

11/07/2022

- Today's lecture is a key step towards the following learning objective: Explain key concepts related to binding free energy calculations. Compare and contrast molecular docking and binding free energy calculations.
- Simulating and analyzing thermodynamic processes

Simulating and analyzing thermodynamic processes

- The lecture will cover
 - importance sampling
 - umbrella sampling
 - thermodynamic processes
 - replica exchange
 - statistical estimators for thermodynamic properties
- At the end of this module, you should be able to answer the following questions:
 - What is a thermodynamic process?
 - What are the benefits of umbrella sampling?
 - What are the benefits of performing replica exchange?
 - How do you know if there are enough states along a thermodynamic process?
- You should be able to calculate
 - expectation values
 - free energy differences between thermodynamic states

Importance Sampling

Importance sampling

- Sampling from one distribution and estimating quantities in another

- $\langle A \rangle_T = \int A(x) p_T(x) dx = \int A(x) \left(\frac{p_T(x)}{p_S(x)} \right) p_S(x) dx = \langle wA \rangle_S$

- $p_T(x)$ is the probability density in the target distribution
- $p_S(x)$ is the probability density in the sampled distribution

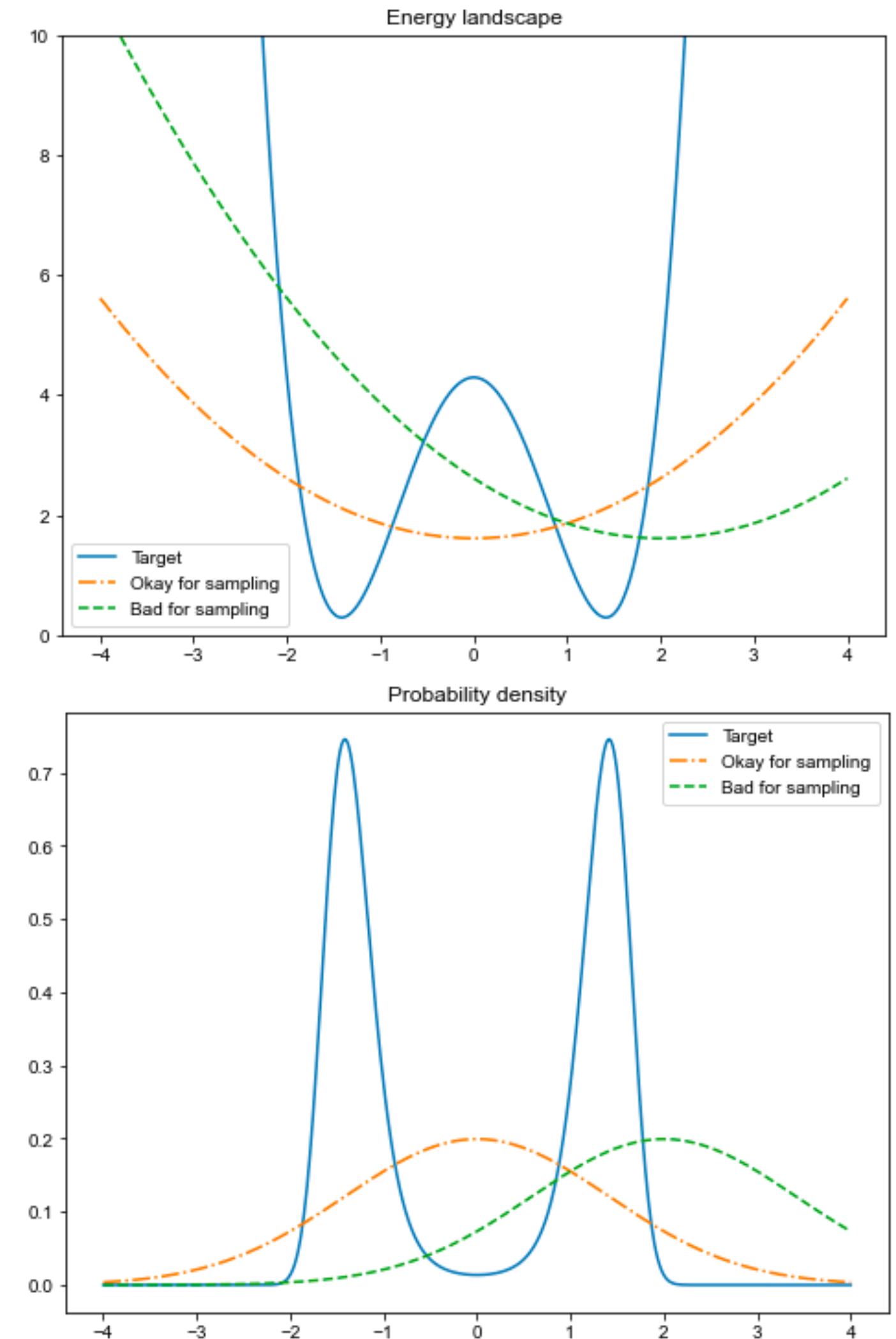
- $w = \left(\frac{p_T}{p_S} \right)$ is the ratio of weights in the two distributions

Why do importance sampling?

- Use one simulation to estimate quantities in multiple thermodynamic states, e.g. different temperatures
- Less computational expense to sample from one distribution, e.g. sample with molecular mechanics and calculate quantities in QM/MM distribution
- Sample from distribution with smaller configuration space, e.g. harmonic restraint towards a crystal structure

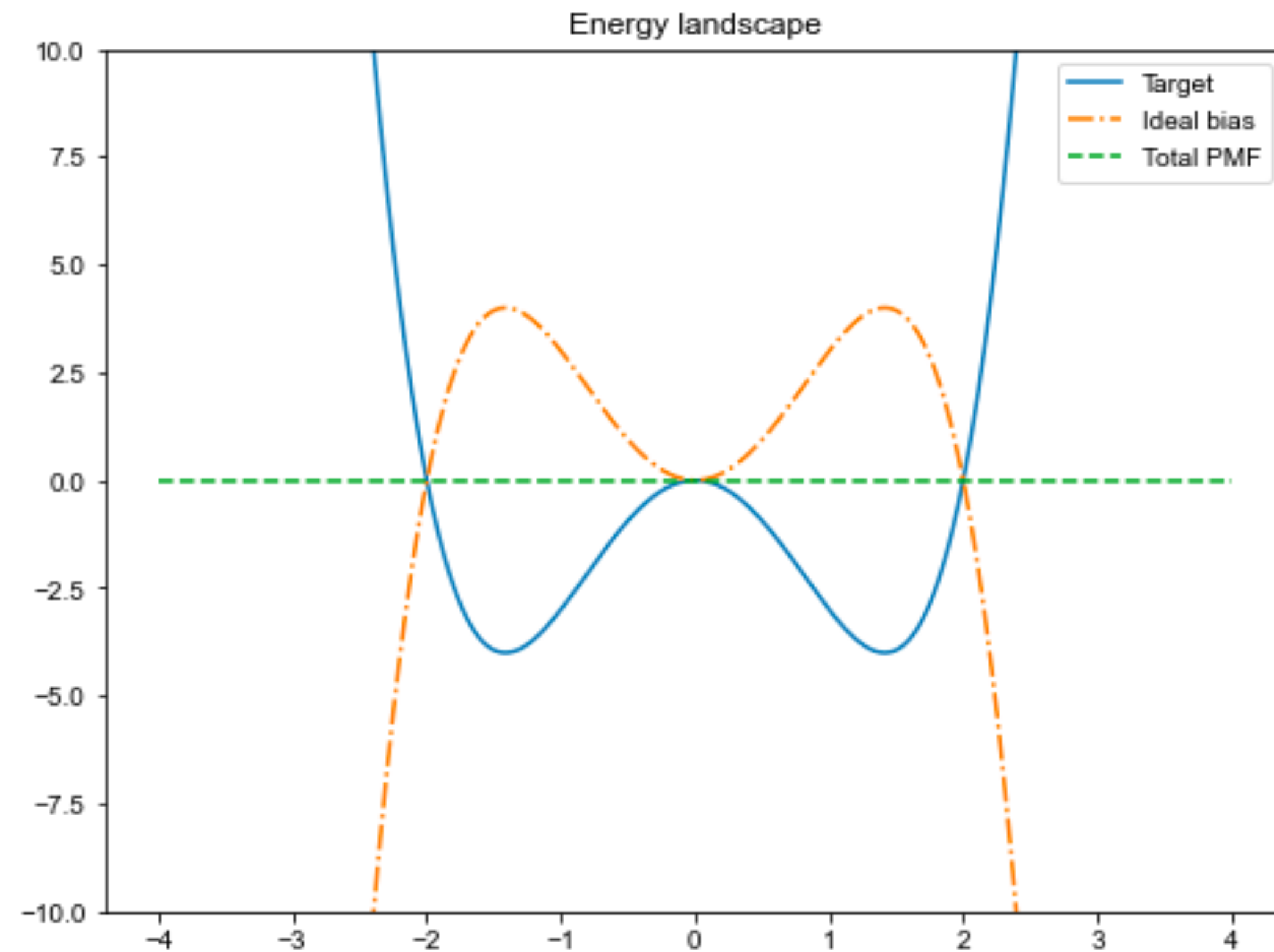
Caveat: Importance sampling

- The target and sampled distribution should be similar
- If the *support* differs, important configurations may be missed
- If the probability density significantly differs
 - the reweighing term can be noisy
 - thermodynamic expectations will require more samples to converge



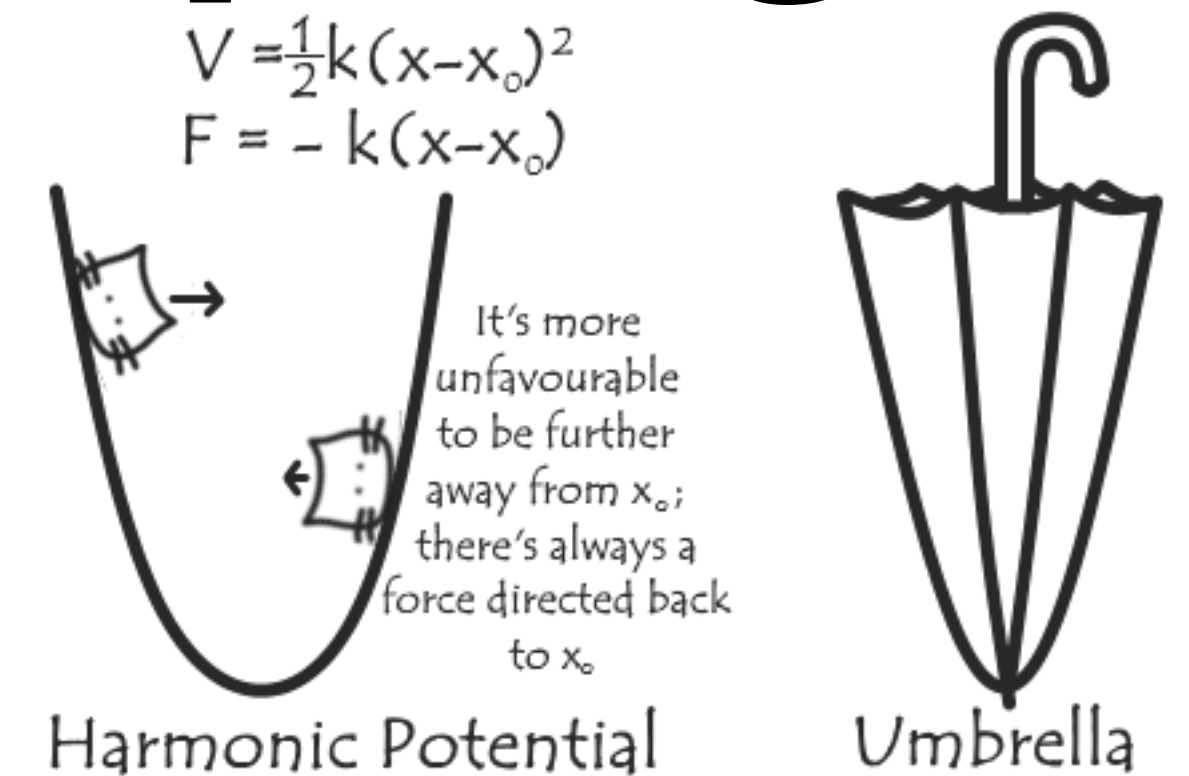
Idealized importance sampling

Covers the entire relevant configuration space along an order parameter

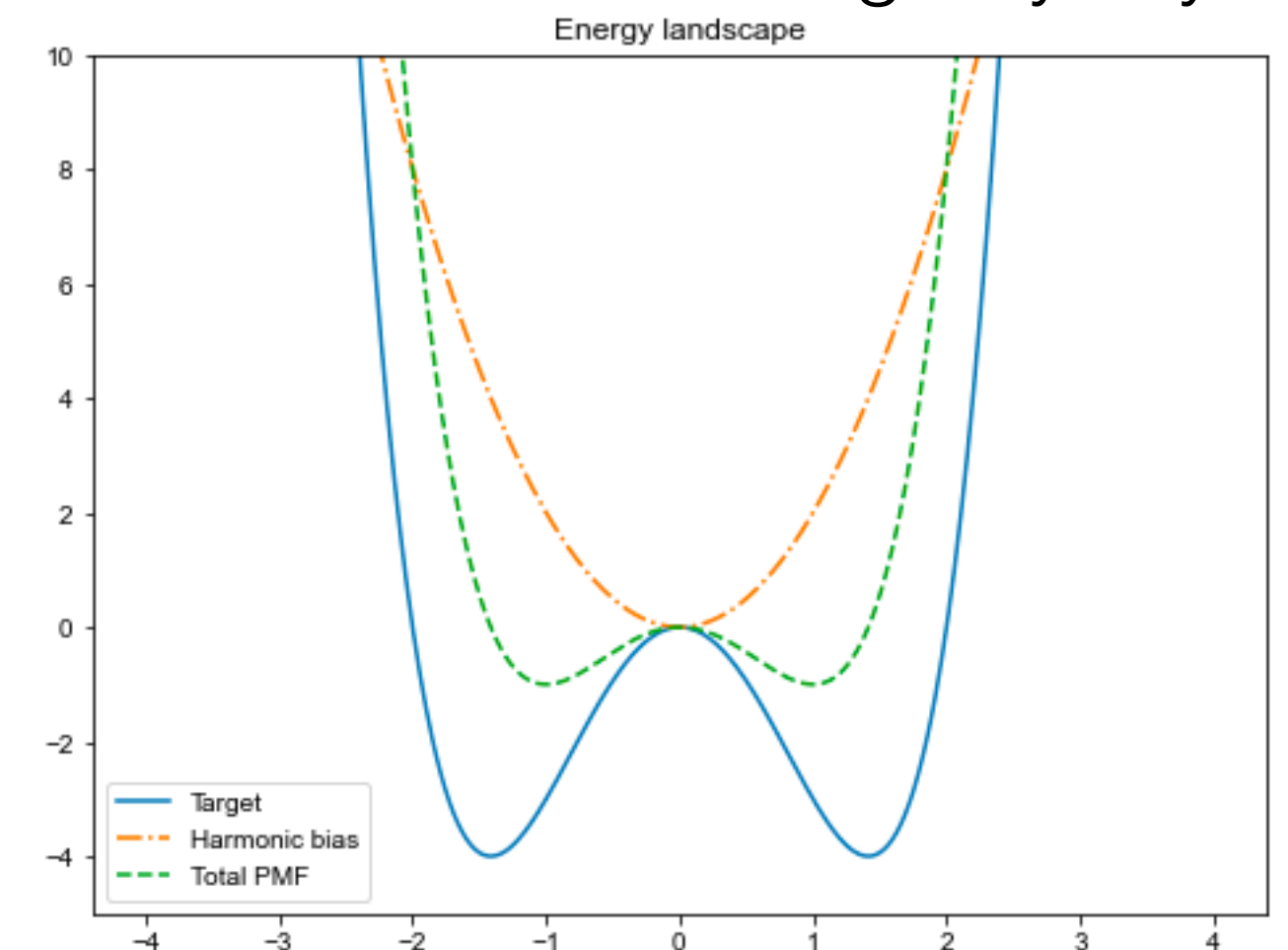


Harmonic umbrella sampling

- Sampling that *covers* the entire relevant configuration space along an order parameter
- Typically involves
 - a harmonic restraint towards a specific value of the order parameter, $U_b(x) = \frac{1}{2}k(z[x] - z_o)^2$
 - k is the spring constant
 - $z[x]$ is the value of the order parameter for configuration x
 - multiple simulations with different spring centers
 - the observed probability distributions should overlap



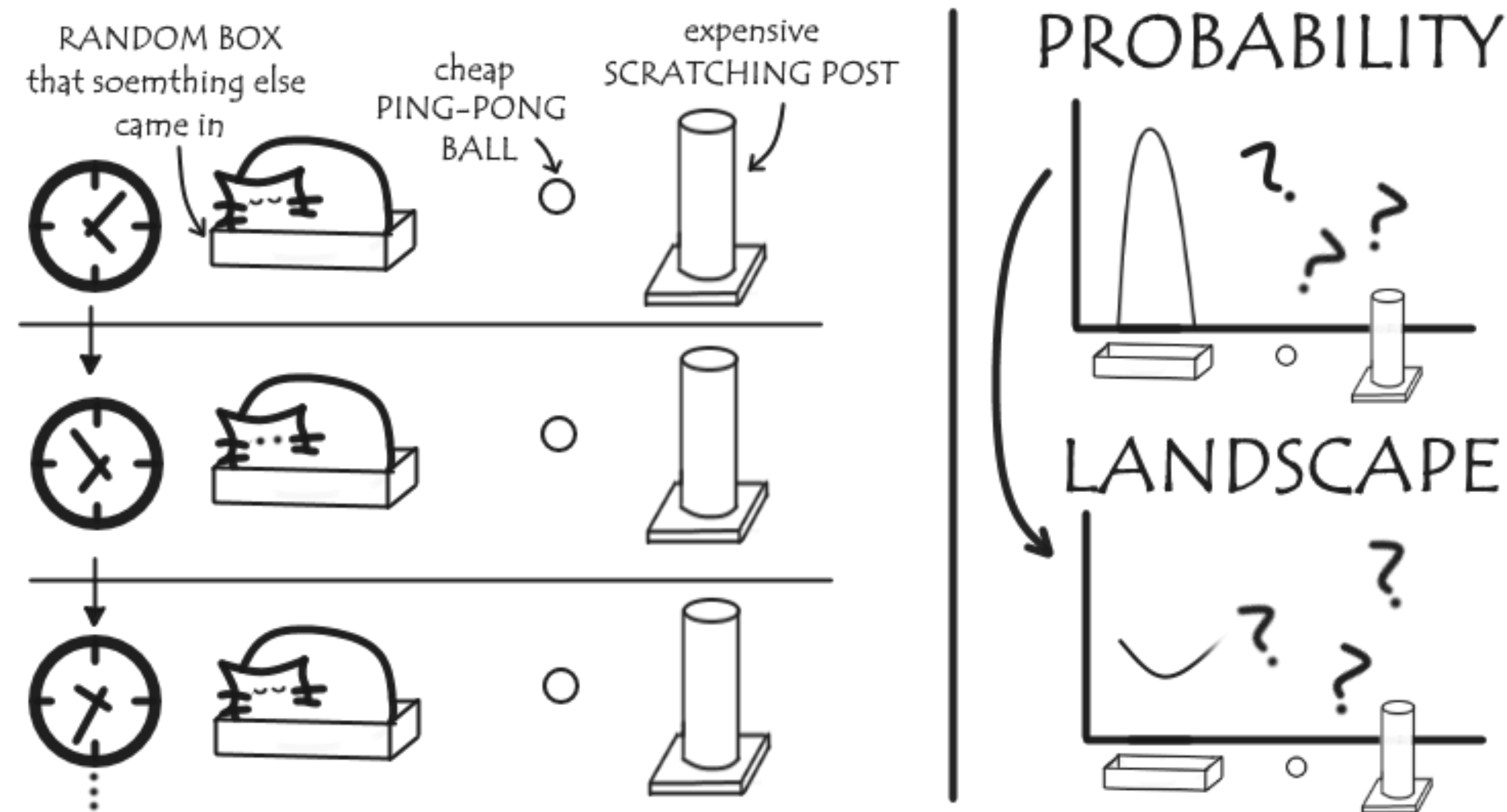
Used with permission from Fiona Naughton: <http://fiona-naughton.github.io/blog/2016/05/25/What-is-this-MD-thing-anyway>



A cat analogy of umbrella sampling

Which toy does the cat like best?

To quantify this, we can look at the fraction of time they spend near a specific toy.

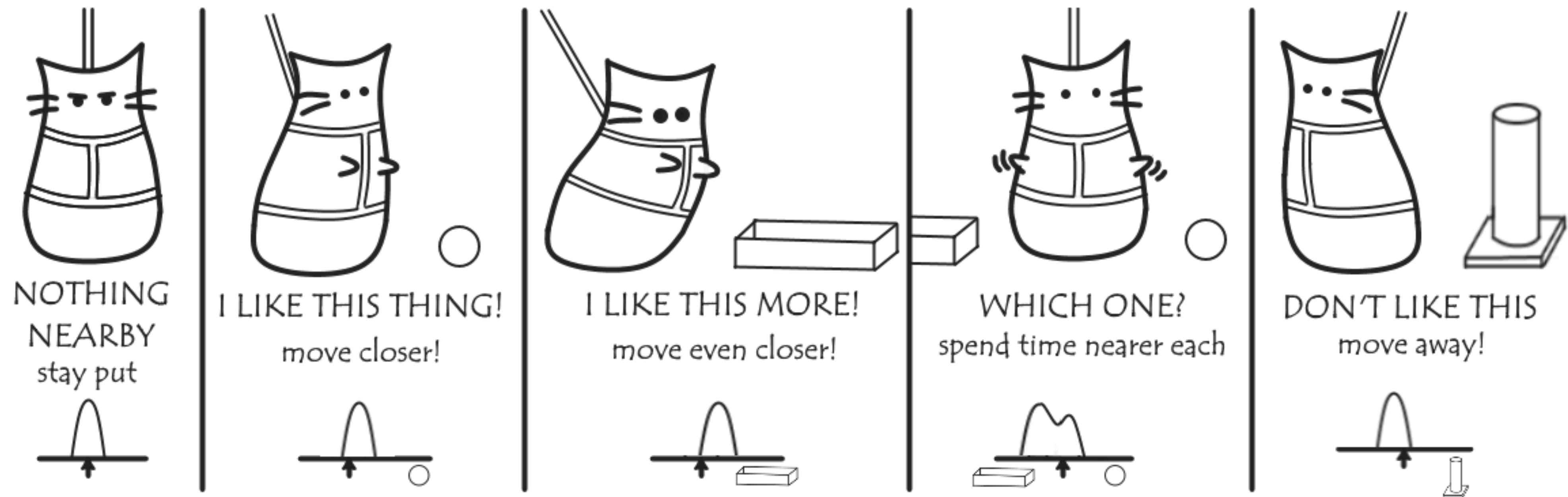


But we can be watching the cat for a long time...

Used with permission from Fiona Naughton: <http://fiona-naughton.github.io/blog/2016/05/25/What-is-this-MD-thing-anyway>

Umbrella sampling with a cat

If we put a restraint on the cat, we can determine what they prefer in a smaller area. Overall, we don't need to watch as long.



Used with permission from Fiona Naughton: <http://fiona-naughton.github.io/blog/2016/05/25/What-is-this-MD-thing-anyway>

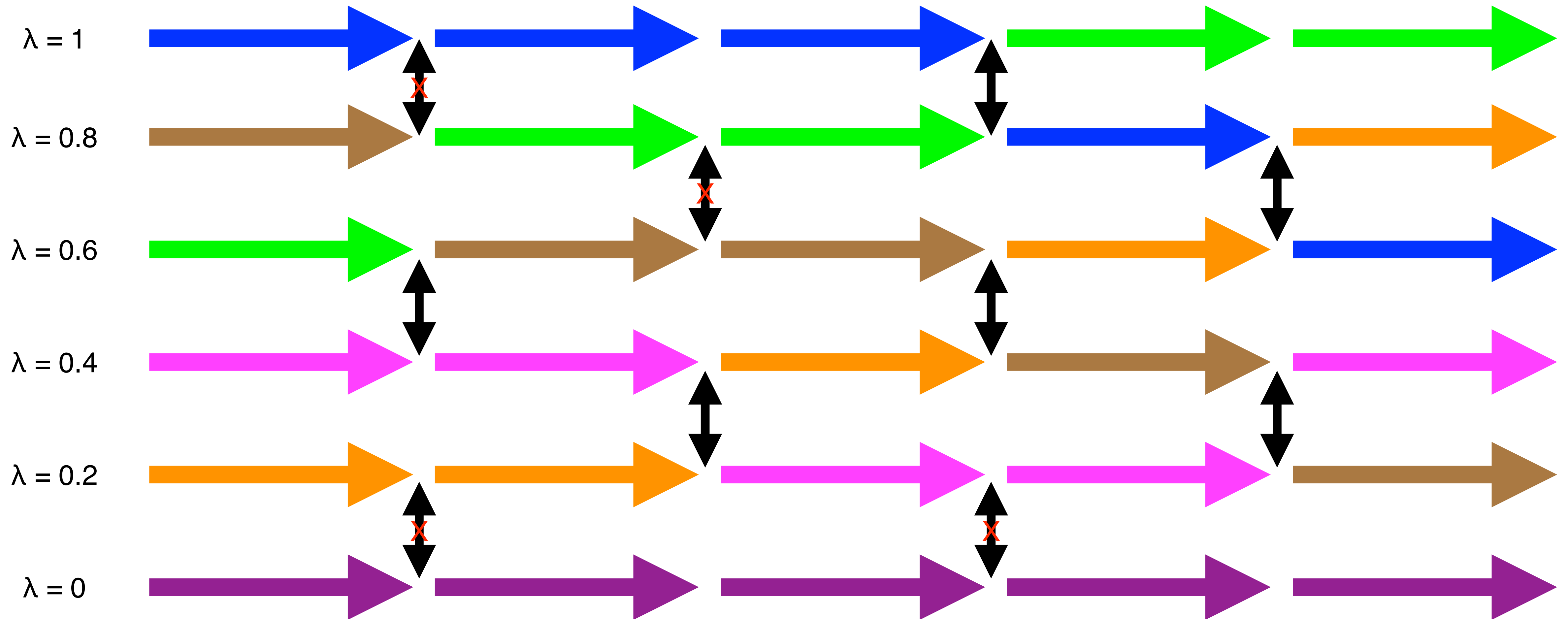
Thermodynamic processes

- A thermodynamic process involves a change in one or more variables that specify a thermodynamic state
- In general and physical chemistry, you learn about specifying a state with
 - temperature
 - pressure/volume
 - number of particles/chemical potential
- Processes include
 - isothermal expansion - a change in volume but not temperature
 - adiabatic expansion - a change in volume without heat transfer in and out of a system
- In molecular simulations, additional variables can define the potential energy
 - harmonic spring constant and center (umbrella sampling)
 - alchemical parameter for
 - transforming one substituent into another
 - decoupling a molecule from solvent

Replica exchange

- Simulations of multiple thermodynamic states with different parameters
 - originally, variation in temperature
 - often used for alchemical coupling
- Configurations from thermodynamic states are periodically swapped
 - Equivalently, thermodynamic parameters are swapped
 - Swapping satisfies detailed balance; both states sample from respective Boltzmann distributions

Swapping

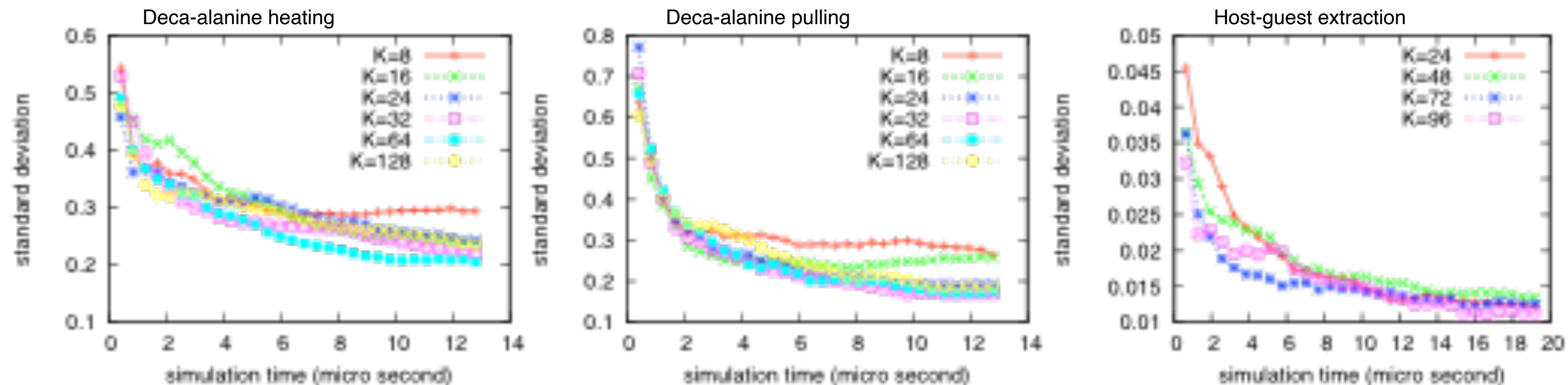


Why perform replica exchange?

- Improve mixing of MCMC chains; higher-entropy states help sample configurations in lower-entropy states
- At high temperature, energetic barriers are crossed more quickly than at low temperature
- Decoupled ligands move freely compared to bound ligands. See http://mypages.iit.edu/~dminh/images/pubs/dock_1hnn.gif.
- Minimal added computational expense

Selecting thermodynamic states

- Thermodynamic state selection has been thought to be a tricky optimization problem
- In Nguyen and Minh (2016) [1]
 - processes were simulated 100x each for each number of states, K
 - the standard deviation of the free energy was evaluated as a function of the total simulation time



- If there are not enough states, the convergence curve levels off
- If there are enough states, the standard deviation of free energy estimates depends on the aggregate simulation time and is insensitive to the number of states.

Review

- What is a thermodynamic process?
- What are the benefits of umbrella sampling?
- What are the benefits of performing replica exchange?
- How do you know if there are enough states along a thermodynamic process for replica exchange?

Estimating thermodynamic properties

Estimating Thermodynamic Properties

- Molecular simulations are used to calculate thermodynamic and kinetic properties
- In general, the thermodynamic properties are
 - expectation values of an observable, including
 - probability of the observable having a certain range of values
 - potential of mean force with respect to the observable
 - free energy differences between thermodynamic states
 - in biomolecular systems, $\Delta G \sim \Delta A$

What is ΔG ?

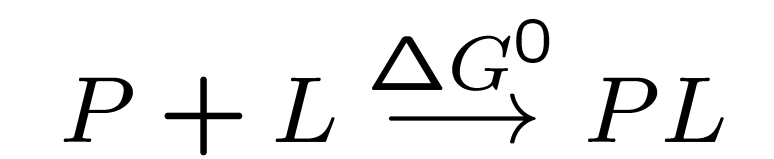
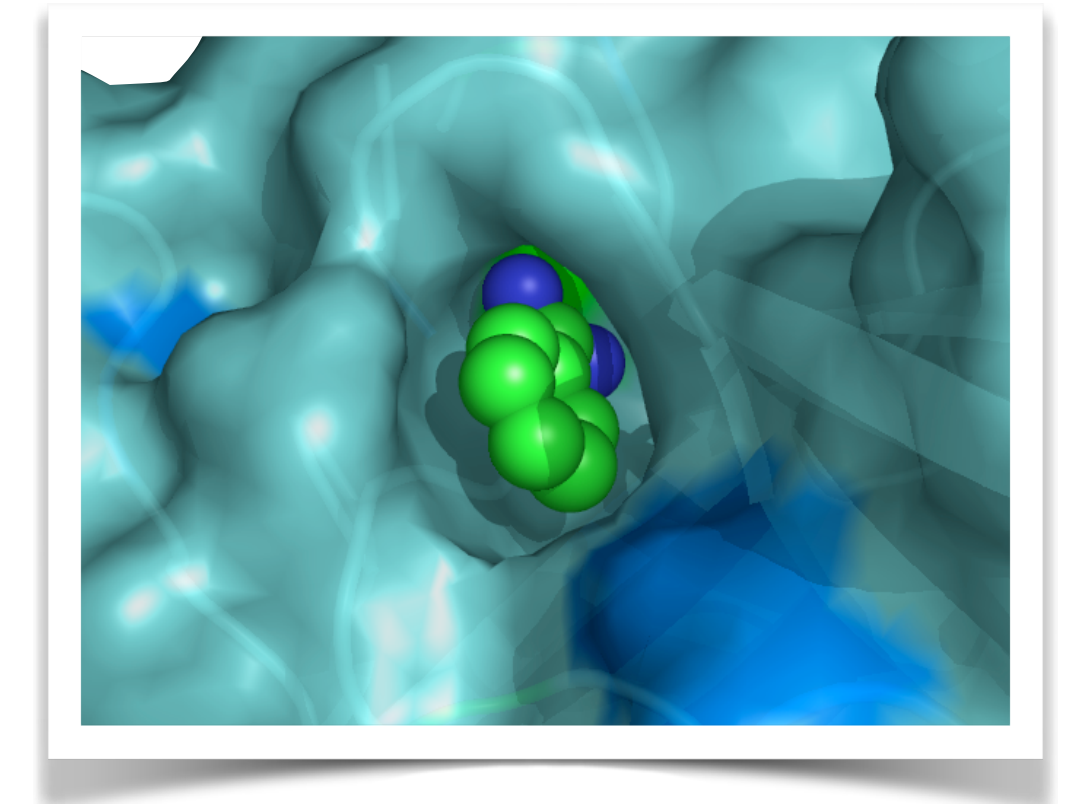
- ΔU is the change in average internal energy.
 - internal energy can be computed for individual structures
 - in biomolecular simulations, internal energy is modeled by the molecular mechanics force field
- $\Delta H = \Delta U + \Delta(pV)$ is the change in enthalpy
 - in biomolecular simulations, change in pV is usually negligible
- ΔG is the Gibbs free energy
 - at constant pressure and temperature, dictates
 - spontaneity and
 - equilibrium constant of process
 - in biomolecular simulation, interest in free energy differences between
 - conformations of a macromolecule
 - thermodynamic states with different temperature, pressure, volume, or other parameters
 - $\Delta G = \Delta H + T\Delta S$, but ΔS is very challenging to compute

What is ΔA ?

- ΔA is the Helmholtz free energy
 - at constant volume and temperature, dictates
 - spontaneity and
 - equilibrium constant of process
- in biomolecular simulation, ΔA and ΔG are usually assumed to be equal

How are free energy calculations useful?

- Noncovalent binding between molecules (see [2])
 - Design molecules to manipulate protein function
 - Recognize toxins
 - Identify enzyme functions
 - Protein design: design binders to target molecule
 - Aid medicinal chemistry, guide synthesis
- Hydration free energies
 - Part of binding free energy & solubility
- Conformational free energies relevant to
 - biological mechanism
 - binding free energy



(from [1])

**How are $\Delta G/\Delta A$ calculated from
molecular simulations?**

Basic Statistical Mechanics

- In the Boltzmann distribution, the probability of a configuration r^N with energy $U_s(r^N)$ is,

$$\pi_s(r^N) \propto \exp [-\beta U_s(r^N)] \text{ (unnormalized)}$$

$$\rho_s(r^N) = \exp [-\beta U_s(r^N)] / Q_s \text{ (normalized)}$$

- A partition function is the normalizing constant of the Boltzmann distribution

$$Q_s = \int \pi_s(r^N) dr^N$$

- The free energy difference is related to a ratio of partition functions

$$\beta(A_1 - A_0) = -\ln \left(\frac{Q_0}{Q_1} \right)$$

The Zwanzig Relation: Derivation

- From before, $\beta(A_1 - A_0) = -\ln \left(\frac{Q_0}{Q_1} \right)$.
- Substituting in partition functions, $\beta(A_1 - A_0) = -\ln \left(\frac{\int e^{-\beta U_1(r^N)} dr^N}{\int e^{-\beta U_0(r^N)} dr^N} \right)$.
- Multiplying by one, $\beta(A_1 - A_0) = -\ln \left(\frac{\int e^{-\beta U_1(r^N) + \beta U_0(r^N) - \beta U_0(r^N)} dr^N}{\int e^{-\beta U_0(r^N)} dr^N} \right)$.
- Defining the potential energy difference $\Delta U(r^N) = U_1(r^N) - U_0(r^N)$,
$$\beta(A_1 - A_0) = -\ln \left(\frac{\int e^{-\beta \Delta U(r^N)} e^{-\beta U_0(r^N)} dr^N}{\int e^{-\beta U_0(r^N)} dr^N} \right)$$

The Zwanzig Relation: In Practice

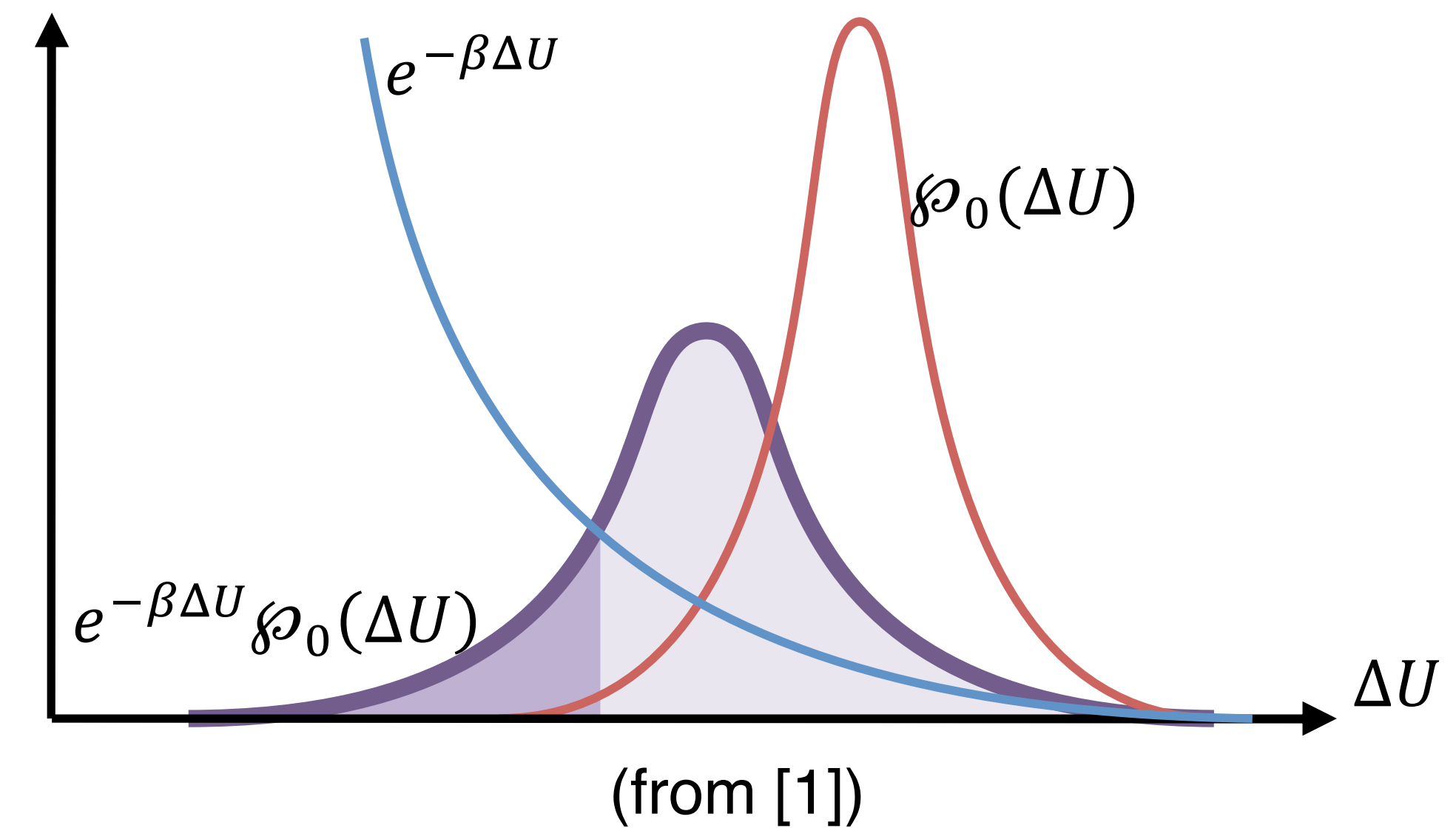
- Using the definition of $\rho_s(r^N)$, $\beta(A_1 - A_0) = -\ln \int \rho_0(r^N) e^{-\beta \Delta U(r^N)} dr^N$.
- The Zwanzig relation [3] is
 - $\beta(A_1 - A_0) = -\ln \langle e^{-\beta \Delta U} \rangle_0$ in a simpler notation.
 - $\beta(A_1 - A_0) = -\ln \langle e^{\beta \Delta U} \rangle_1$ can be derived with similar steps
- This shows us that
 - The free energy difference can be computed based on an average over configurations taken from one of the states of interest
 - We can generate these configurations with MC or MD
 - The free energy comes from evaluating the energies of these configurations in both potentials U_0 and U_1 , and taking an appropriate average of the energy difference

The Zwanzig Relation: Limitations

- In terms of an integral over the distribution of ΔU (instead of over $\rho_o(r^N)$) the Zwanzig relation is,

$$\beta(A_1 - A_0) = -\ln \int e^{-\beta\Delta U} \rho_0(\Delta U) d\Delta U.$$

- Sampling is from the red curve
- Accurate estimation requires the purple curve
- The calculation will not be accurate if U_0 and U_1 are very different!
- Potential energies will be different if the states access different parts of configuration space



Other ways to calculate ΔG

- Multistate Bennett Acceptance Ratio (MBAR) [5]
 - estimates free energies and thermodynamic expectations from a series of states
 - extension of Bennett Acceptance Ratio (BAR) [4], which uses data from two states
 - Proven to be statistically optimal
- Thermodynamic integration is based on the fundamental theorem of calculus, integrating one the derivative of the free energy with respect to a parameter
- All of the methods require thermodynamic states with configuration space overlap, meaning that
 - similar configurations have similar energies
 - the most relevant configuration space is similar

References

- [1] Nguyen, T. H.; Minh, D. D. L. Intermediate Thermodynamic States Contribute Equally to Free Energy Convergence: A Demonstration with Replica Exchange. *Journal of Chemical Theory and Computation* 2016, 12 (5), 2154–2161. <https://doi.org/10.1021/acs.jctc.6b00060>.
- [2] Many parts of this module were adapted from a lecture by David Mobley (https://github.com/MobleyLab/drug-computing/tree/master/uci-pharmsci/lectures/free_energy_basics) under the CC BY 4.0 license. The lecture is part of the Drug Discovery Computing Techniques course (PharmSci 175/275) at UC Irvine.
- [3] Zwanzig, R. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *Journal of Chemical Physics* 1954, 22 (8), 1420.
- [4] Bennett, C. H. Efficient Estimation of Free-Energy Differences from Monte Carlo Data. *Journal of Computational Physics* 1976, 22 (2), 245–268.
- [5] Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *Journal of Chemical Physics* 2008, 129 (12), 124105.