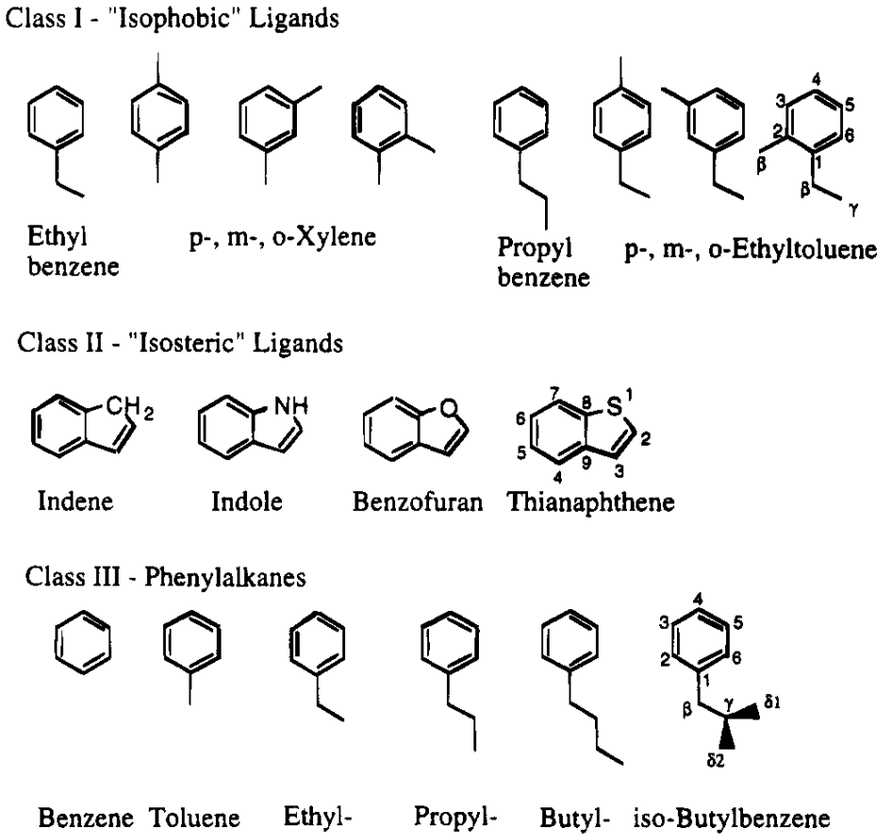
T4 lysozyme L99A:
A model binding site for small hydrophobic ligands

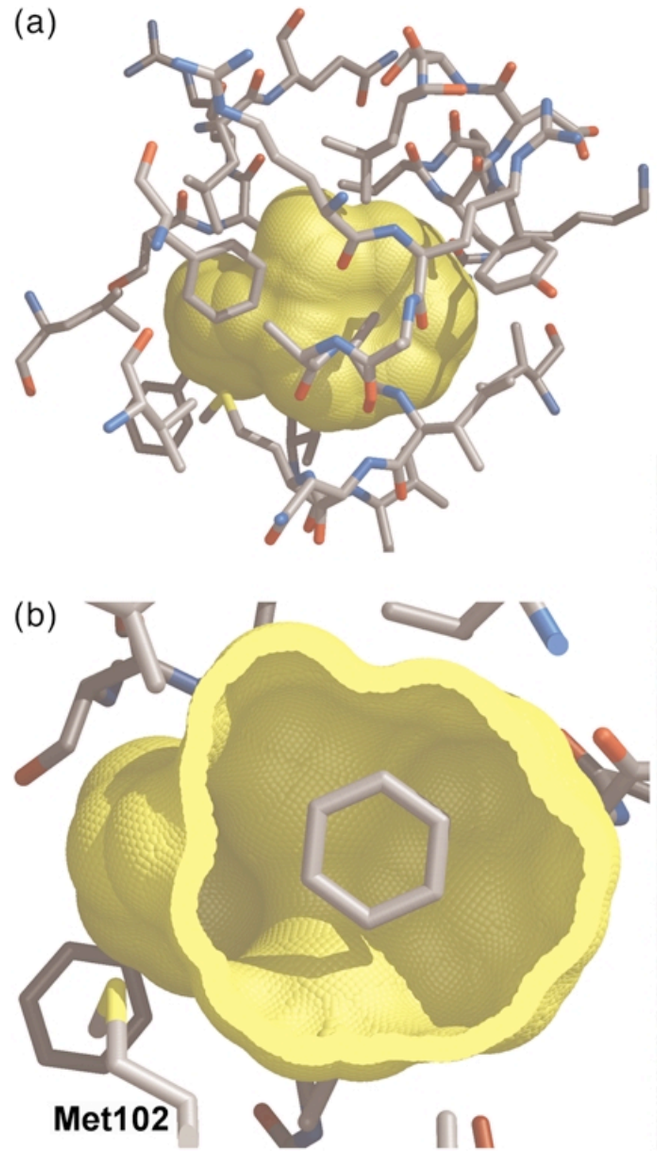




Wei, Baase, Weaver, Matthews, and Shoichet. JMB 322:339, 2002.

T4 lysozyme L99A:
A model binding site for small hydrophobic ligands

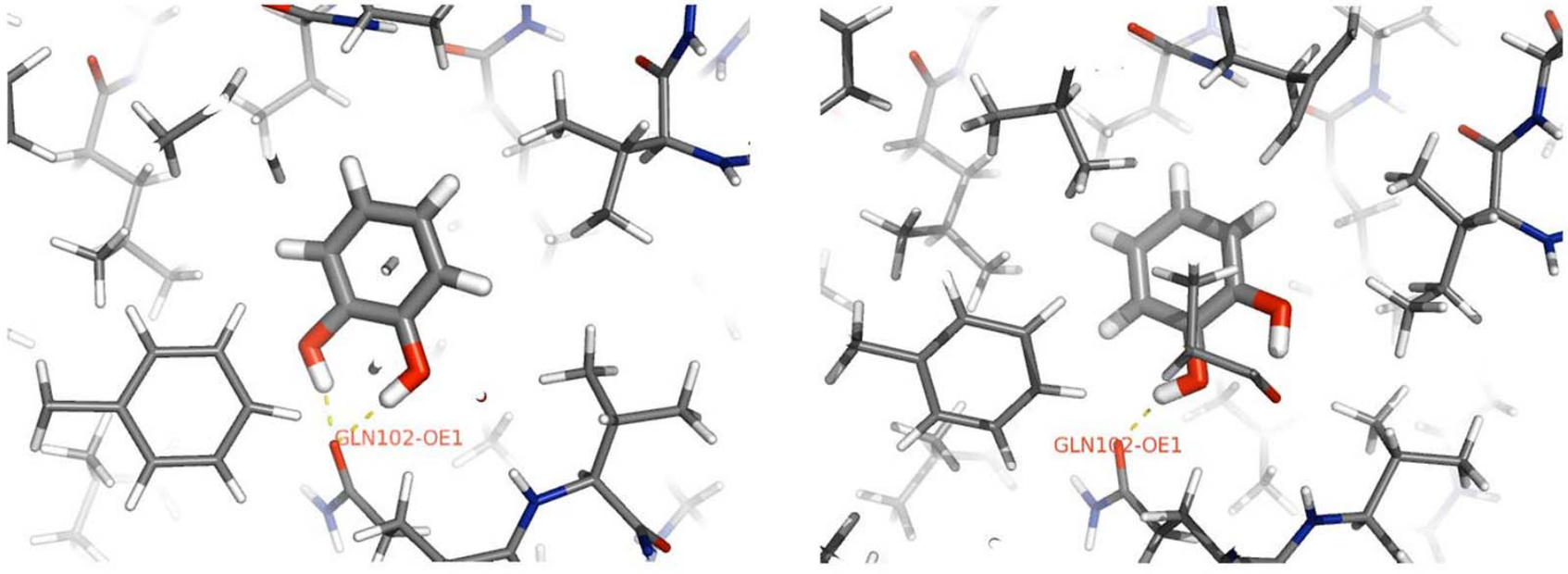




Can we quantitatively predict binding affinity of known ligands?

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

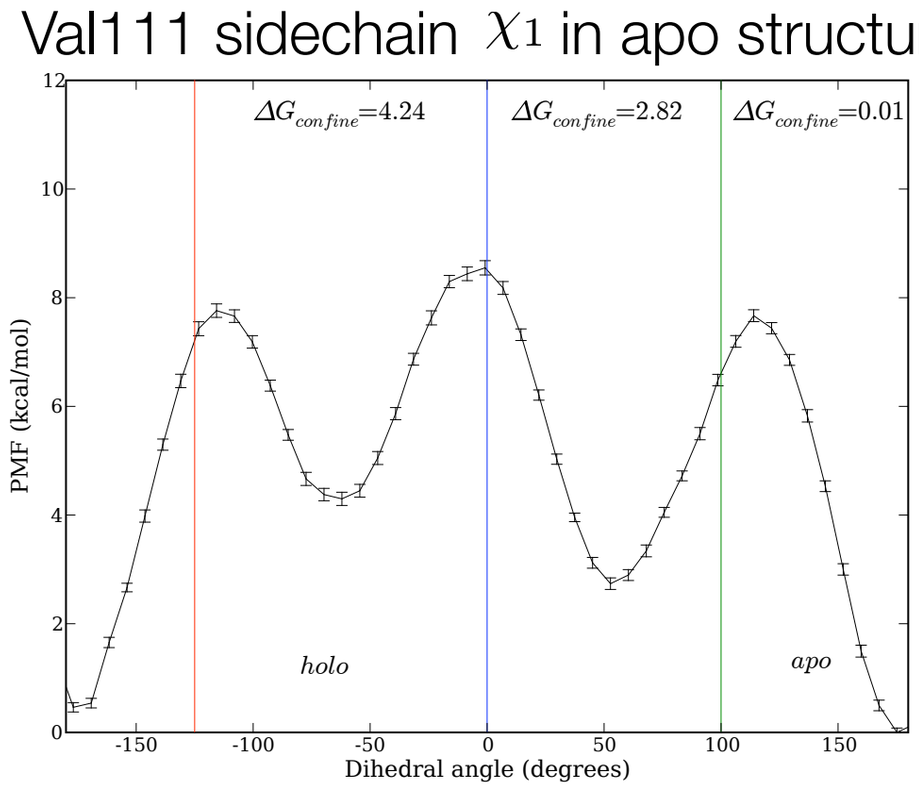
Multiple protein and ligand conformations and conformational changes can contribute to binding

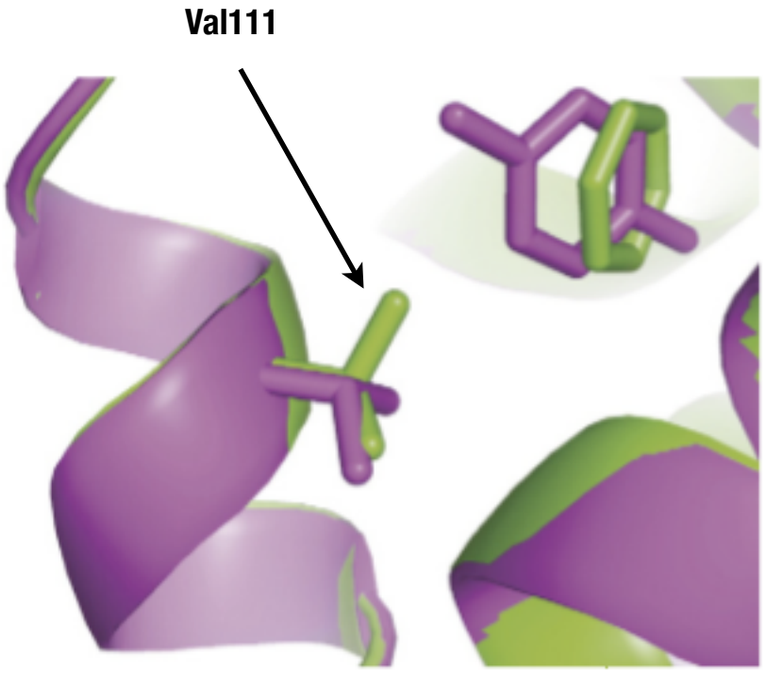


Difference in affinity between different bound orientations is only ~1 kT

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

Multiple protein and ligand conformations and conformational changes can contribute to binding

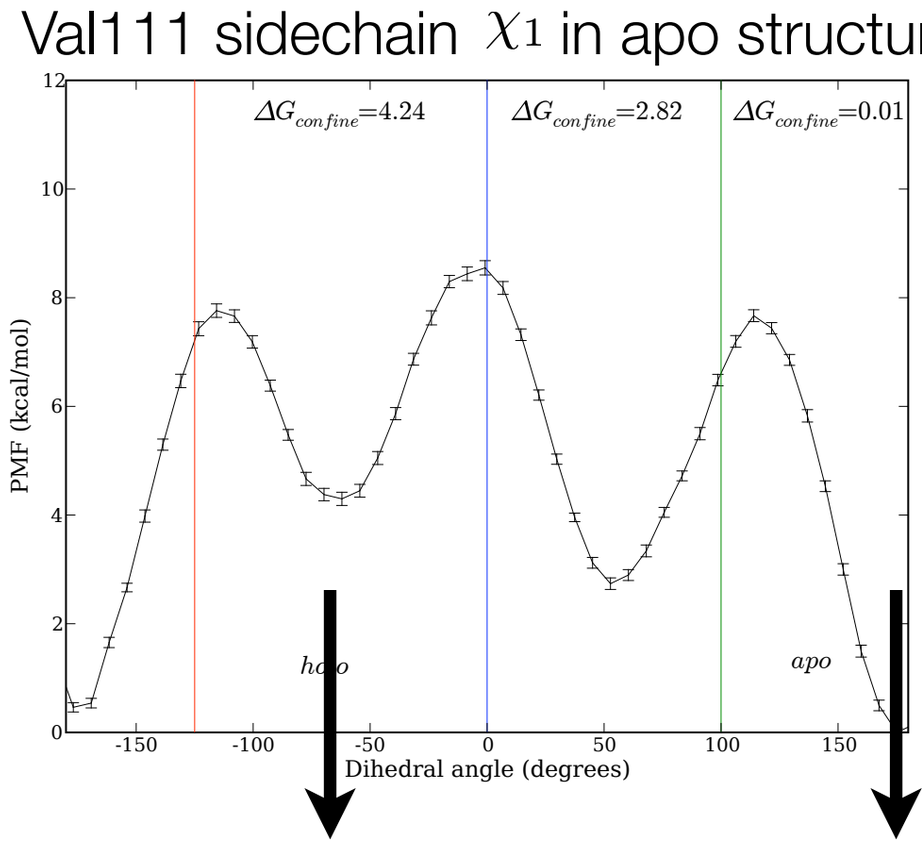


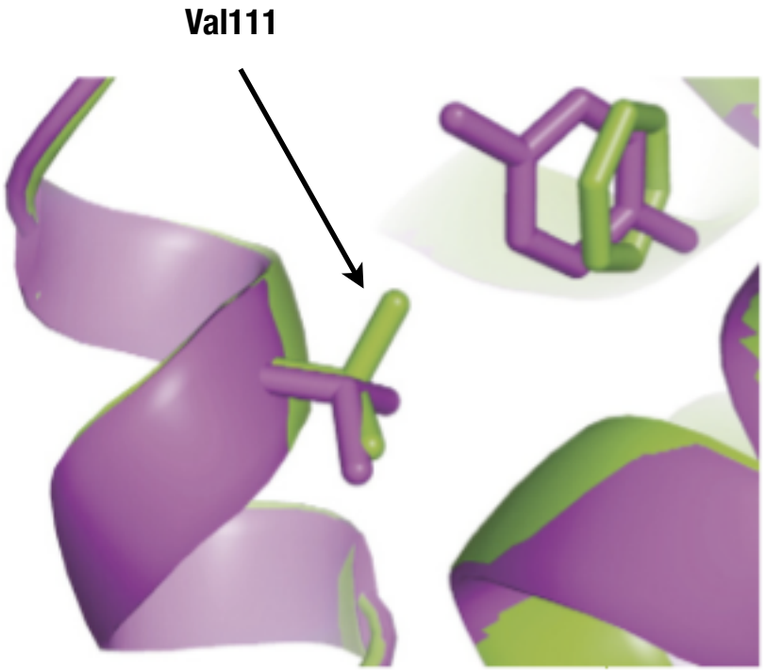


p-xylene vs. benzene bound to T4 lysozyme L99A

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

Multiple protein and ligand conformations and conformational changes can contribute to binding



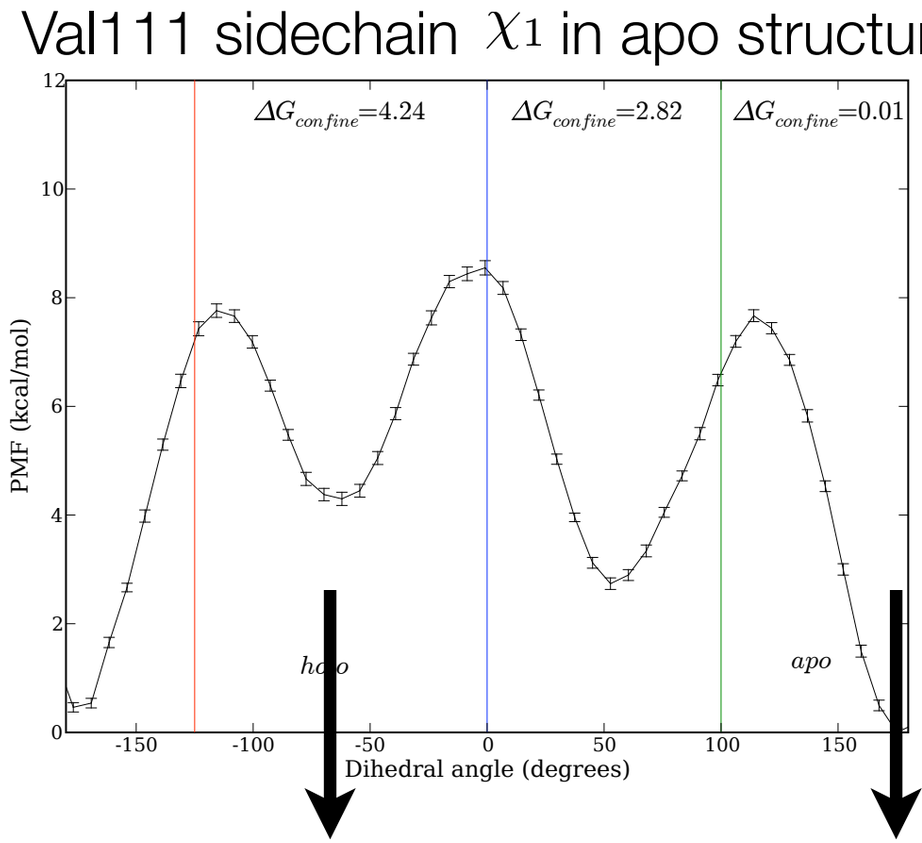


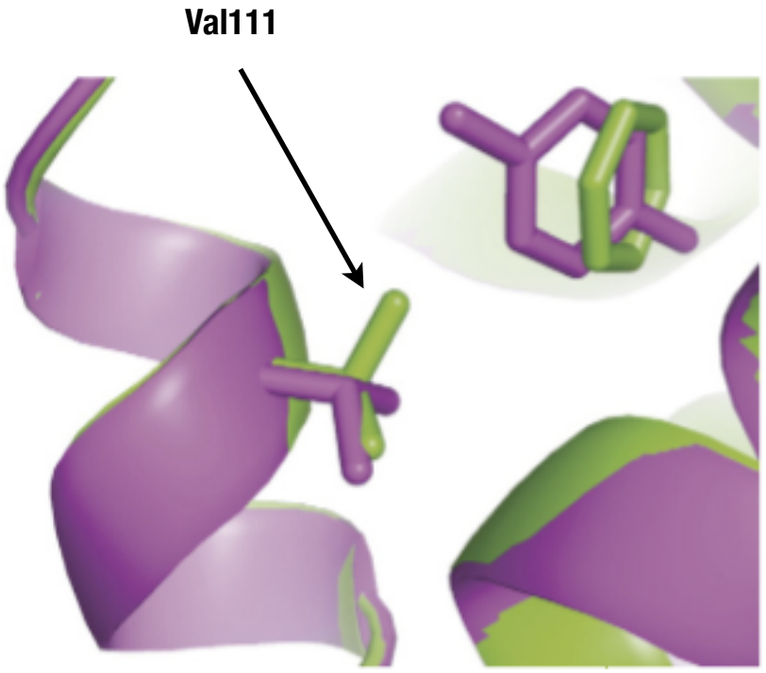
p-xylene vs. benzene bound to T4 lysozyme L99A

binding free energy -7.3 kcal/mol -3.0 kcal/mol

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

Multiple protein and ligand conformations and conformational changes can contribute to binding





p-xylene vs. benzene bound to T4 lysozyme L99A

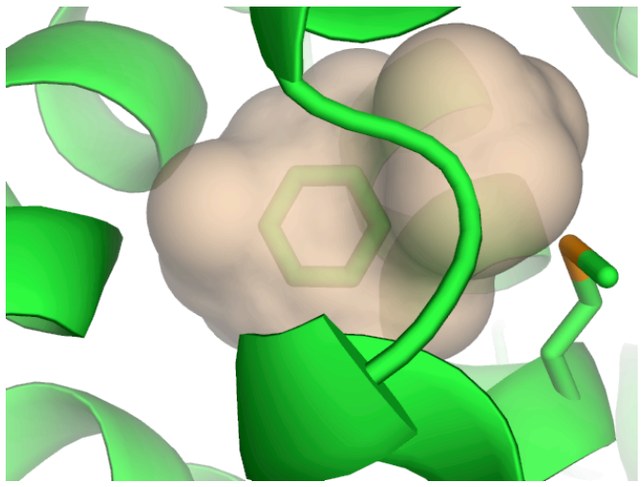
binding free energy	-7.3 kcal/mol		-3.0 kcal/mol
confinement free energy	4.2 kcal/mol		0.0 kcal/mol
net	-3.1 kcal/mol		-3.0 kcal/mol
release	-0.3 kcal/mol		-0.6 kcal/mol
total binding free energy	-3.4+-0.3	≈	-3.6+-0.3

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

How accurately can we reproduce known binding affinities?

Retrospective test on known, neutral small-molecule ligands of T4 lysozyme L99A

only *apo* structure was used



model binding site in T4 lysozyme L99A

Ligand	$\Delta G^o_{expt.}$ kcal/mol	$\Delta G^o_{calc.}$ kcal/mol	$\Delta G^o_{calc.} - \Delta G^o_{expt.}$ kcal/mol
2,3-benzofuran	-5.46 ± 0.03	-3.53 ± 0.06	1.93 ± 0.07
benzene	-5.19 ± 0.16	-4.56 ± 0.20	0.63 ± 0.26
ethylbenzene	-5.76 ± 0.07	-6.36 ± 0.18	-0.60 ± 0.19
indene	-5.13 ± 0.01	-1.75 ± 0.07	3.38 ± 0.07
indole	-4.89 ± 0.06	-0.42 ± 0.08	4.47 ± 0.10
isobutylbenzene	-6.51 ± 0.06	-5.01 ± 0.20	1.50 ± 0.21
n-butylbenzene	-6.70 ± 0.02	-4.87 ± 0.14	1.83 ± 0.14
n-propylbenzene	-6.55 ± 0.02	-5.88 ± 0.11	0.67 ± 0.12
o-xylene	-4.60 ± 0.06	-1.27 ± 0.18	3.33 ± 0.19
p-xylene	-4.67 ± 0.06	-3.54 ± 0.17	1.13 ± 0.18
toluene	-5.52 ± 0.06	-4.58 ± 0.12	0.94 ± 0.14
phenol	> -2.74	-1.26 ± 0.09	N/A
2-fluorobenzaldehyde	> -2.74	-2.92 ± 0.14	N/A
RMS error:		2.24	
Correlation, R:		0.72	

Mobley, Graves, Chodera, McReynolds, Shoichet, and Dill. JMB 371:1118, 2007.

How accurately can we predict unknown binding affinities?

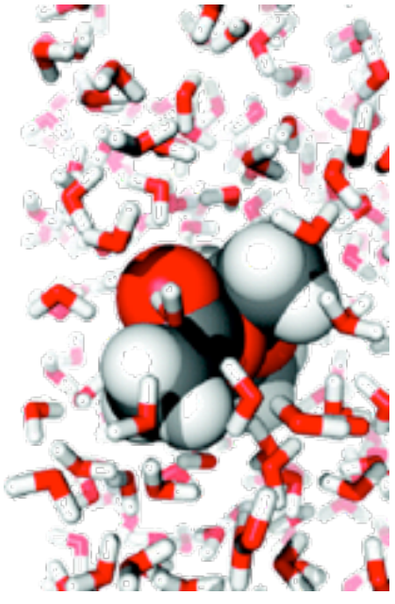
Ligand	DOCK Score (kcal/mol)	Prediction[1]	ΔG^o_{calc}[2] (kcal/mol)	ΔT_m (^{o}C)	Experiment	$\Delta G^o_{expt.}$ (kcal/mol)
1,2-dichlorobenzene	-19.99	Binder	-5.66 ± 0.15	2.90	Binder	-6.37
n-methylaniline	-17.29	Binder	-5.37 ± 0.11	1.00	Binder	-4.70
1-methylpyrrole	-15.27	Binder	-4.32 ± 0.08	2.20	Binder	-4.44
1,2-benzenedithiol	-18.51	Binder	-2.79 ± 0.13	2.50	Binder	N.D.
thieno-[2,3-c]pyridine	-18.81	Nonbinder	-2.56 ± 0.07	-0.40	Nonbinder	N.D.

Table 4

prediction accuracy ~ 0.5 kcal/mol

Mobley, Graves, Chodera, McReynolds, Shoichet, and Dill. JMB 371:1118, 2007.

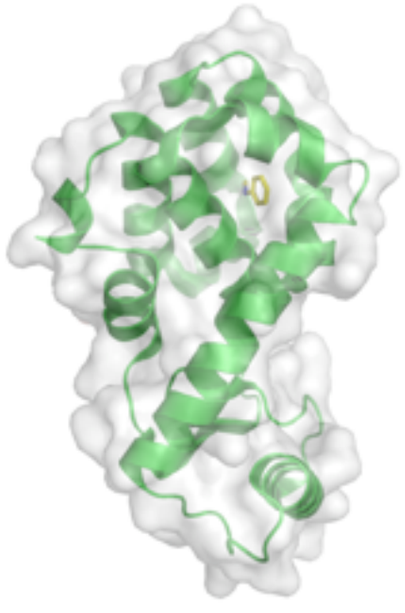
How far are we toward targets relevant to pharma?



hydration free energies of small neutral molecules

1.23±0.01 kcal/mol [502]
(Mobley et al., 2008)

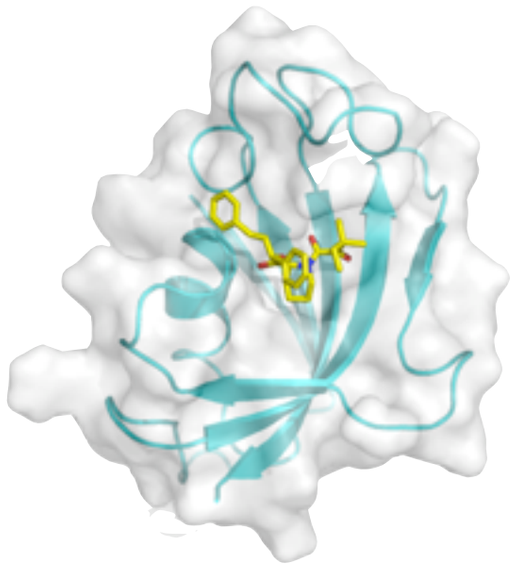
1.33±0.05 kcal/mol [17]
(Nicholls and Mobley et al., J Med Chem)



small apolar ligands T4 lysozyme L99A

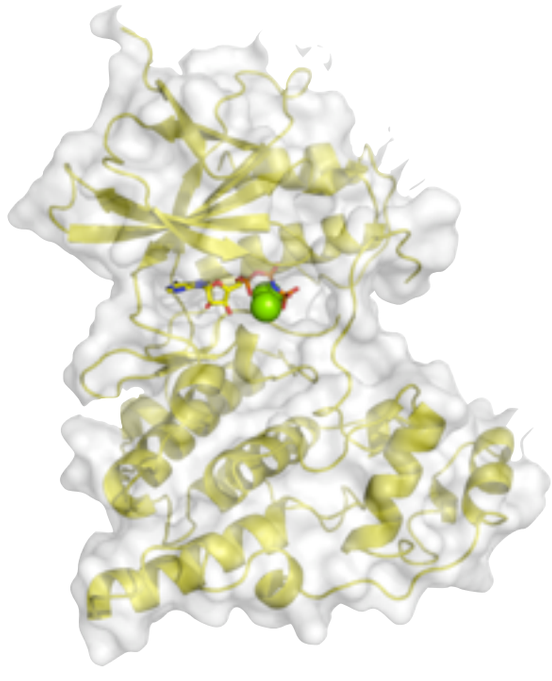
1.89±0.04 kcal/mol [13]
(Mobley and Graves et al., JMB 2007)

0.6±0.2 kcal/mol [3]
(Mobley and Graves et al., JMB 2007)



polar ligands FKBP12

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



JNK3 kinase

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

(not to scale)

Specific goals for YANK

Accelerate free energy calculations on commodity GPUs using OpenMM

The other extreme is Anton: an expensive, special-purpose machine for MD.

Test how accurate implicit solvent models are for ligand-binding free energy calculations

Largely unknown, only ~2 papers: Essex J Med Chem 49:7427, 2006; Roux Proteins 74:996, 2009.

Determine how critical statistical mechanics is for "rescoring" docked poses

Many rescoring methods use GB models, but all focus on energies of minimized conformations

Make free energy calculations both more realistic and more "idiot-proof"

Automatic sampling of protonation states, tautomers, salt concentrations

Provide a testbed for enhanced sampling algorithms that mix Monte Carlo and MD

Mix in efficient Monte Carlo moves; exploit Markov chain Monte Carlo algorithms

Build a tool for pharma and academia that fills a special "niche" between dock/rescore and FEP

Aim to replace unreliable, deficient methods like GB rescoring, MM-PBSA, MD

Alchemical free energy calculations have stricter requirements than “typical” MD applications

---

“My simulations is fine as long as gromacs didn’t segfault [too many times]” is not sufficient.

Free energies often exquisitely sensitive to details that seem “unimportant” in straightforward MD:
- **e.g. choice of PME parameters, long-range dispersion corrections, polarization convergence tol**

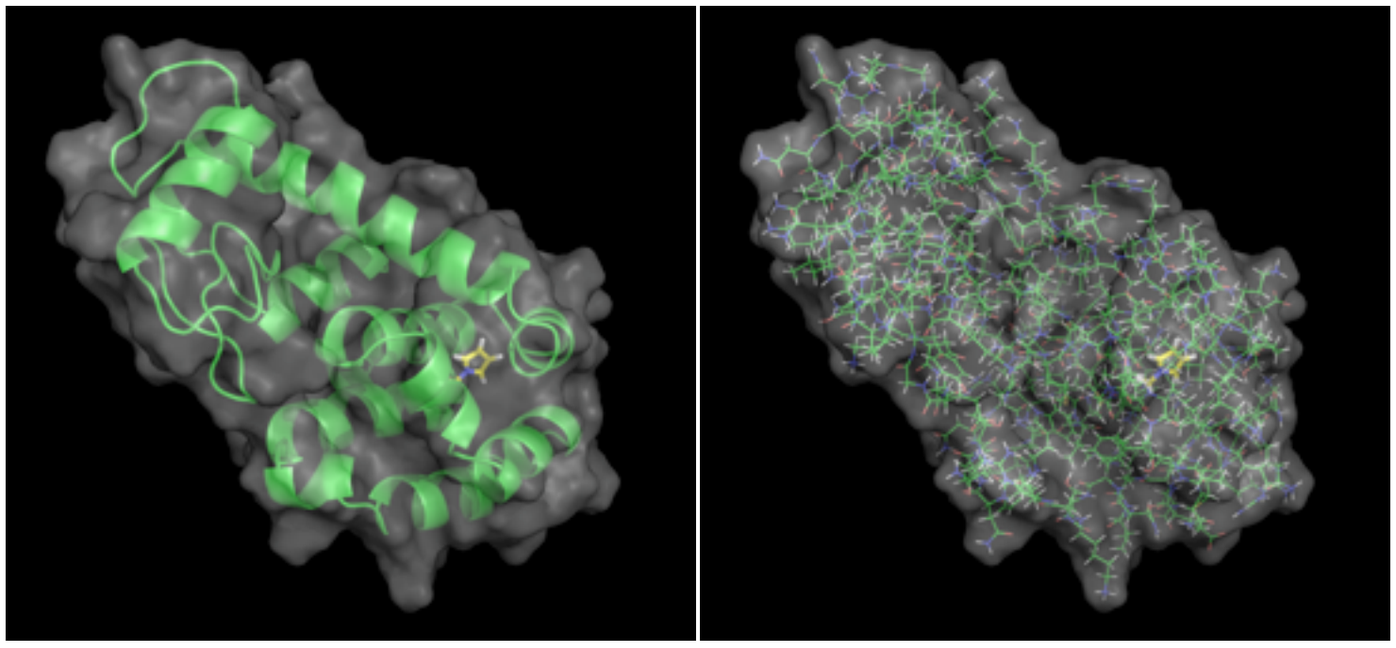
What we need from OpenMM:

Correct equilibrium must be sampled in order for relationships used in estimating free energy differences (e.g. BAR) to provide a meaningful estimate:
- **forces must be accurate derivatives of potential**
- **integrator and thermostats must generate correct ensemble**

Potential energy differences between alchemical states must be computed precisely
- **e.g. early versions of gromacs used single-precision accumulators for PME energy**

Milestone 1 : Feasibility
Does OpenMM provide sufficient speed and accuracy?

Implement AMBER crd/prmtop reader
Test energy conservation and timing for a “real system” where only change is GB model



T4 lysozyme L99A + 1-methylpyrrole [2616 atoms]
AMBER parm96 / GAFF + AM1-BCC charges
same system treated by Mobley et al.
except OBC GBSA instead of explicit solvent

also, test on more realistic systems:
- trypsin inhibitors [charged ligands]
- CDK2 kinase [real target]

Conditional pass

Erroneous force computation on NVIDIA GeForce 9600M GT and GeForce GT 120,
despite passing all CUDA and OpenMM tests. Only obvious with force calculations or constant-energy simulations; wouldn’t notice a problem with NVT simulations until your protein started to unfold!

NVIDIA GeForce GTX 285 gives correct forces and stable NVE dynamics with ~ 32 ns/day performance
Compare to Intel Core 2 Duo 2.4 GHz with single-core performance ~ 36 ps/day!

Milestone 2 : Accuracy of GBSA models
How much additional error is introduced by use of GBSA?

Repeat same free energy calculations as Mobley at al. for T4 lysozyme and other systems
Determine how much GBSA model degrades performance
Just need code fast enough for comparison to be feasible -- optimize later.

scale charges and GB intrinsic radii:

$$q_i \rightarrow (1 - \lambda_{elec}) \cdot q_i$$
$$\sigma_i^{GB} \rightarrow (1 - \lambda_{elec}) \cdot \sigma_i^{GB}$$

no soft-core support, so scale LJ radii and well depth:

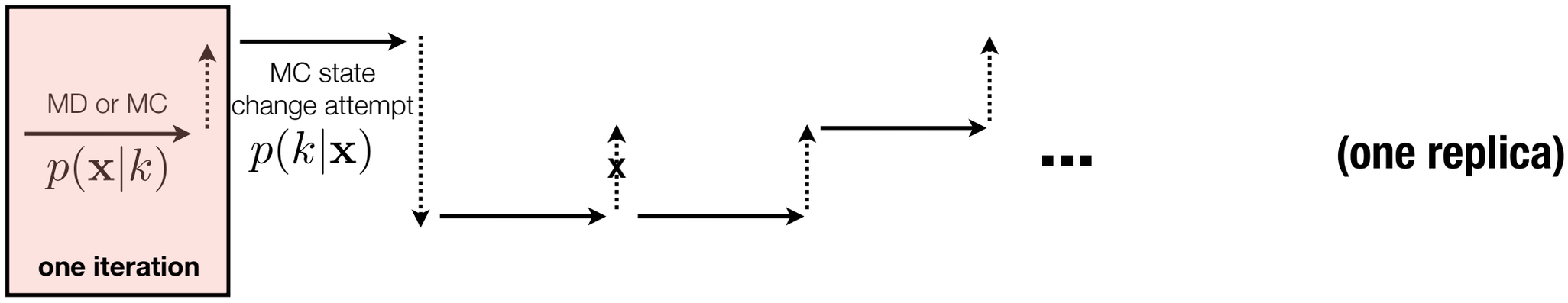
$$\sigma_i \rightarrow (1 - \lambda_{LJ}) \cdot \sigma_i$$

$$\epsilon_i \rightarrow (1 - \lambda_{LJ}) \cdot \epsilon_i$$

Basic algorithm for free energy calculations:

```
for each alchemical state k = 1:K
  for each iteration n = 1:N
    integrate T steps of dynamics with thermostat
    for each alchemical state l = 1:L
      compute potential energy of current configuration at state l and store it
reprocess all data with MBAR
```

Replica-exchange can reduce correlation times and provides a “built-in” convergence estimate



Original algorithm:

```
for each alchemical state k = 1:K
  for each iteration n = 1:N
    integrate T steps of dynamics with thermostat
    for each alchemical state l = 1:L
      compute potential energy of current configuration at state l and store it
reprocess all data with MBAR
```

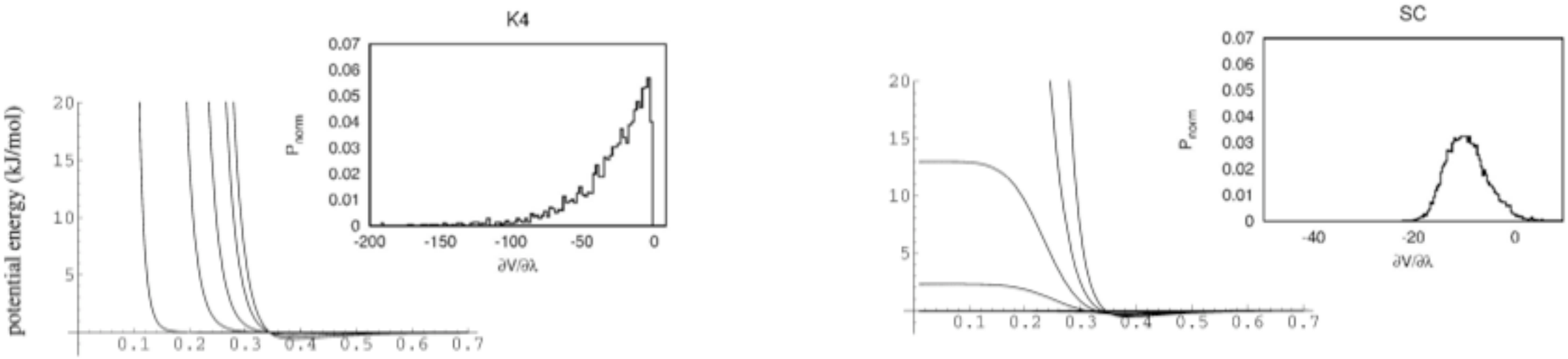
New algorithm:

```
for each iteration n = 1:N
  for each replica k = 1:K
    mix states among replicas (Gibbs sampling)
    integrate T steps of dynamics with thermostat
    for each alchemical state l = 1:L
      compute potential energy of current configuration at state l and store it
reprocess all data with MBAR
```

If each replica initialized from a different conformation (e.g. docked pose, receptor model), we have a “built-in” convergence check: Calculation is only converged after estimates of binding affinity from all replicas agree within statistical error.

Milestone 3 : Critical efficiency / usability improvements

Soft-core Lennard-Jones interactions critical to reducing variance in estimated free energies in explicit solvent calculations; likely important for implicit as well.



Steinbrecher T, Mobley DL, and Case DA. JCP 127:214108, 2007.
Pitera JW and van Gunsteren WF. Mol. Sim. 28:45, 2002.

Could potentially be implemented with lookup tables, which also allows “fused” intermediates [e.g. Beutler and van Gunsteren JCP 101:1417, 1994].

GPU-accelerated energy calculations to reduce bottleneck

Kai is near nearly finished with these

Automatic ligand parameterization / system setup scripts

Switch to Python codebase

Critical for programmer efficiency; sanity

Will require near-complete exposure of OpenMM API

Helping out Randy with pyOpenMM: https://simtk.org/home/pyopenmm

Milestone 4 : Treating realistic systems, making calculations “idiot-proof”, enhancing sampling

---

Constant-pH simulations

Easily implemented in GB via simple Monte Carlo moves where proton parameters are modified. [Mongan et al. J Comput Chem 25:2038, 2004.]

Debye-Hückel screening for monovalent counterions

Already incorporated into AMBER GB models [Srinivasan et al. Theor. Chem. Acc. 101:426, 1999]

Monte Carlo sampling of tautomers

If aqueous tautomer ratios are known or can be predicted, MC moves allow switches between

Semi-grand-canonical sampling of divalent (or all explicit) counterions?

Easily implemented in MCMC framework

Efficient torsion MC moves for sampling sidechain conformational changes

e.g. “smart-darting MC”, with move set of rotamer displacements [JCP 114:6994, 2001]

All of these methods simply are different MC moves that “plug in” to the same MCMC framework.

Milestone 5 and beyond

Integrated setup pipeline

There's no reason we can't incorporate the entire MODELLER + MCCE pipeline we use to build in missing residues and heavy atoms and select protonation states.

Markov state models

One way to overcome slow conformational dynamics of proteins: Divide-and-conquer
Related to concepts of "conformational selection"

Expanded-ensemble methods for sampling over huge libraries of compounds

Find the best binders in a combinatorially-constructed library

Enhanced sampling methods such as NML

Collaborators and funding

Stanford
Vijay Pande
Imran Haque

Vertex
Kim Branson
Pat Walters
Mark Murcko

Stanford SimBios
Peter Eastman
Mark Friedrichs

University of Virginia
Michael Shirts

University of New Orleans
David Mobley

Funding
California Institute for Quantitative Biosciences (QB3)
Distinguished Postdoctoral Fellowship

Code and data available at http://www.choderalab.org

The standard state for binding free energies

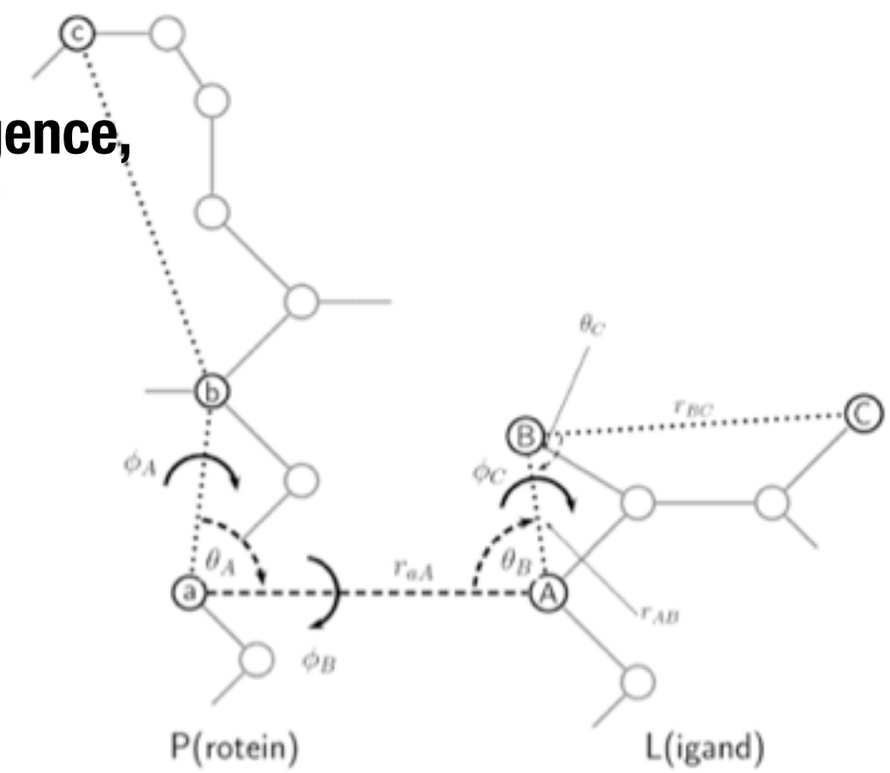
Association constant is not unitless:

$$K_a = \frac{[PL]}{[P][L]} = \frac{e^{-\beta \Delta G_a}}{(1\,\mathrm{M})}$$

Free energy of binding is defined with respect to a reference state
Standard reference state is 1M ligand (1 ligand / 1660 A^3)

Luckily, if you use Boresch's ligand restraints to aid convergence, he has already computed the free energy correction for you!

$$\Delta A_r^{\mathrm{VBA},0} = -kT \ln\left[\frac{8\pi^2 V^0}{r_{\mathrm{aA},0}^{\ 2} \sin\theta_{\mathrm{A},0} \sin\theta_{\mathrm{B},0}} \frac{(K_r K_{\theta_\mathrm{A}} K_{\theta_\mathrm{B}} K_{\phi_\mathrm{A}} K_{\phi_\mathrm{B}} K_{\phi_\mathrm{C}})^{1/2}}{(2\pi kT)^3}\right]$$



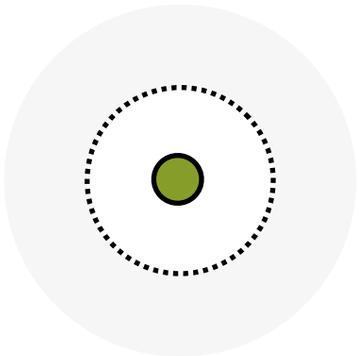
Absolute Binding Free Energies: A Quantitative Approach for Their Calculation
Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M.
J. Phys. Chem. B.; **(Article); 2003**; 107(35); 9535-9551. DOI: 10.1021/jp0217839

Anisotropic long-range dispersion correction

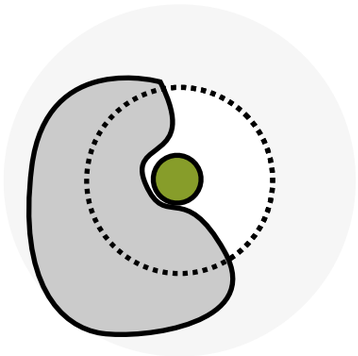
Simulations in solvent must be run with long-range dispersion correction to ensure results are not sensitive to choice of cutoff.

```
; Apply long range dispersion corrections for Energy and Pressure =
DispCorr                 = AllEnerPres
```

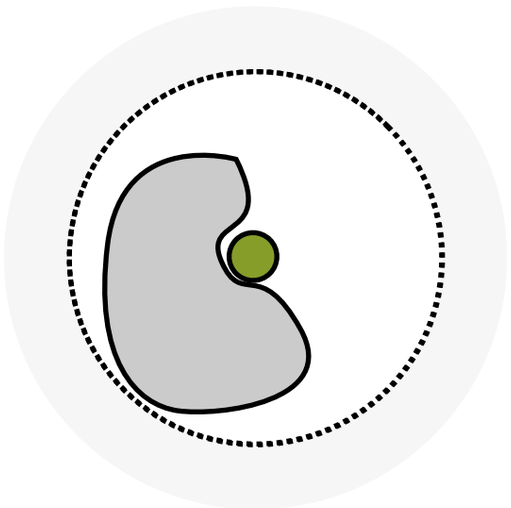
This correction assumes isotropic distribution of Lennard-Jones sites throughout system.



isotropic assumption holds



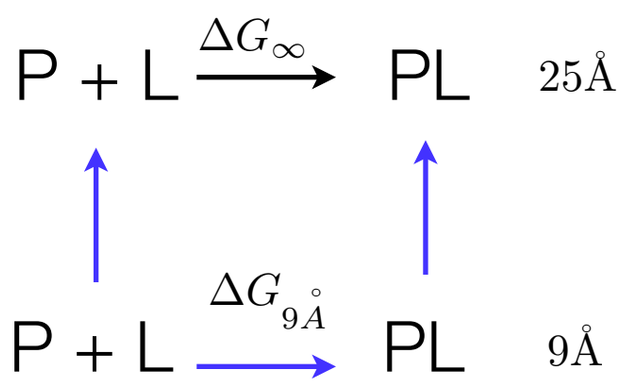
isotropic assumption fails



isotropic assumption holds

Instead, we have to enlarge cutoff so that isotropic assumption holds

An explicit postprocessing step recomputes energies with large cutoff and estimates perturbation free energies using EXP.



Can make a difference of 3 kcal/mol, depending on number of ligand atoms

M. R. Shirts*, **D. L. Mobley***, J. D. Chodera, and V. S. Pande. "Accurate and efficient corrections for missing dispersion interactions in molecular simulations", *J. Phys. Chem. B* **111**:13052-13063 (2007).