Alchemical free energy calculations in OpenMM

Lee-Ping Wang
Stanford Department of Chemistry
OpenMM Workshop, Stanford University
September 7, 2012

Special thanks to: John Chodera, Morgan Lawrenz

Outline

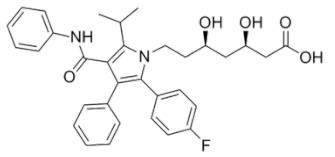
- **Introduction to free energy calculations**
  - Applications: Solvation and binding free energies
  - Methods: Reaction coordinate vs. alchemical methods
  - Accuracy and benchmarking
- Part I: Running the alchemical simulations
  - Alchemical intermediates
  - Soft-core nonbonded force
  - Restraints
- Part II: Analyzing the data
  - Multistate Bennett acceptance ratio
- Part III: Hydration free energy exercise

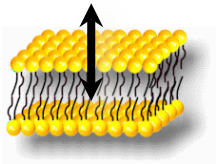
Introduction to free energy calculations

Free energy differences are central to many questions in biological and pharmaceutical chemistry.

How does compound X partition between different environments?

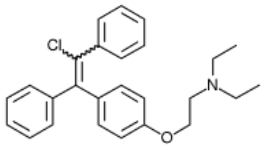
e.g. How well does Lipitor partition between octanol and water?

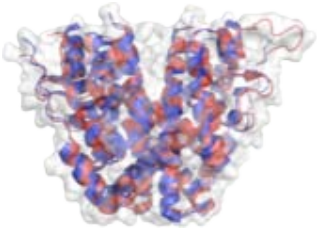




How tightly does compound X bind protein Y? How selectively?

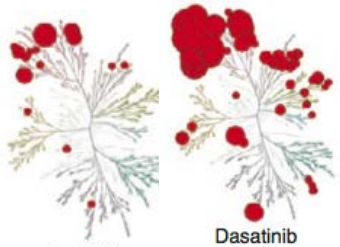
e.g. How well does clomifene bind/discriminate ERα/ERβ?





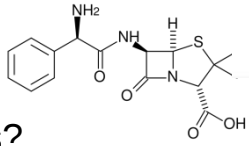
How do modifications in compound X modulate affinity and selectivity?

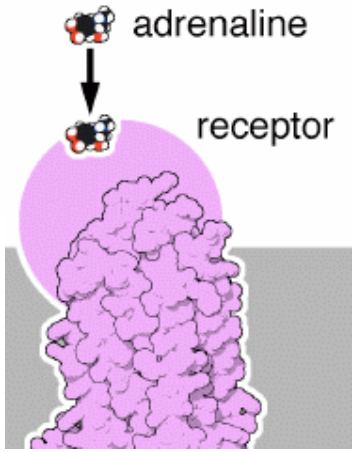
e.g. Why does selectivity of imatinib differ from that of other kinase inhibitors?



How do target mutations affect binding affinities?

e.g. How do clinically isolated β-lactamase resistance mutations restore ability to hydrolyze β-lactam antibiotics?

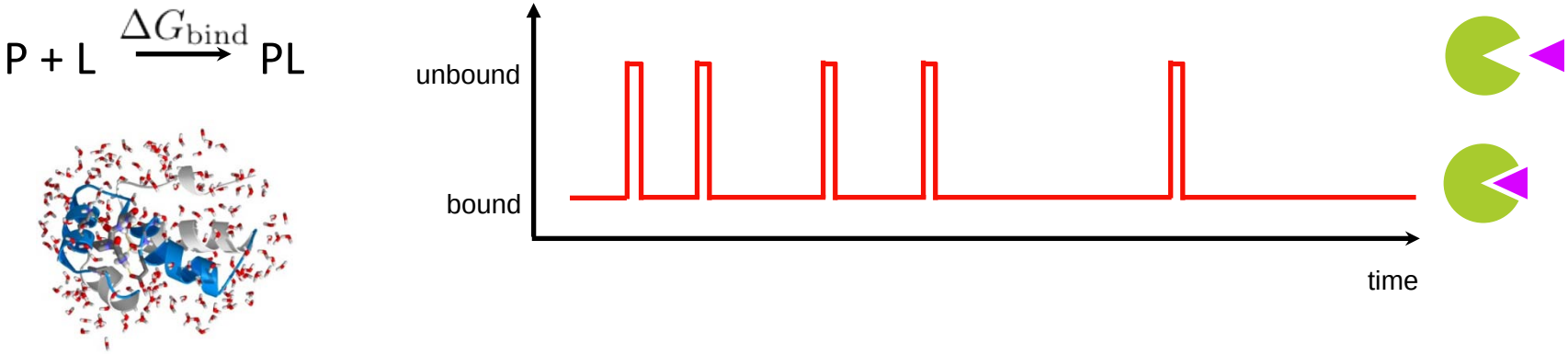




Calculation of free energies

Without specialized simulation methods,
free energy calculations would be impractically slow.

Observing binding events from a normal MD simulation:



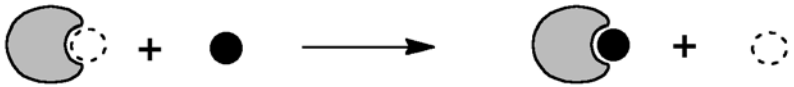
For typical drug off-rates (10^{-4} s^{-1}),
trajectories would need to be impractically long (hours),
requiring ~10^{9} CPU-years to simulate.

Calculation of free energies

Free energy differences are calculated efficiently by simulating a path between reactant and product.

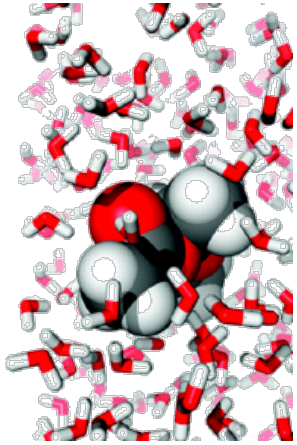
- Potential of Mean Force: Track the free energy along a specified reaction coordinate

- Alchemical: Simulate along an unphysical pathway to get free energy differences between two physical endpoints



Accuracy of free energy calculations

The accuracy of a free energy calculation depends on the complexity of the problem.

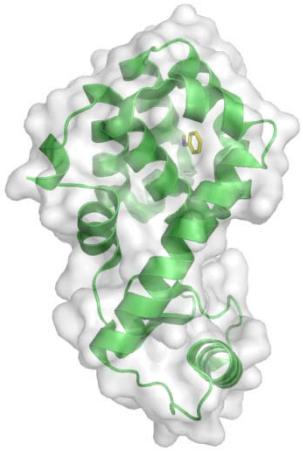


hydration free energies
of small neutral molecules

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

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

solvent only
small, neutral molecules
fixed protonation states

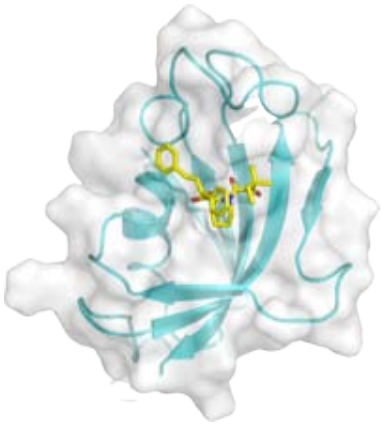


small apolar ligands
T4 lysozyme L99A

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)

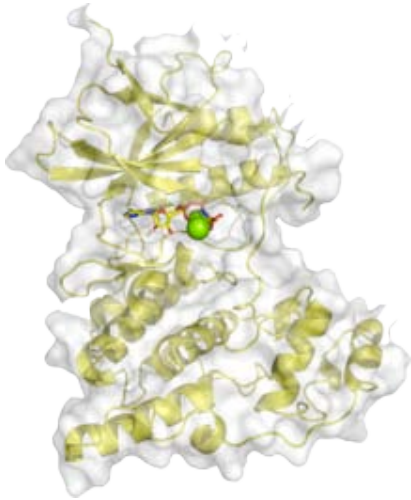
small, rigid protein
small, neutral ligands
fixed protonation states
multiple sidechain orientations
multiple ligand binding modes



polar ligands
FKBP12

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

small, rigid protein
fixed protonation states
larger drug-like ligands
rotatable bond in ligands



JNK3 kinase

6.3 kcal/mol [44]
(Haque, Chodera, Shirts, Mobley, Pande)

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

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

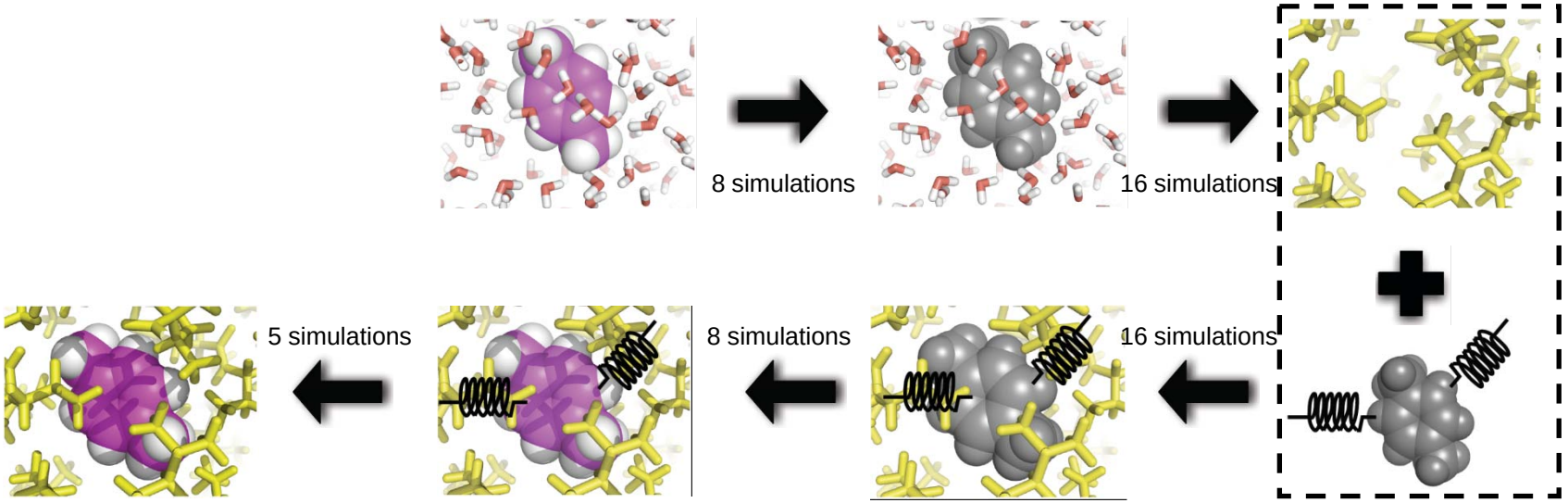
easy
hard

Outline

- Introduction to free energy calculations
  - Applications: Solvation and binding free energies
  - Methods: Reaction coordinate vs. alchemical methods
  - Accuracy and benchmarking
- **Part I: Running the alchemical simulations**
  - Alchemical intermediates
  - Soft-core nonbonded force
  - Restraints
- Part II: Analyzing the data
  - Multistate Bennett acceptance ratio
- Part III: Hydration free energy exercise

Alchemical free energy calculations

Alchemical simulations progress through a series of intermediates between initial and final states.

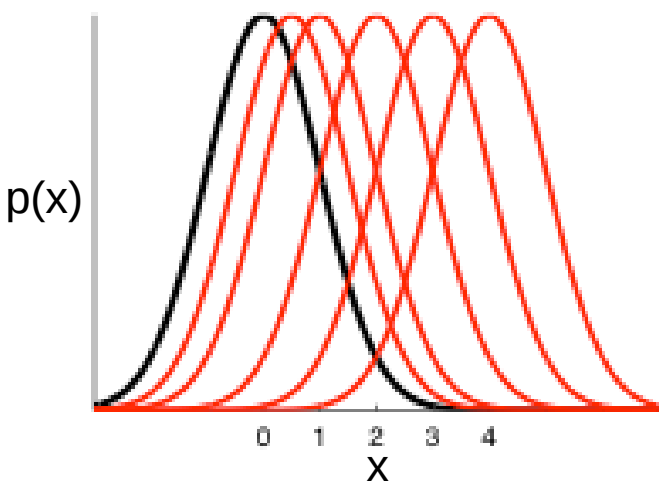


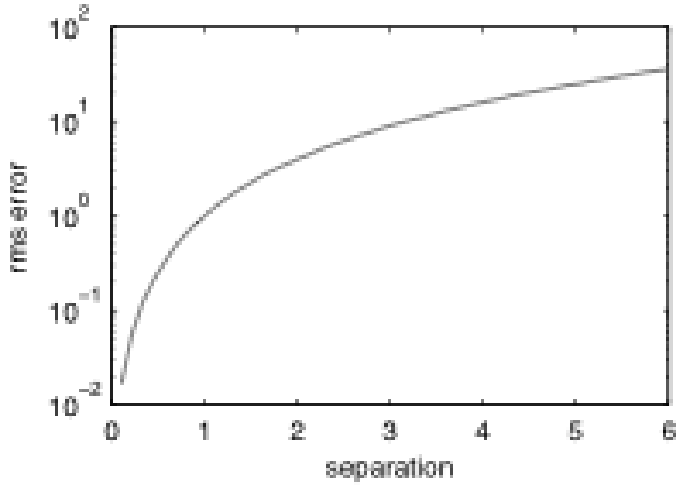
1. Atomic charges on ligand are switched off
2. VdW interaction between ligand and water is switched off
   Harmonic restraints are switched on
3. VdW interaction between ligand and protein is switched on
4. Atomic charges on ligand are switched on
5. Harmonic restraints are removed

Graphics from David Mobley

Alchemical free energy calculations

The accuracy of alchemical simulations relies on the phase space overlap of neighboring intermediates.





Intermediate states are introduced to ensure a contiguous chain of good overlap.

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

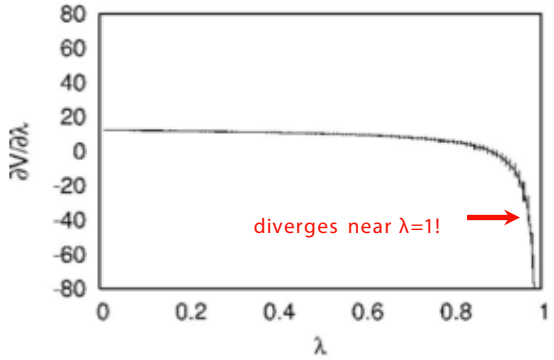
Soft-core Lennard Jones potentials

The Lennard-Jones repulsion is switched off gently, greatly improving the overlap between intermediates.

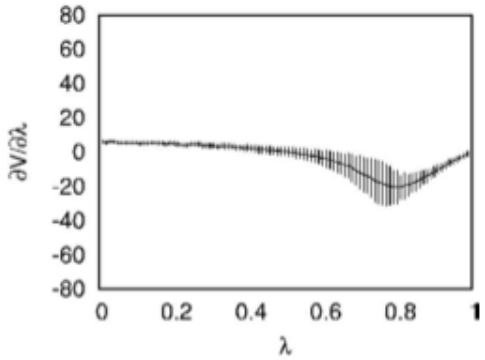
Example: Using thermodynamic integration, the free energy difference is the area under the curve. Lennard-Jones is off at λ=1.

$$\Delta F = \int_0^1 d\lambda' \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\lambda'}$$

Standard (linear) decoupling



Soft core decoupling



$$U(r;\lambda) = 4\epsilon(1-\lambda)\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

$$U(r;\lambda) = 4\epsilon(1-\lambda)\left[\frac{1}{[\alpha\lambda + (r/\sigma)^6]^2} - \frac{1}{[\alpha\lambda + (r/\sigma)^6]}\right]$$

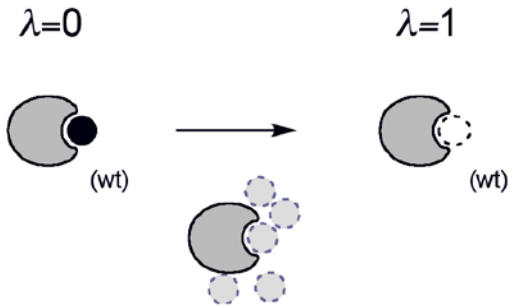
T. A. Steinbrecher, D. L. Mobley, and D. A. Case. JCP 127:214108 (2007)

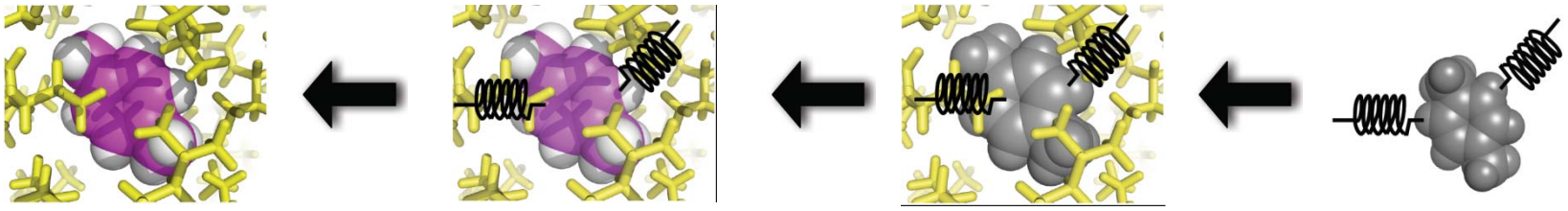
(More efficient forms known, but all based on same idea.)

Restraint potentials

Restraint potentials eliminate convergence issues related to translational / rotational motion.

Alchemical intermediates with weakened interactions may encounter convergence issues from the ligand diffusing out of the binding pocket or otherwise "moving around".



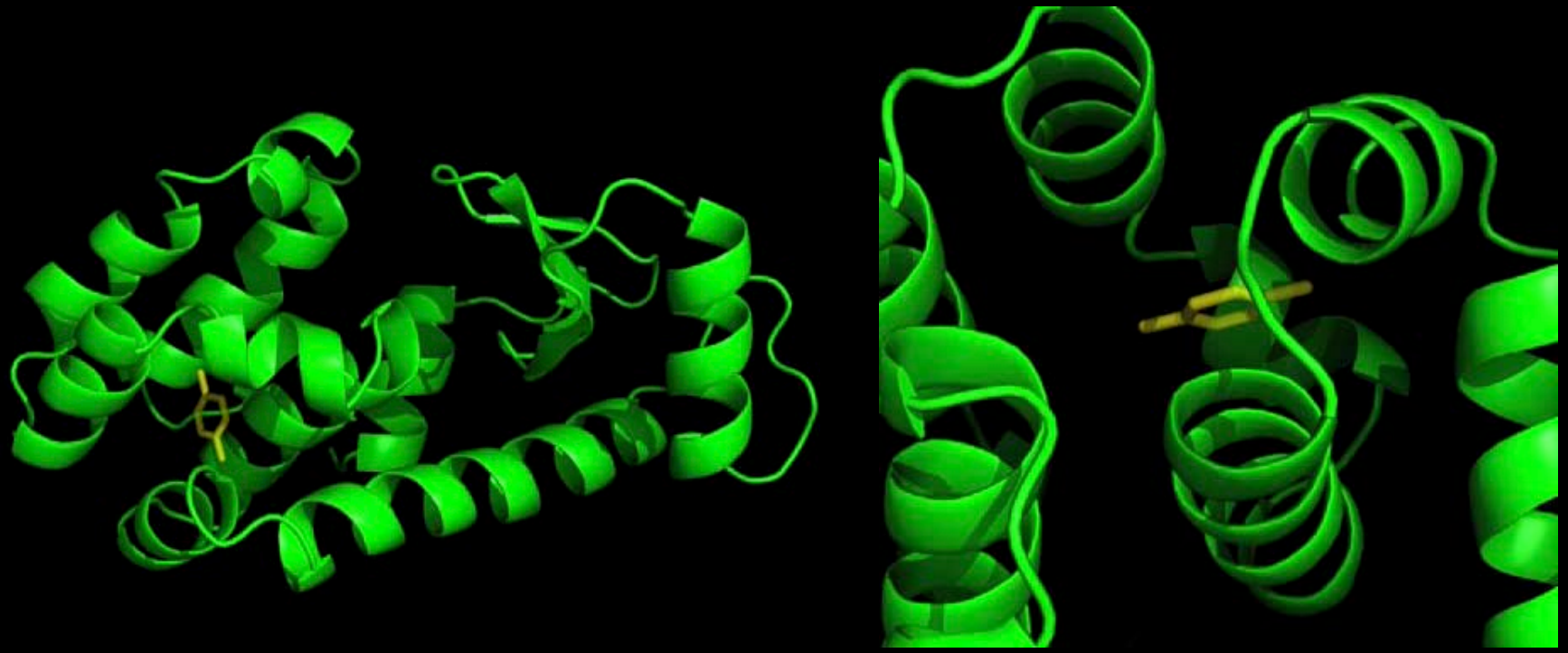


Harmonic restraint potentials keep the ligand in place;
the free energy of restrained degrees of freedom are added back in analytically.

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

Demonstration of alchemical intermediates

Using enhanced sampling methods, multiple binding events and reorientation of the ligand are observed.



[closeup]

p-xylene bound to T4 lysozyme L99A

Outline

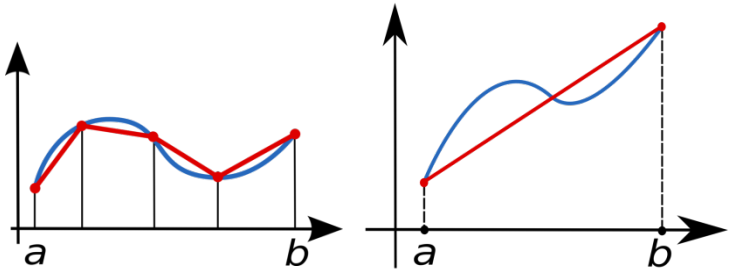
- Introduction to free energy calculations
  - Applications: Solvation and binding free energies
  - Methods: Reaction coordinate vs. alchemical methods
  - Accuracy and benchmarking
- Part I: Running the alchemical simulations
  - Alchemical intermediates
  - Soft-core nonbonded force
  - Restraints
- **Part II: Analyzing the data**
  - Multistate Bennett acceptance ratio
- Part III: Hydration free energy exercise

Estimating free energy differences from data

Not all estimators are created equal (in terms of bias and statistical efficiency).

Thermodynamic integration

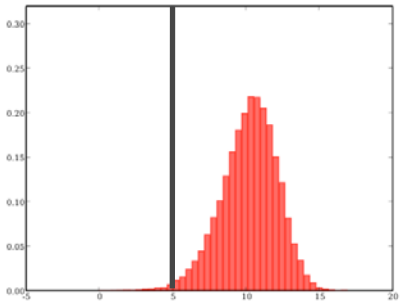
$$\Delta F = \int_{\lambda_1}^{\lambda_2} d\lambda' \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda'} \approx \frac{\Delta\lambda}{2} \left[\left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda_1} + \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda_2}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



quadrature error (bias) difficult to quantify

Free energy perturbation / exponential reweighting

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

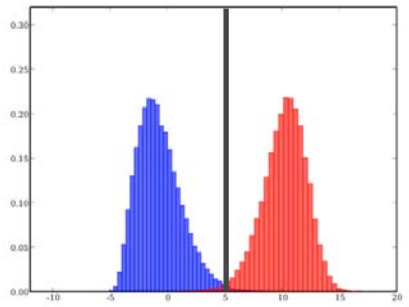


Zwanzig RW. JCP 22:1420, 1954.
Shirts MR and Pande VS. JCP 122:144107, 2005.

suffers from large bias and variance

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.

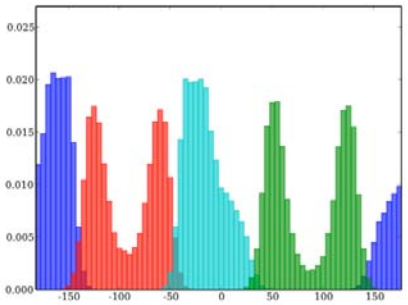
only applicable to two states - can't we use all the data?

Multistate Bennett acceptance ratio (MBAR)

The MBAR method provides statistically optimal estimates of free energy differences.

Statistically optimal analysis of samples from multiple equilibrium states

Michael R. Shirts (Department of Chemistry, Columbia University), John D. Chodera (Department of Chemistry, Stanford University)
http://arxiv.org/abs/0801.1426



$$\hat{f}_i = -\ln \sum_{j=1}^{K} \sum_{n=1}^{N_j} \frac{\exp[-u_i(\boldsymbol{x}_{jn})]}{\sum_{k=1}^{K} N_k \exp[\hat{f}_k - u_k(\boldsymbol{x}_{jn})]}$$

$$\delta^2 \Delta \hat{f}_{ij} = \hat{\Theta}_{ii} - 2\hat{\Theta}_{ij} + \hat{\Theta}_{jj}$$

$$\hat{\boldsymbol{\Theta}} = \mathbf{W}^{\mathrm{T}}(\mathbf{I}_N - \mathbf{W}\mathbf{N}\mathbf{W}^{\mathrm{T}})^{+}\mathbf{W}$$

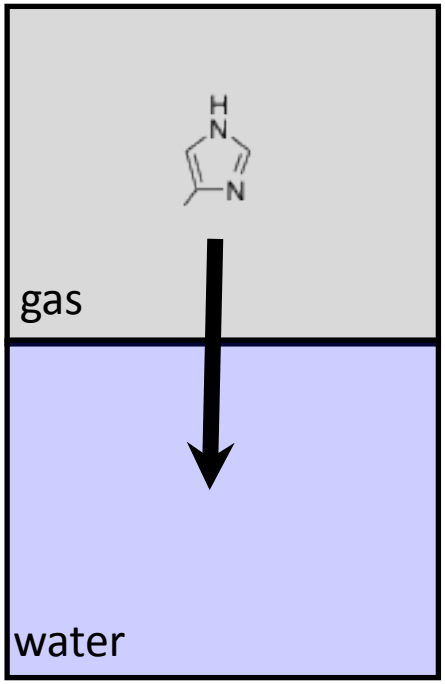
http://simtk.org/home/pymbar

Outline

- Introduction to free energy calculations
  - Applications: Solvation and binding free energies
  - Methods: Reaction coordinate vs. alchemical methods
  - Accuracy and benchmarking
- Part I: Running the alchemical simulations
  - Alchemical intermediates
  - Soft-core nonbonded force
  - Restraints
- Part II: Analyzing the data
  - Multistate Bennett acceptance ratio
- **Part III: Hydration free energy exercise**

Hydration free energy exercise

Hydration free energies are relatively simple to compute compared to binding free energies.



Benefits:

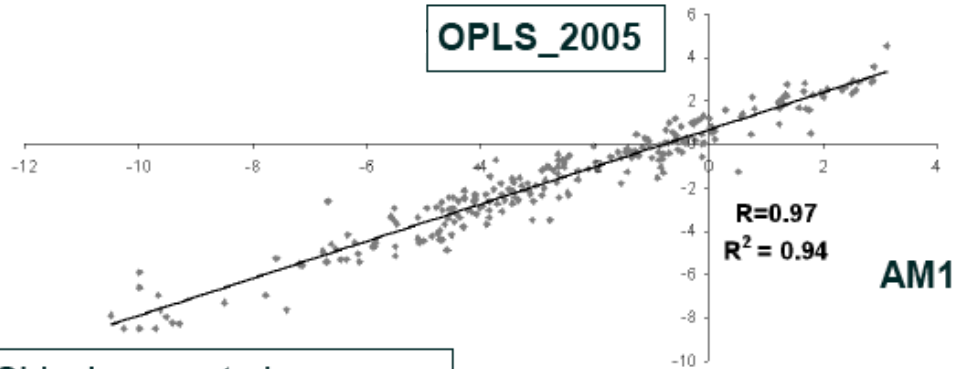
- Test the accuracy of half of binding reaction (withdrawal from water)
- A simple, tractable system for studying accuracy of calculations

Note: Run `simulate-water.py` now, it will take a few minutes…

D. L. Mobley, E. Dumont, J. D. Chodera and K. A. Dill. J. Phys. Chem. B. 111:2242-2254 (2007) (DOI)

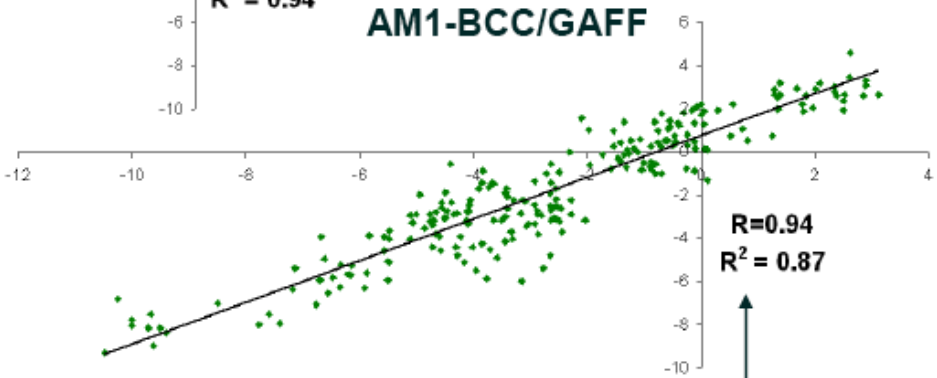
Some hydration free energy results

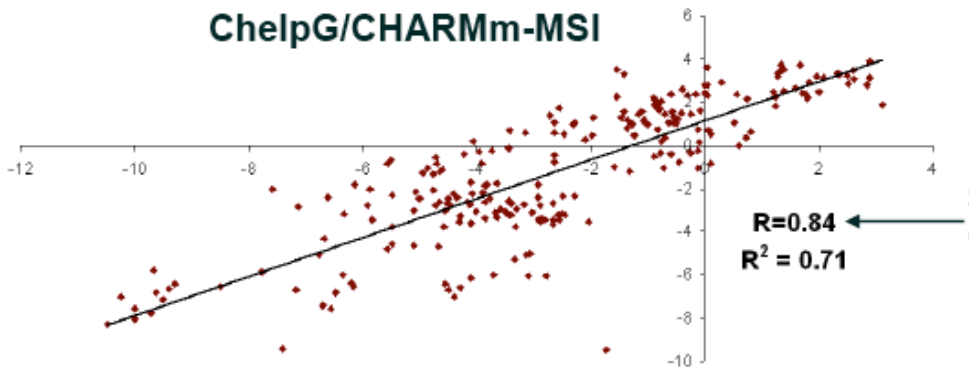
Hydration free energies are often used for force field benchmarking.



Correlation of Experimental (x-axis) and Calculated (y-axis) absolute solvation free energies.

Shivakumar et al., J. Chem. Theory Comput. 6, 1509, **2010**



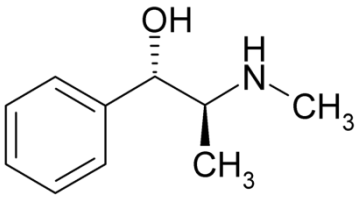


Shivakumar D, Deng Y, Roux B; J. Chem. Theory Comput. 5; 919; 2009

Slide by: David Rinaldo, Applications Scientist at Schrodinger

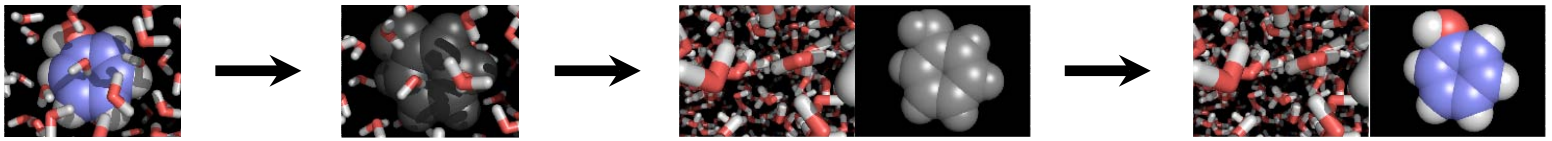
SCHRÖDINGER.

Choice of intermediates



	annihilation (remove interactions within solute)	decoupling (keep interactions within solute)
electrostatics	potentially large ΔG short correlation times requires vacuum recharging calculation	smaller ΔG longer correlation times due to bare charges no vacuum recharging calculation required 3 PME evaluations per timestep, tricky to use
Lennard-Jones	requires simulation in vacuum	no vacuum simulation required eliminates potentially unphysical conformations

Note that the literature is full of extremely confusing terminology, like “double decoupling”, which is actually annihilation...



“annihilation” vs “decoupling” discussed in:

M. R. Shirts, D. L. Mobley, and J. D. Chodera. "Alchemical free energy calculations: Ready for prime time?",Annual Reports in Computational Chemistry 3:41-59 (2007).

Example calculation

Perform a simple solvation free energy simulation.

Example calculation 1: Hydration free energy of a water molecule

- Run `simulate-water.py` . What the script does:
  1. Set up the OpenMM simulation, minimize the energy
  2. Create several alchemical intermediates
  3. Run a simulation for each alchemical intermediate
  4. Post-process the simulation trajectories, evaluating the energy of each alchemical state over each trajectory
  5. Save the data to disk

- Run `analyze.py --ncfile output.nc`
  1. Run MBAR analysis on output data.
  2. Get hydration free energy!

```
#########################################################################
### At 298.00 K, the solvation free energy is: -5.594 +- 0.514 kcal/mol ###
#########################################################################
```

Example calculation

How about a more interesting molecule?

Example calculation 2: Hydration free energy of ethanol

- Copy `simulate-water.py` to `ethanol.py`
  1. Change the number of alchemical atoms.
  2. Load the ethanol force field and PDB files.
  3. Use the `Modeller` class to create a solvated box (or use provided solvated box).

- Run `analyze.py --ncfile output.nc`
  1. Run MBAR analysis on output data.
  2. Get hydration free energy!
- Finished? Try these:
  1. Experiment with the number of alchemical intermediates, solvent box size and the simulation length. How to optimize accuracy vs. efficiency?
  2. Set up your own molecule of interest using AMBER tools.