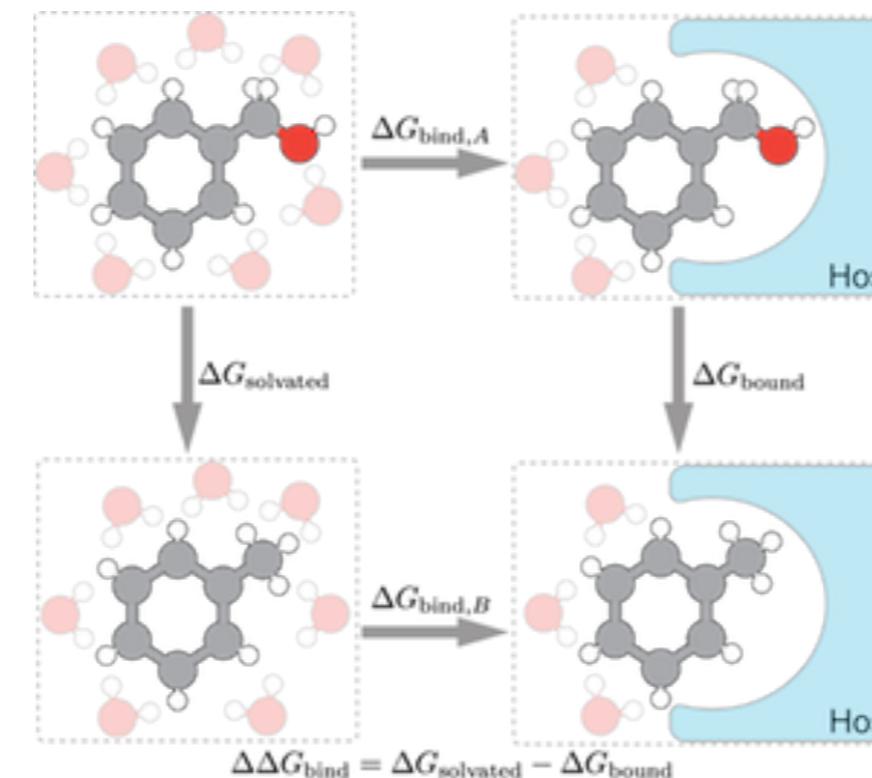




THE UNIVERSITY
of EDINBURGH

Alchemical free energy simulations with BioSimSpace



Antonia Mey – antonia.mey@ed.ac.uk

School of Chemistry
University of Edinburgh, UK

May 2019 – University of Bristol

The plan

- 14:00 - 14:30 Introduction to Free Energy energy simulations
- 14:30 - 15:00 Exercise 01: Setting up an alchemical simulations
- 15:00 - 15:15 Coffee
- 15:15 - 15:30 Free energy analysis
- 15:30 - 16:00 Exercise 02: Free energy analysis
- 16:00 - 17:00 Exercise 03: Advanced free energy analysis

BioSimSpace server

notebook.biosimspace.org

Update BioSimSpace



Logout Control Panel

Files Running Clusters

Select items to perform actions on them.

The screenshot shows a Jupyter Notebook interface. On the left, there is a file browser with a list of items: 0 files, 1 folder named 'biosimspace'. The 'biosimspace' folder is expanded, showing sub-folders: 'biosimspace_workshop', 'demo', 'free_energy_workshop', and 'python_and_data_workshop'. To the right of the file browser is a context menu with the following options: 'Upload', 'New ▾', and a refresh icon. Below these options are sections for 'Notebook:' (Python 3), 'Other:' (Text File, Folder, Terminal), and a timestamp '15 hours ago'. A red arrow points from the text 'open terminal' to the 'Terminal' option in the context menu.

open terminal

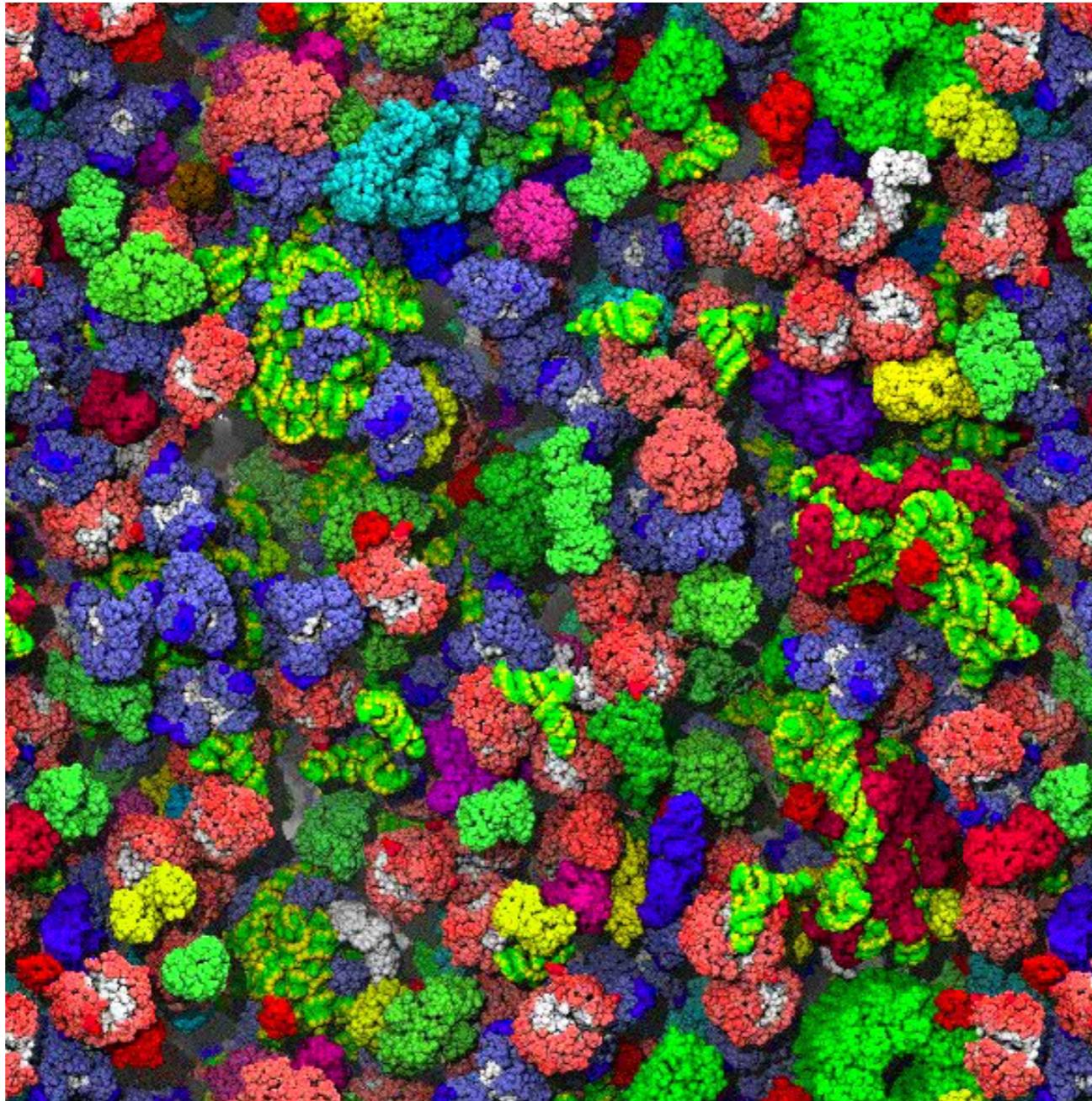


Logout Control Panel

```
jovyan@jupyter-6ee22867-2d3d9e-2d4f78-2d8420-2d9d15e7e0e79a:~$ update_biosimspace
```

The crowded cell

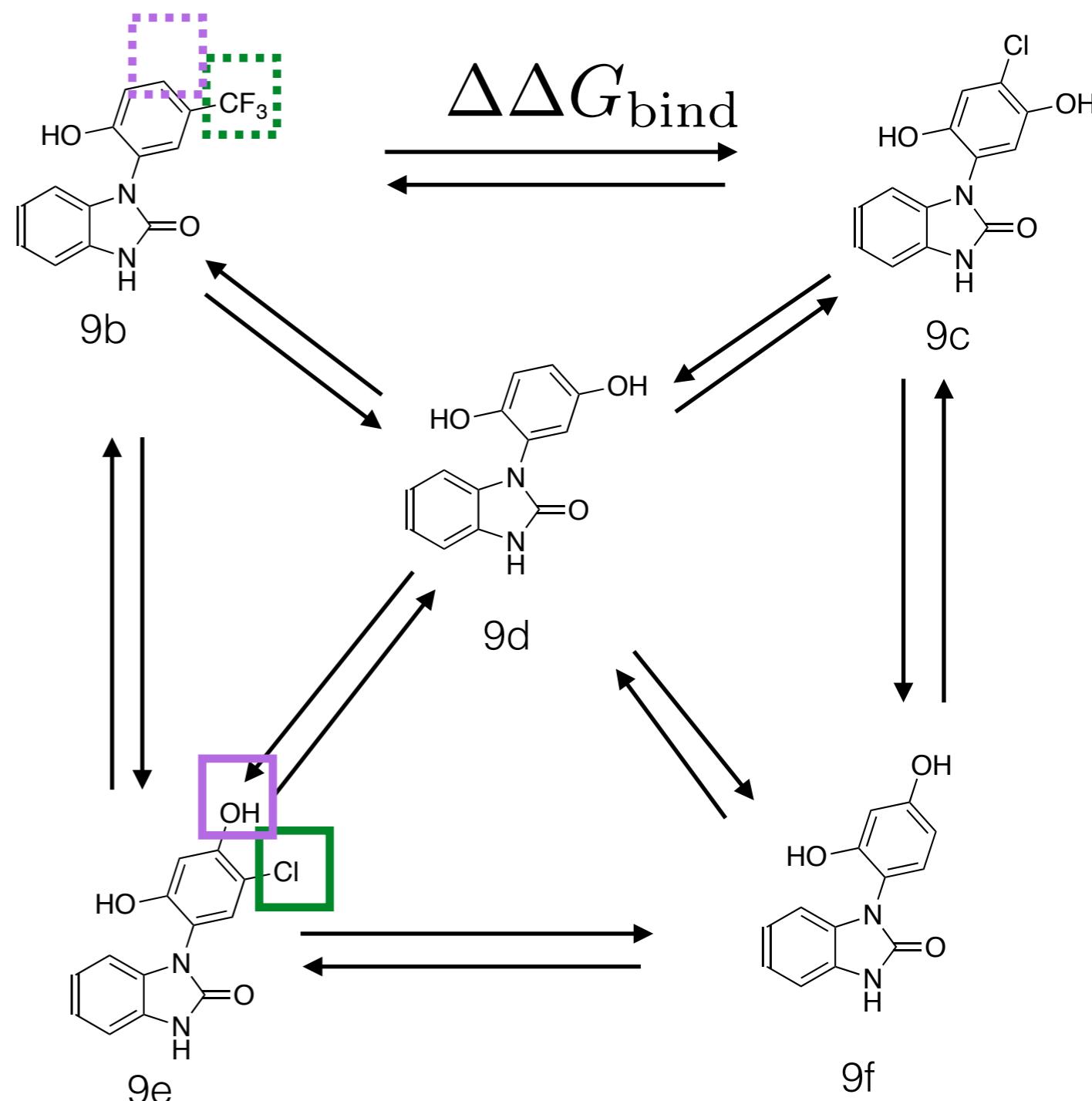
~2 million proteins in an E coli cell



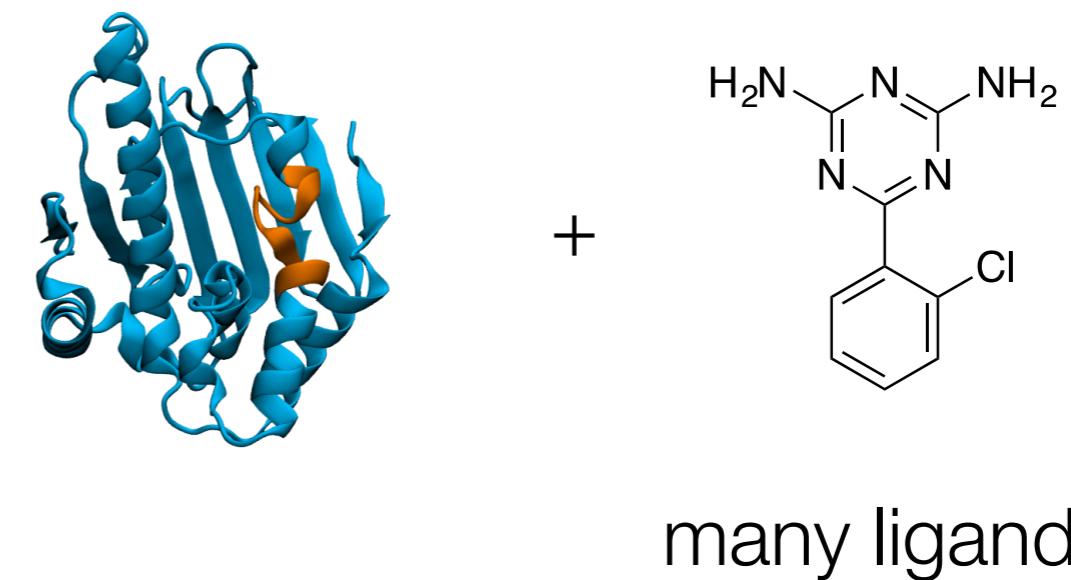
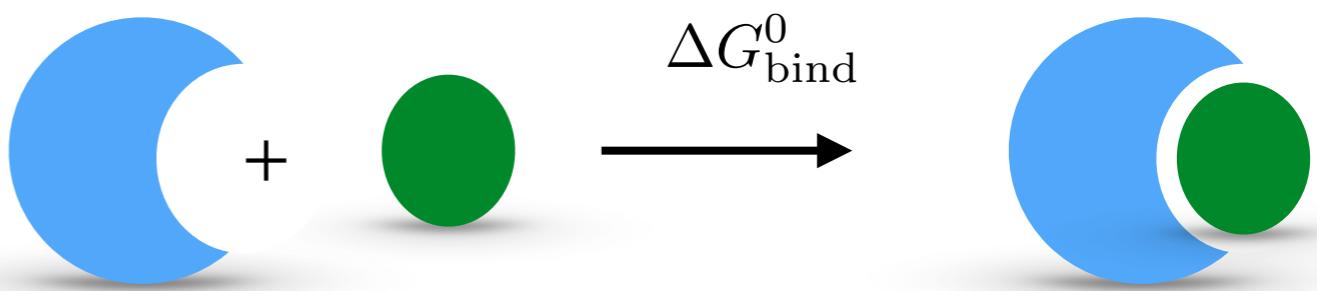
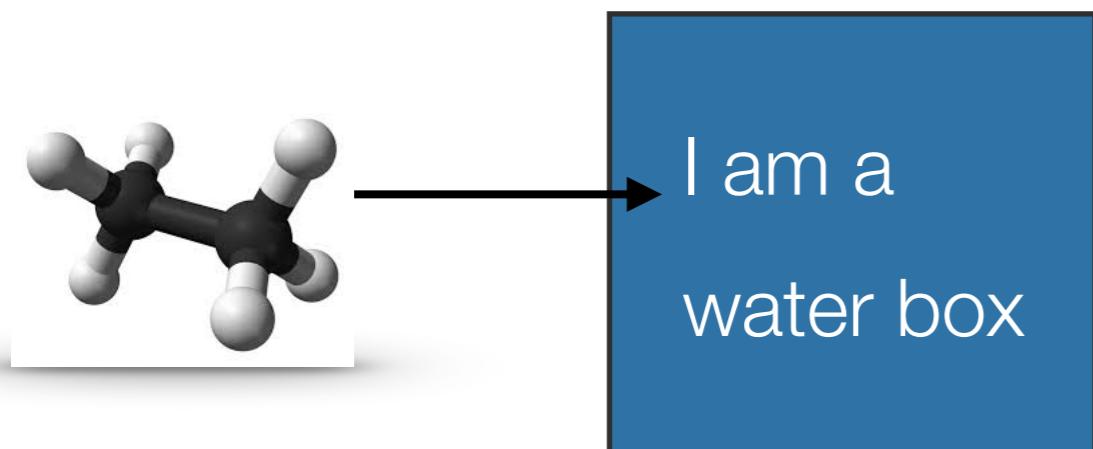
MCGuffee et al PLoS Comput Biol (2010)

Free energy predictions

Dataset	Cresset/ UoE		Wang et al. ²		Song et al. ³	
	R	MUE	R	MUE	R*	MUE*
Thrombin	0.88 ± 0.04	0.35 ± 0.04	0.71 ± 0.24	0.76 ± 0.13	0.76	0.46
TYK2	0.87 ± 0.02	0.60 ± 0.04	0.89 ± 0.07	0.75 ± 0.11	0.57	1.07
PTP1B	0.83 ± 0.04	0.84 ± 0.06	0.80 ± 0.08	0.89 ± 0.12	0.71	1.06
JNK1	0.81 ± 0.02	0.85 ± 0.04	0.85 ± 0.07	0.78 ± 0.12	0.47	1.07
MCL1	0.79 ± 0.02	1.30 ± 0.06	0.77 ± 0.05	1.16 ± 0.10	0.65	1.52
CDK2	0.69 ± 0.09	1.02 ± 0.08	0.48 ± 0.19	0.91 ± 0.12	0.47	0.97
BACE	tbd	tbd	0.78 ± 0.07	0.84 ± 0.08	0.43	1.20
p38	tbd	tbd	0.65 ± 0.09	0.80 ± 0.08	0.38	1.20



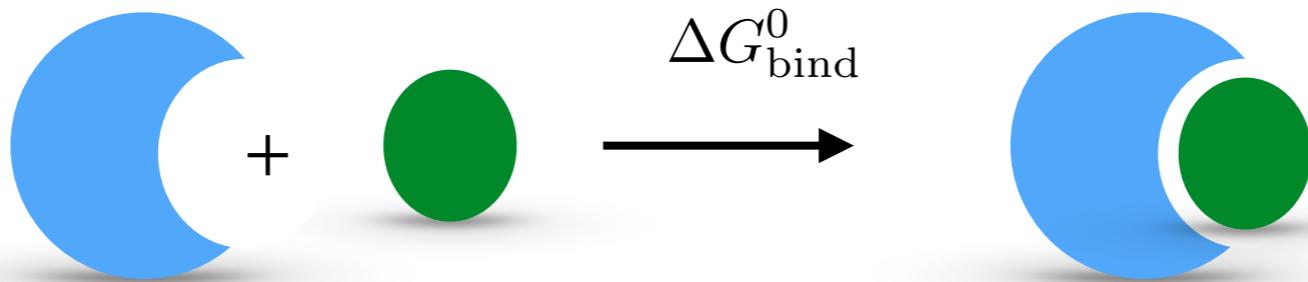
Problems we want to tackle



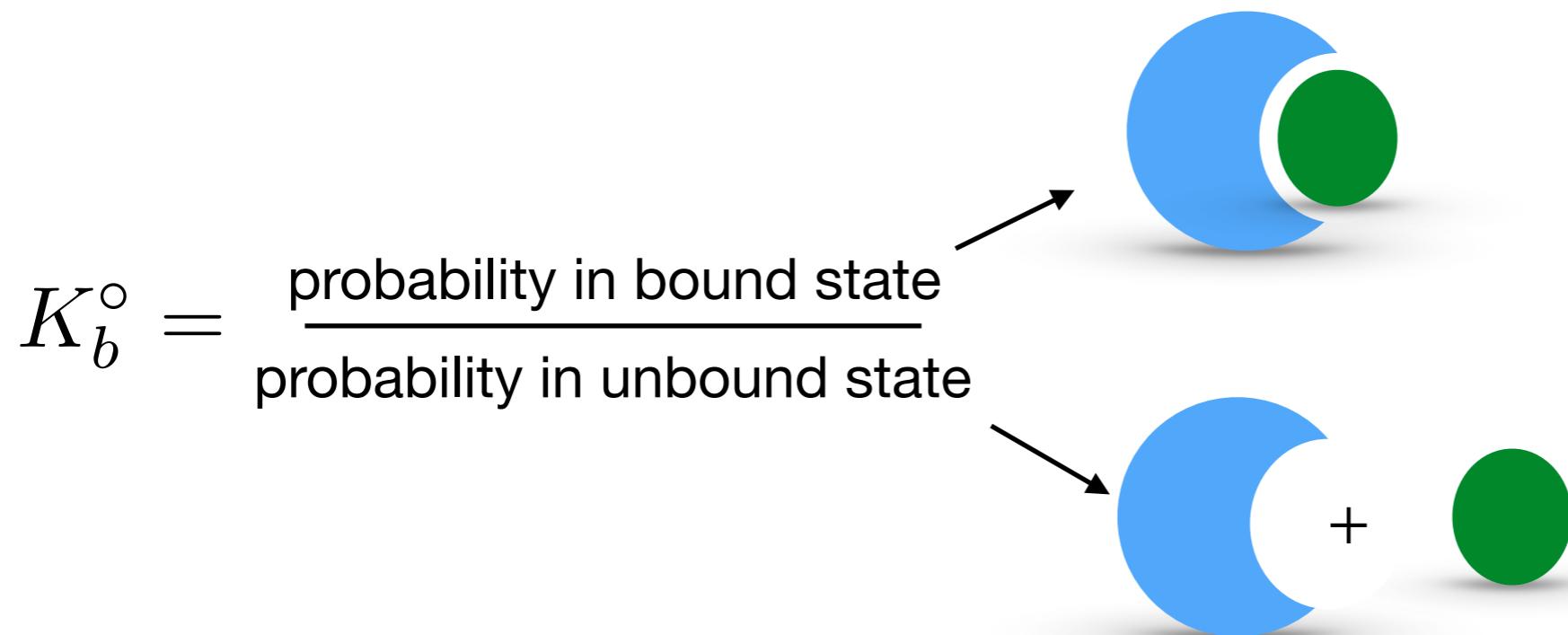
Compute the free energy of hydration

Predict affinity of a protein and ligand to bind

What is a free energy of binding?

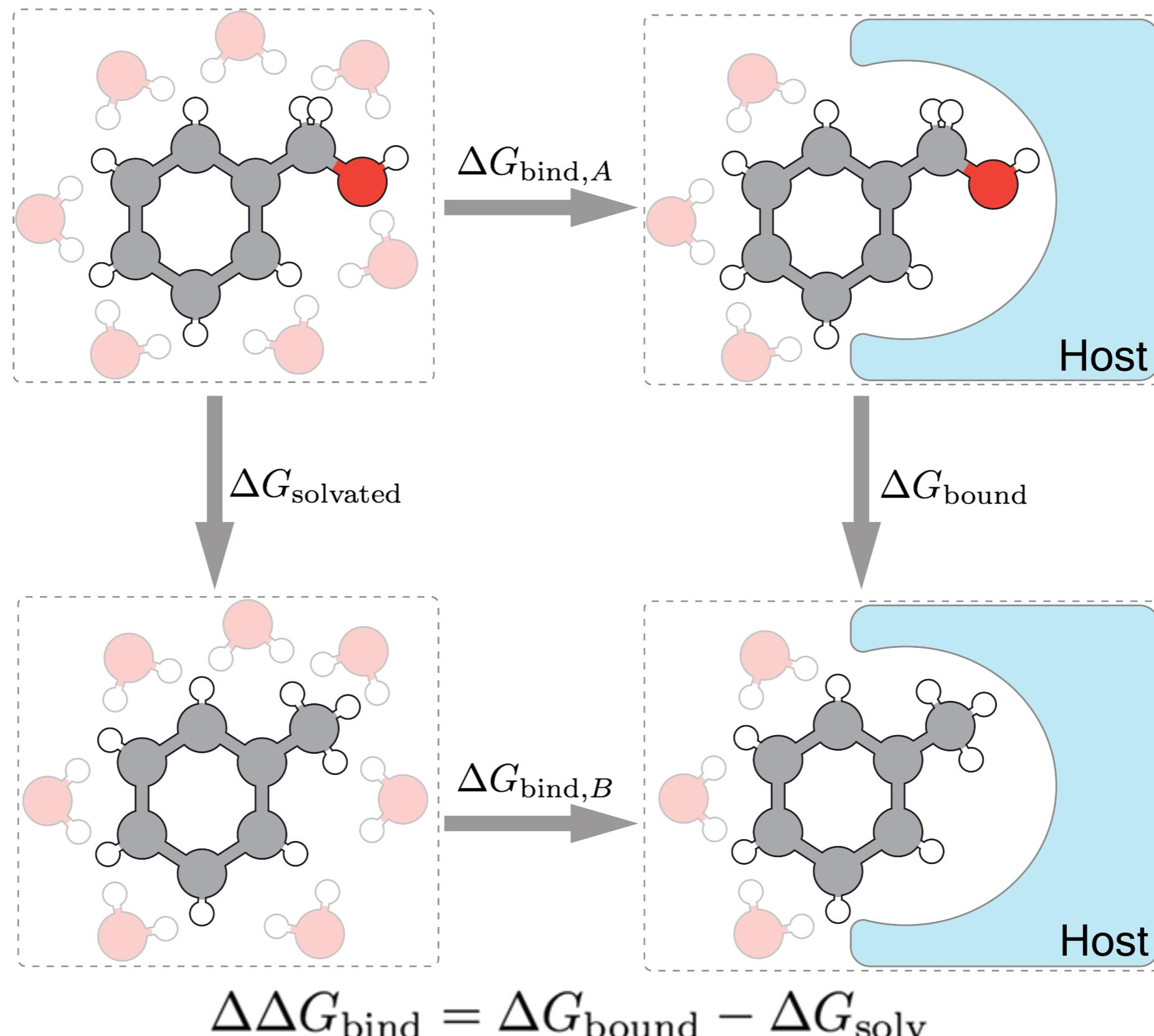


$$\Delta G_{\text{bind}} = -k_B T \ln K_b^\circ$$

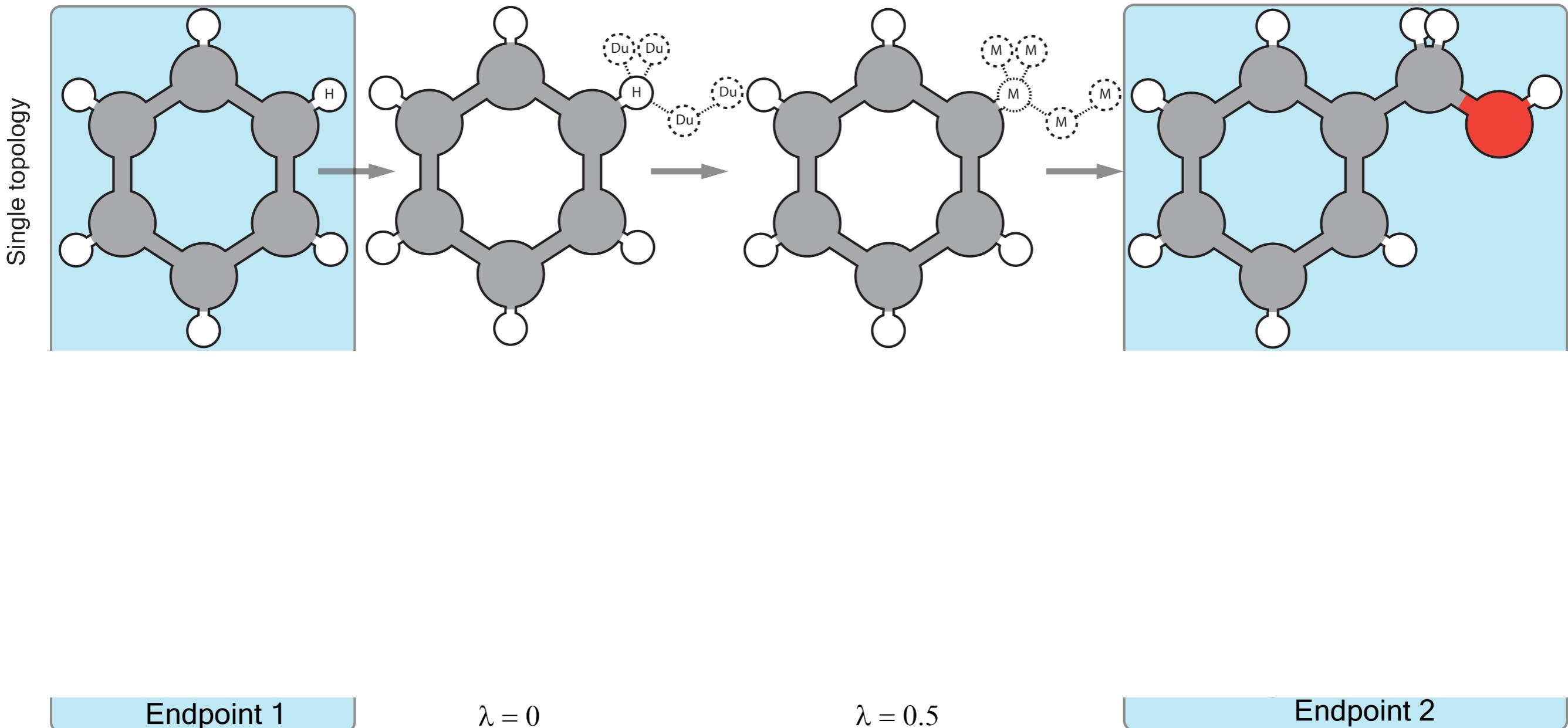


Sample this using
Molecular dynamics?

The alchemical pathway



How do we sample this alchemical pathway?



$$U(\lambda, \mathbf{x}) = (1 - \lambda)U_0(\mathbf{x}) + \lambda U_1(\mathbf{x}) + U_{\text{unaffected}}$$

MD and forcefields

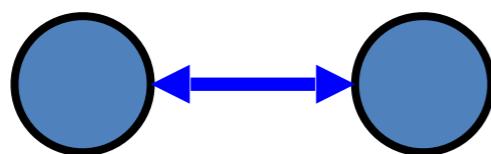
$$U = \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{\text{torsions}} \frac{V_n}{2} [1 + \cos(n\phi - \delta)] \\ + \sum_{\text{improper}} V_{imp} + \sum_{\text{LJ}} 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) + \sum_{\text{elec}} \frac{q_i q_j}{r_{ij}},$$

- The functional form and parameter set define a force field.
- Commonly used force fields include:
 - CHARMM (Chemistry at Harvard Molecular Mechanics)
 - AMBER (Assisted Model Building with Energy Refinement)
 - OPLS (Optimised Potentials for Liquid Simulations)

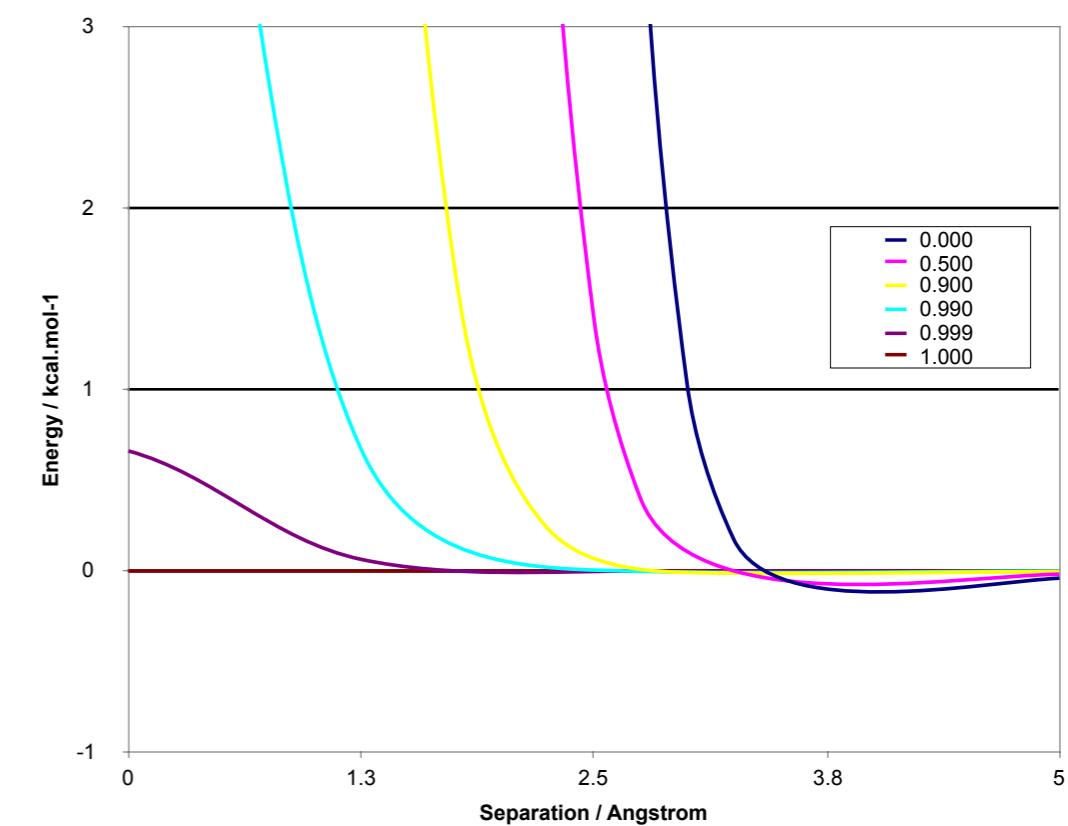
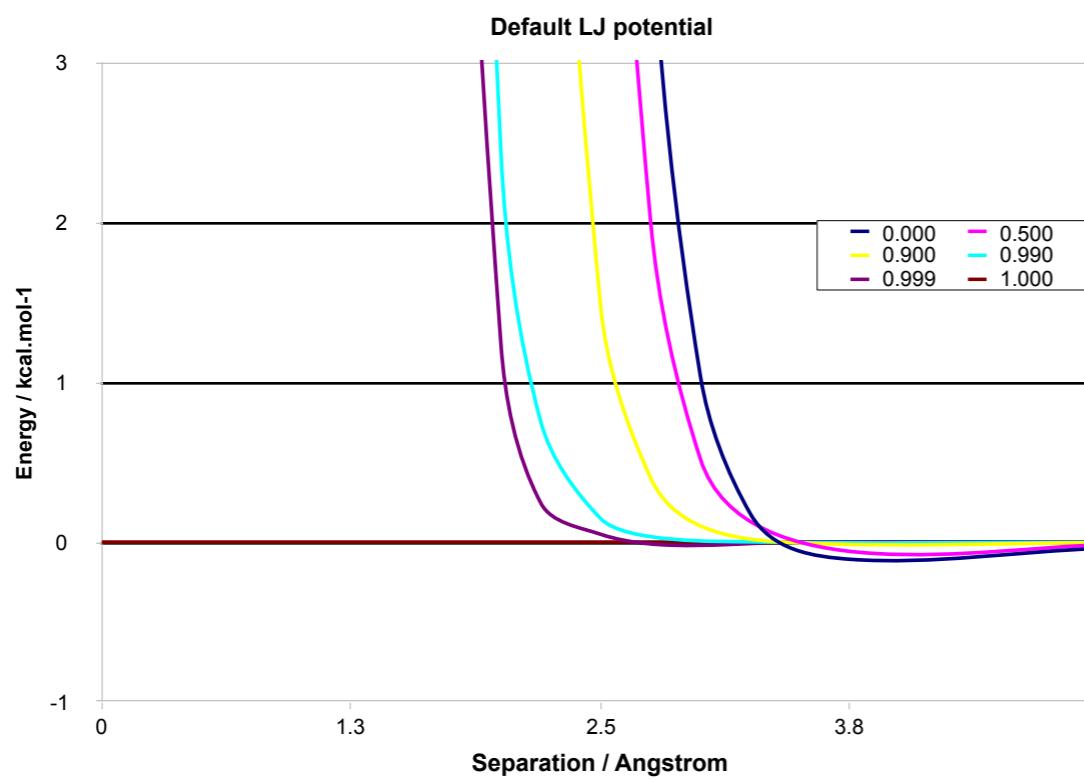
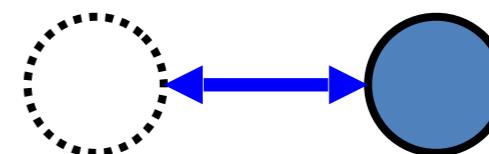
Softcore potential for dummy atoms

Zacharias et al. *J. Chem. Phys.* 100, 9025-, 1994

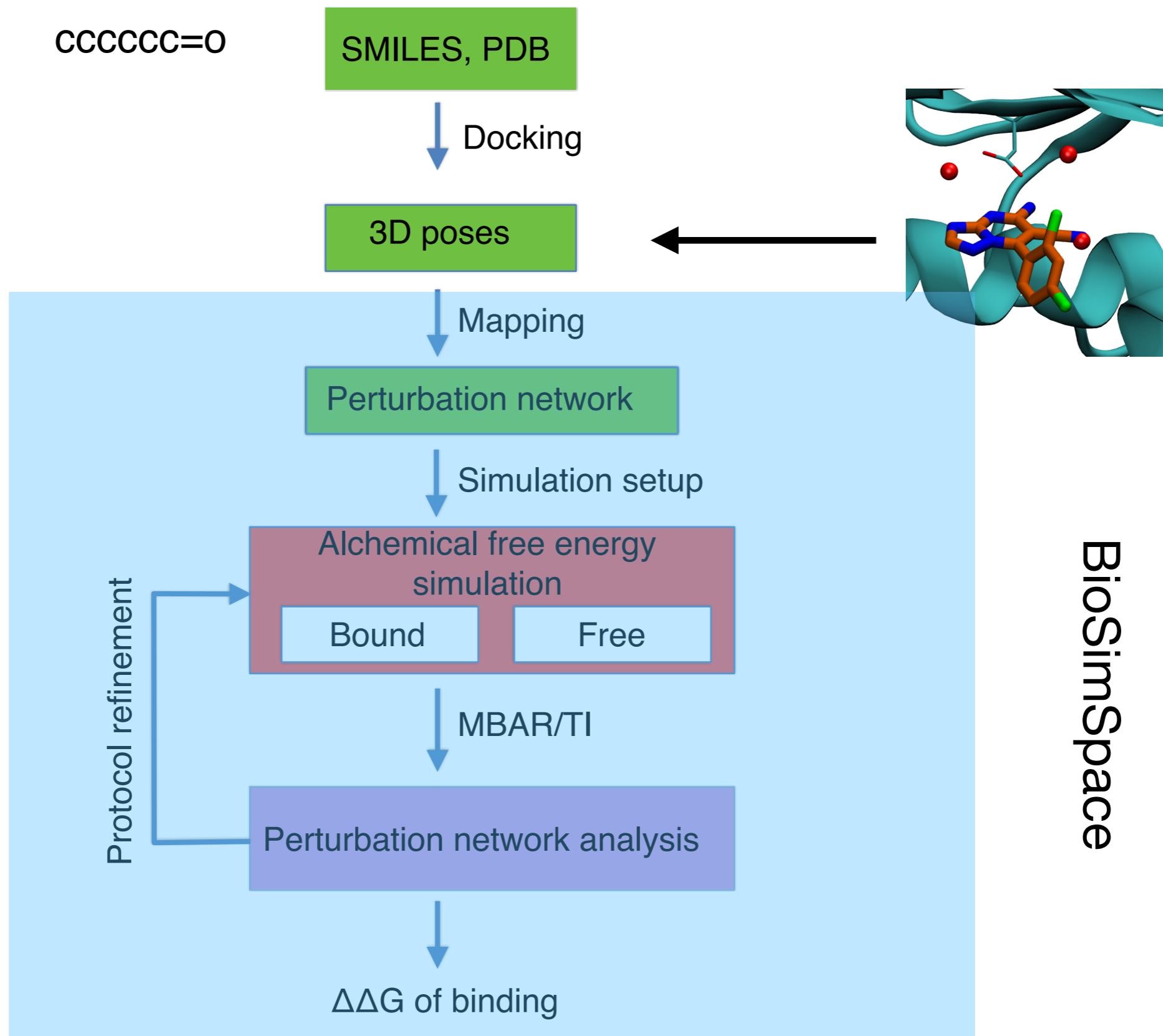
$$U_{\text{nonbonded},\lambda} = (1 - \lambda)4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}^{12}}{(\lambda\delta\sigma_{ij} + r_{ij}^2)^6} \right) - \left(\frac{\sigma_{ij}^6}{(\lambda\delta\sigma_{ij} + r_{ij}^2)^3} \right) \right] + \frac{(1 - \lambda)^n q_i q_j}{4\pi\epsilon_0 \sqrt{(\lambda + r_{ij}^2)}}$$



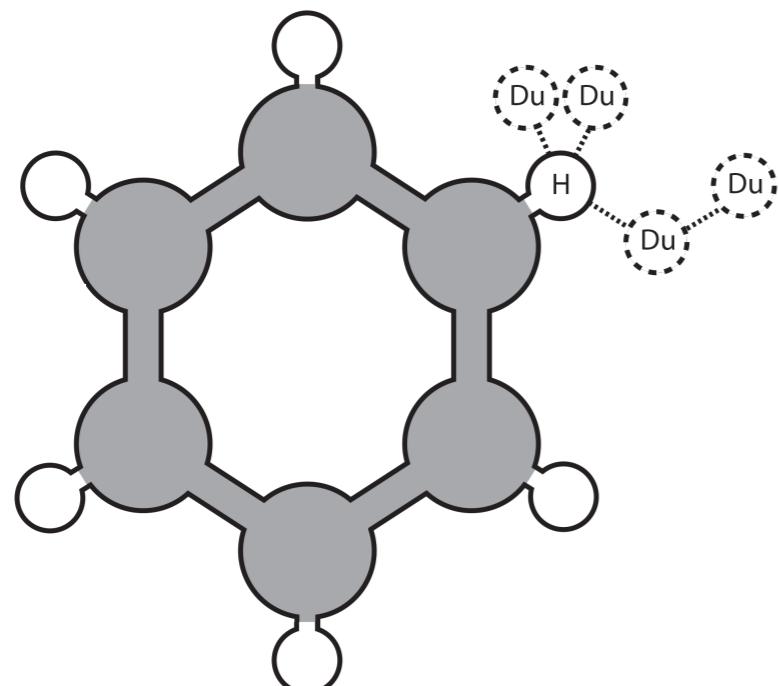
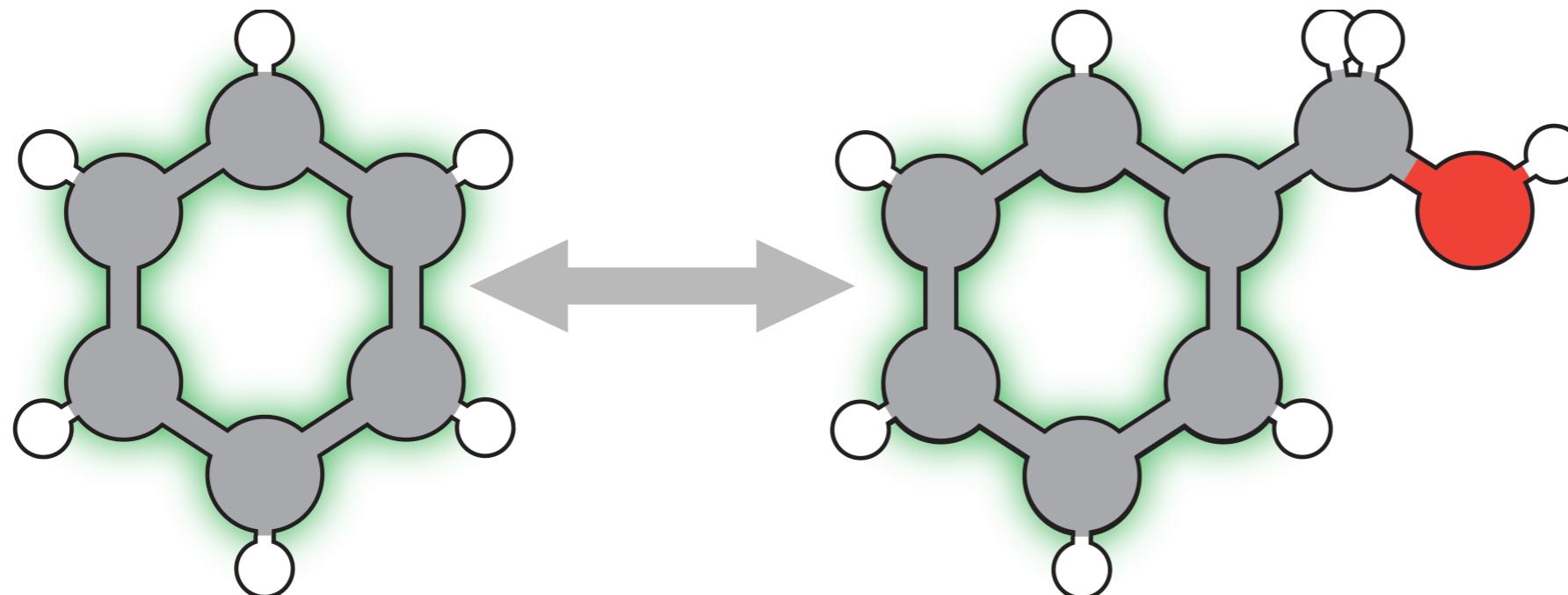
Dummy atoms!



Binding free energies in practice

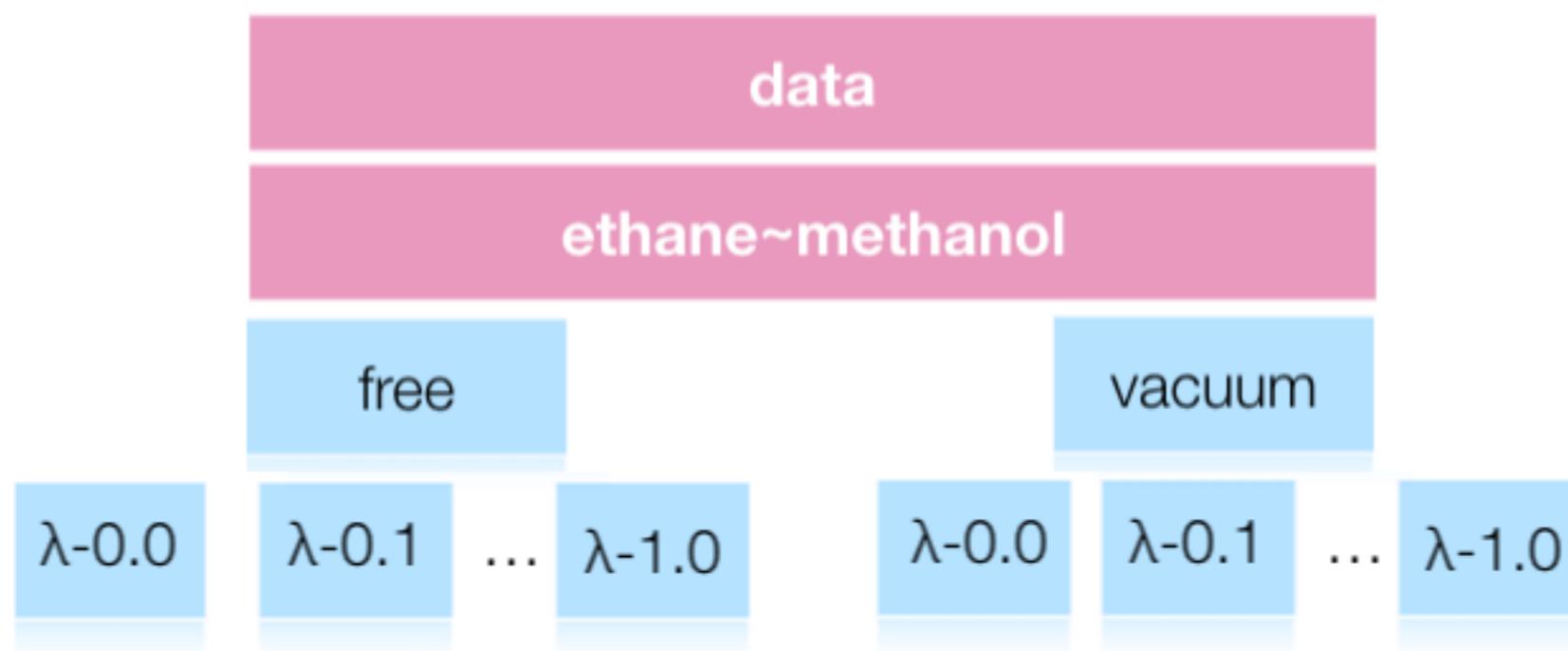
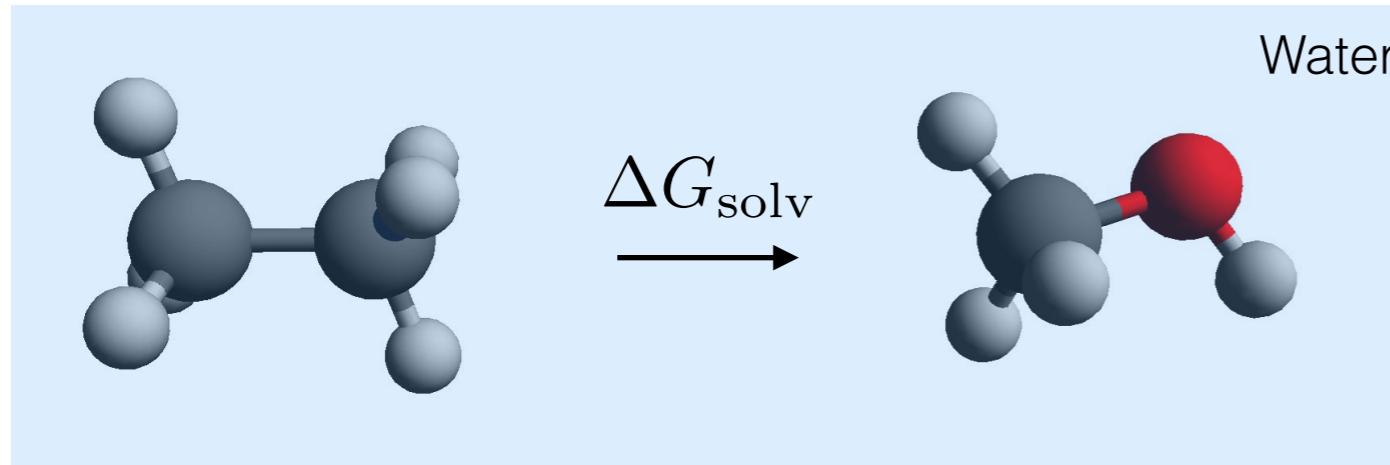


BioSimSpace can do the mapping between molecules



BioSimSpace holds a merged molecule and can write intermediates

BioSimSpace sets up all the simulation structure

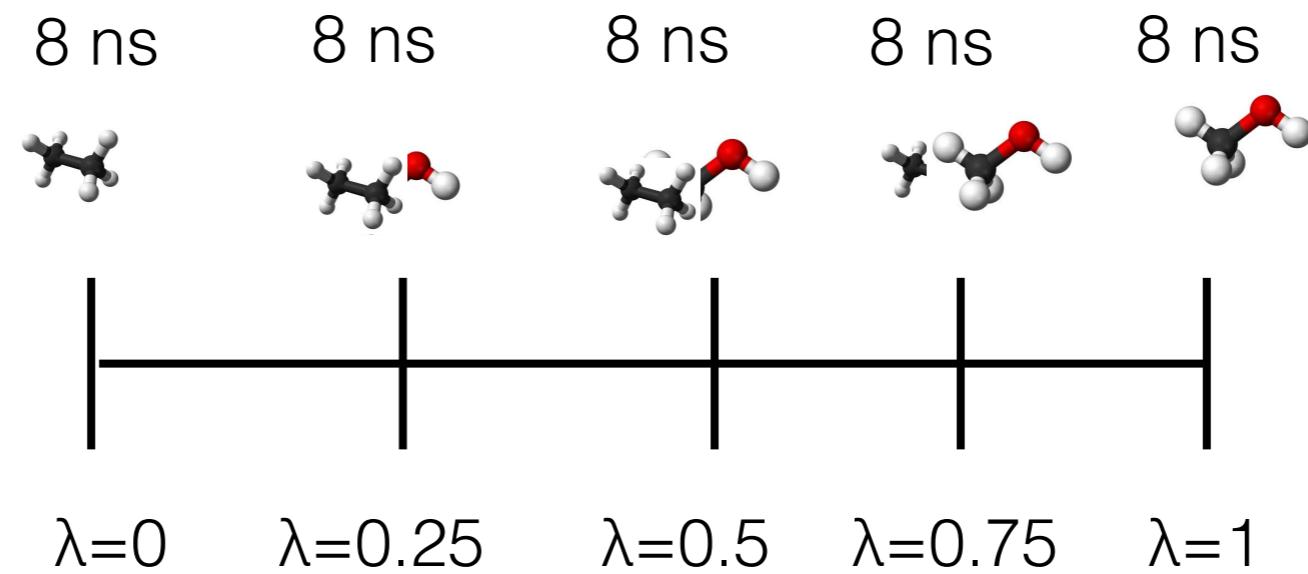
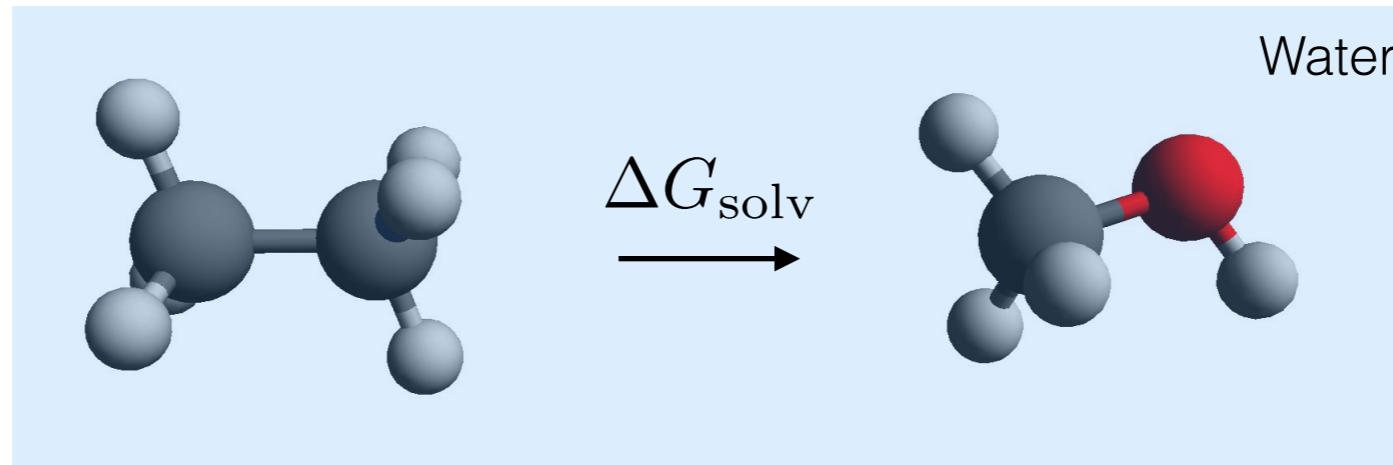


Analysis

What methods have you heard about?

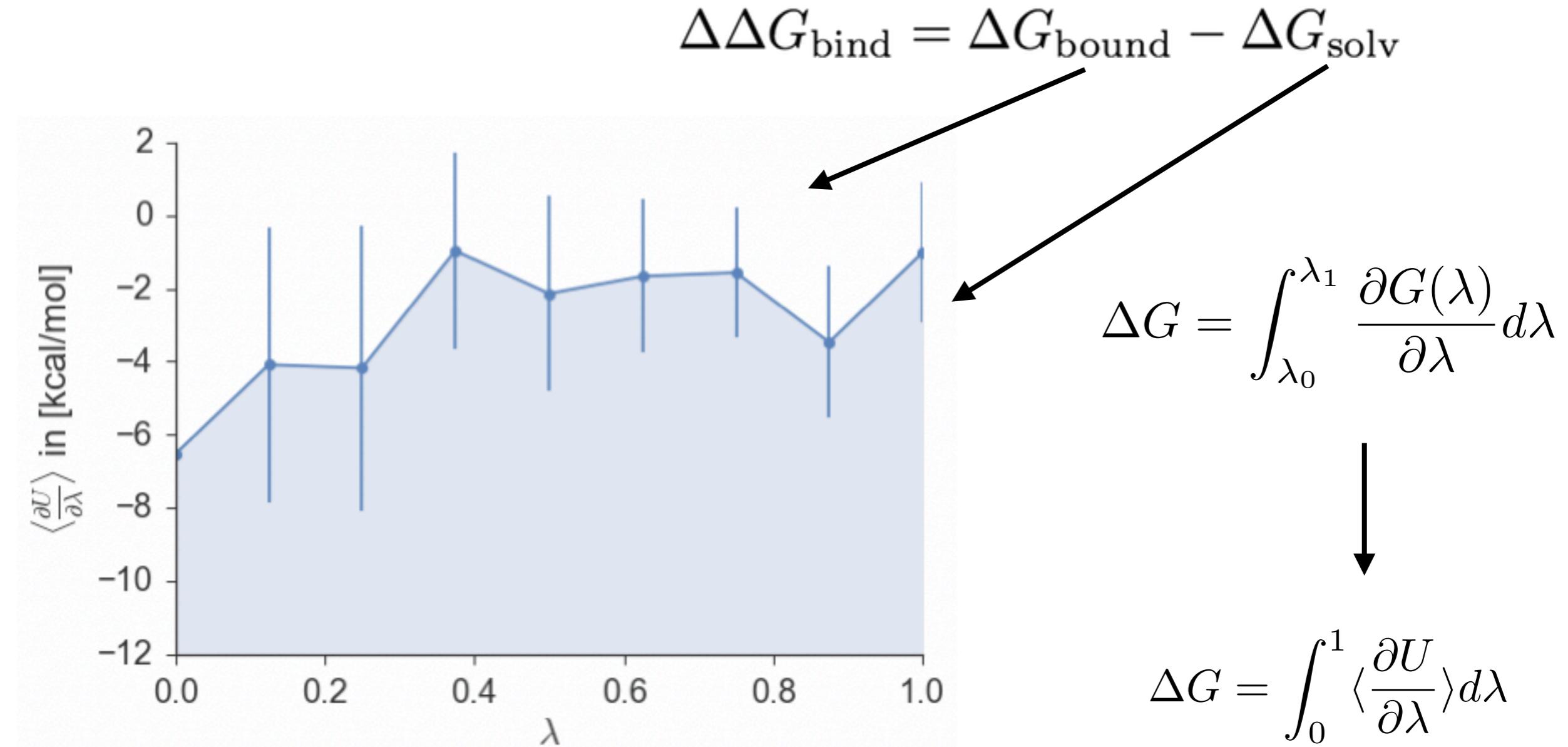
What do you know about them?

Analysis using the Zwanzig equation



$$\Delta G_{AB} = G(A) - G(B) = -k_B T \ln \langle \exp \left(-\frac{U_B - U_A}{k_B T} \right) \rangle_A$$

Thermodynamic integration



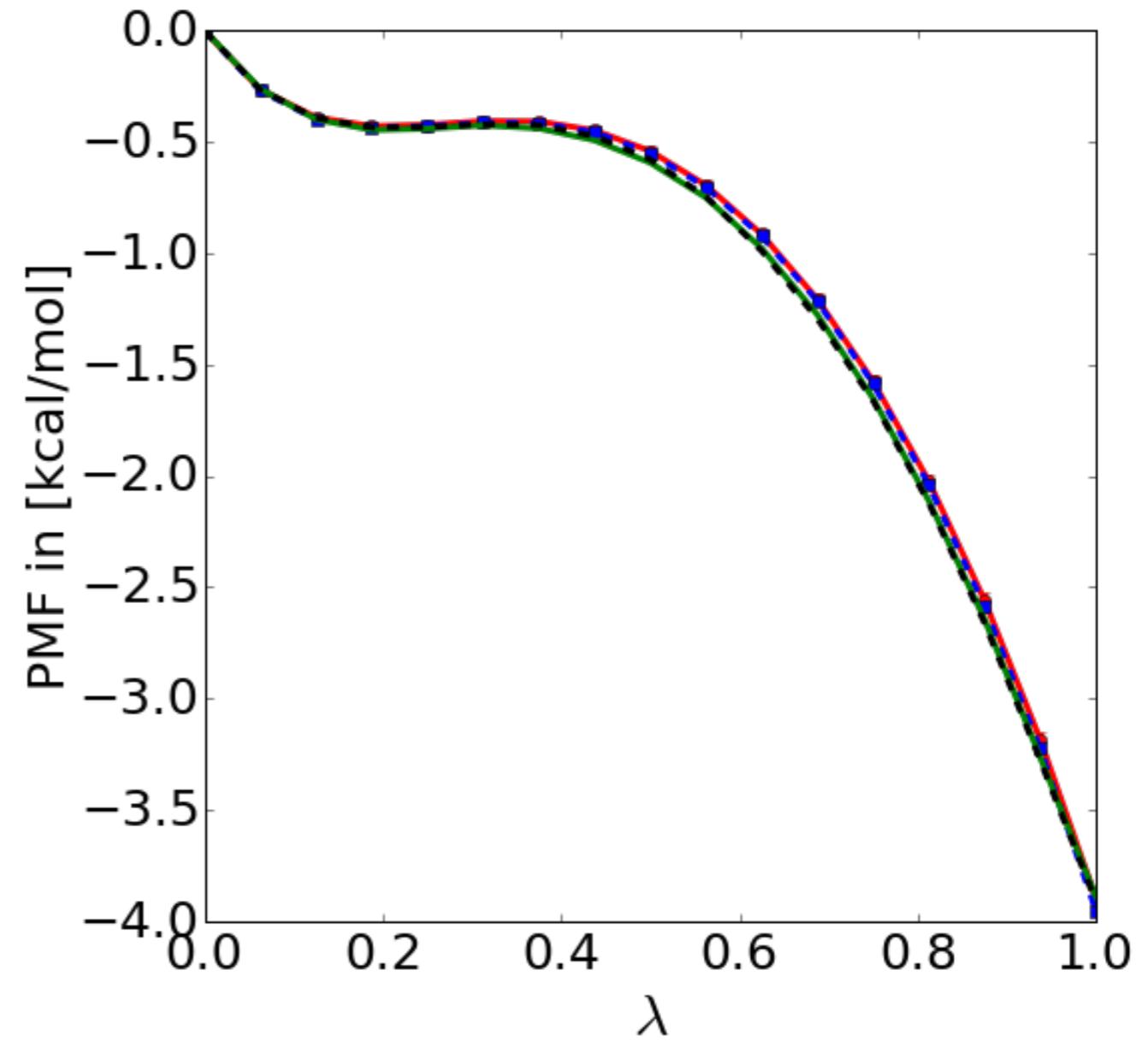
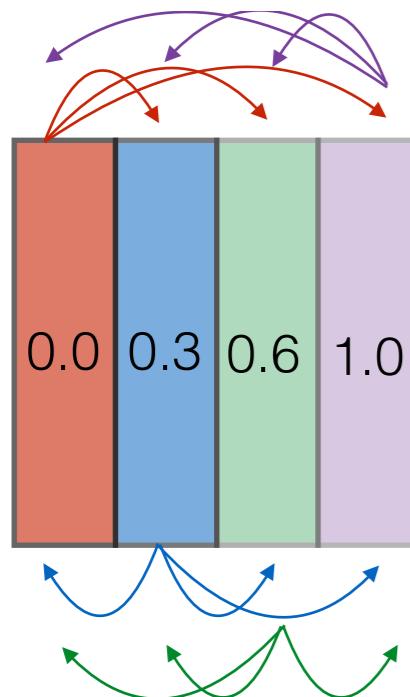
The integral represents the area under the curve.
There are different ways in which one can numerically integrate.

$$\mathcal{G}(\lambda_1) = -k_B T \ln \sum_{j=1}^N \sum_{n=1}^{K_j} \frac{\exp(-U(\lambda_i)/kT)}{\sum_{k=1}^N N_k \exp((\mathcal{G}(\lambda_k) - U(\lambda_k))/kT)}$$

MBAR is a reweighting estimator

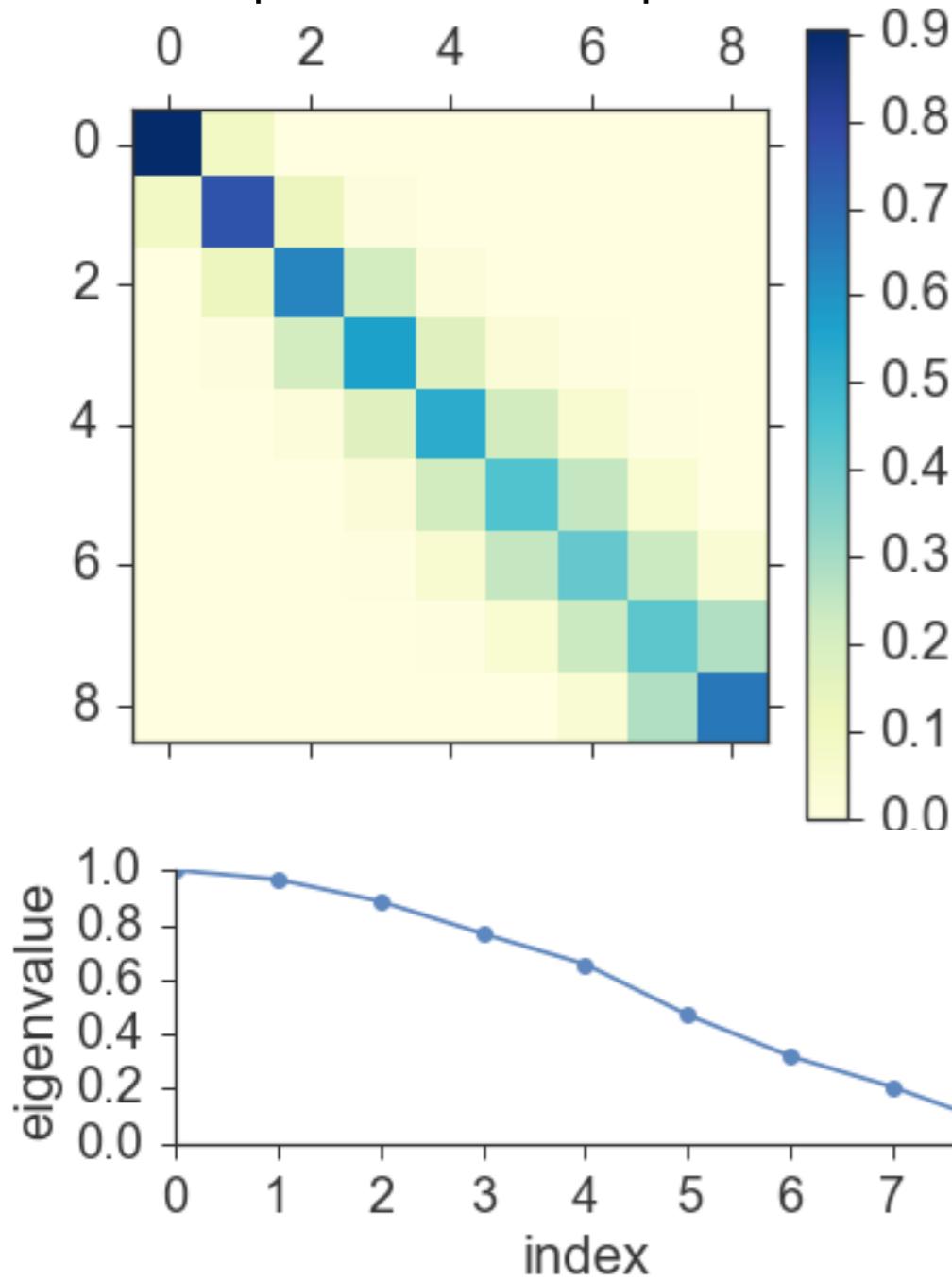
The free energy is correct up to an additive constant, which will cancel out when evaluating free energy differences.

$$\Delta G_{AB} = G(\lambda_1) - G(\lambda_0)$$



Overlap

Overlap matrix estimates the phase space overlap



Define a weight matrix \mathbf{W} containing the weight of each sample x_n .

$$W_{n,i}(x_n) = \frac{\exp(\beta F_i - U_i(x_n))}{\sum_{k=1}^K N_k \exp(\beta F_k - \beta U_k(x_n))}$$

Probability of $p_i(x_n)$ of x_n occurring at lambda window i .

$$\mathbf{O} = \mathbf{W}^T \mathbf{W} \mathbf{N}$$

\mathbf{N} is a diagonal matrix with the number of samples collected at each lambda window.

Best Practices

- Overlap
 - No element should be zero in the diagonal and above and below the main diagonal
 - None zero elements should at least be > 0.03
 - The number of samples needed for a good estimate will increase with a decrease in overlap.
 - Many eigenvalues close to 1 are a bad sign for good overlap and a good free energy estimate.
- Compare different estimators
- Run multiple independent runs
- Subsample your data

Analysis in BioSimSpace

```
# Solvate in a 60 angstrom box of TIP3P water.  
solvated = BSS.Solvent.tip3p(molecule=system, box=3*[ 60*BSS.Units.Length.angstrom])  
  
# Create the free energy protocol.  
protocol = BSS.Protocol.FreeEnergy(runtime=4*BSS.Units.Time.nanosecond, num_lam=9)  
  
# Initialise the binding free energy object.  
freenrg = BSS.FreeEnergy.Binding(solvated, protocol, work_dir="Binding_benzene_o_xylene" )
```

freenrg.run() → run the simulation

freenrg.analyse() → analyse the simulation

wraps around `analyse_freenrg mbar`

Thank you

MOVIE SCIENCE
MONTAGE



ACTUAL SCIENCE
MONTAGE

