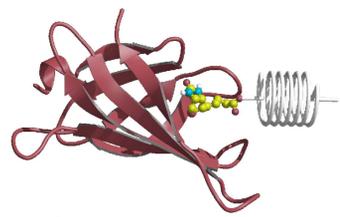# Determining Free Energy Differences Through Variational Morphing

Martin Reinhardt, Helmut Grubmüller
Department of Theoretical and Computational Biophysics
Max Planck Institute for Biophysical Chemistry, Göttingen
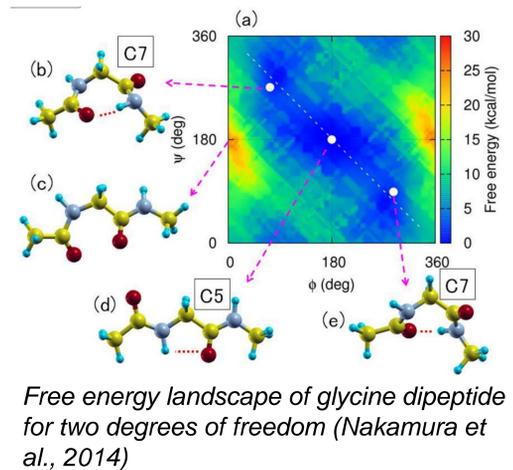martin.reinhardt@mpibpc.mpg.de

Abstract

The Free Energy Perturbation (FEP) method is widely used to estimate the free energy difference between two given states of a many-body system, such as a macromolecule, where Molecular Dynamics or Monte Carlo simulations are typically used for sampling. Because insufficient phase space overlap of the two states can lead to large biases in the results, the calculation is commonly split up into several smaller steps using intermediate states defined along a so-called morphing path - typically a linear interpolation between the start and the end states' Hamiltonians. Here, using FEP, we develop an optimization scheme to find the path that yields the most accurate free energies from the much more general class of paths that includes also non-linear interpolations. For a broad range of one-dimensional test systems, we show that the free energy sampling error can be decreased compared to linear interpolations by more than 50 %. Further, correlations arising from using the same sample in one state to evaluate differences to more than one other state can be taken into account. Finally, we show that the optimization is still valid for a high number of intermediate states with very sparse sampling for each, enabling further optimization by trading off sampling versus number of intermediate states.
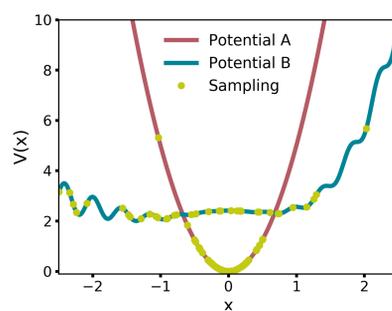
The Problem of Free Energies

- **Free Energy Differences:** Drive (almost) everything:
  - Conformational Changes
  - Protein folding
  - Binding processes
- Definition microscopic: $G = -k_B T \ln \int_{-\infty}^{\infty} e^{-H(\mathbf{x})/k_B T} d\mathbf{x}$

  $H(\mathbf{x})$: Hamiltonian of state $\mathbf{x}$

- Macroscopic: $G = U - TS$

- Problem: often **high dimensional**, cannot integrate over full phase space
- But for two systems: often only **difference** required

Free energy landscape of glycine dipeptide for two degrees of freedom (Nakamura et al., 2014)

Determining Free Energy Differences

- **Basic Approach** (Robert Zwanzig 1954)[1]:
  - Create sample in system A, e.g. through Molecular Dynamics Simulations
  - Evaluate Hamiltonians in A and B and take average of exponentially weighted difference
  - In many steps called **Free Energy Perturbation (FEP)**

$$\Delta G = -\frac{1}{\beta} \ln \left\langle \exp\left(-\beta\left[H_B(\mathbf{x},\mathbf{p}) - H_A(\mathbf{x},\mathbf{p})\right]\right)\right\rangle_A$$
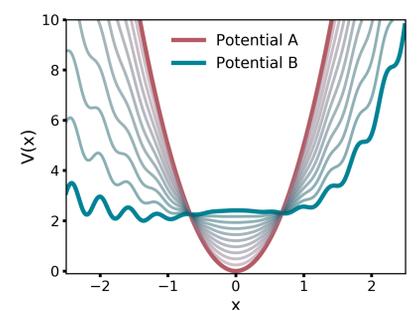

- Results **deviate** from true value:
  - Finite sampling: often do not sample full significant phase space
  - Bias introduced (Jensen inequality):

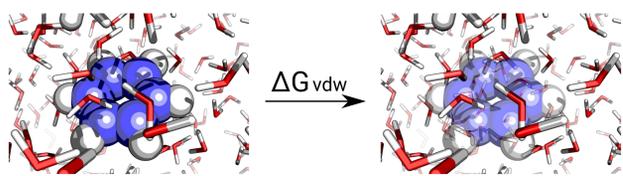 $$\ln\langle e^{-\Delta V}\rangle \neq \langle \ln \frac{1}{N}\sum_i e^{-\Delta V(x_i)}\rangle$$

- **Improvement**: Use **intermediate** steps. Sum up small free energy differences to total:
  $$\Delta G = \sum_{i=1}^{N_{int}+1}\Delta G_i$$
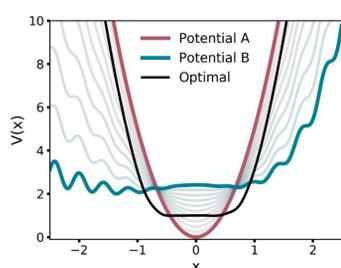

Alchemical Transformations

- When using **Molecular Dynamics** Simulations:
  - Intermediate states are often **"unphysical"** (e.g. an atom has parameters in between those of carbon and oxygen)
  - Hence called **alchemical**, we "transform" one molecule into another on the way
  - Gives **better statistical properties** than physical intermediates

Finding the Minimum Variance Pathway

- So far, modelers are mostly restricted to linear interpolations[2]:
  $$H_{int}(\mathbf{x}) = (1-\lambda)H_A(\mathbf{x}) + \lambda H_B(\mathbf{x}), \quad 0 \leq \lambda \leq 1$$
- However, only requirement: Start at $H_A(\mathbf{x})$ and end at $H_B(\mathbf{x})$, **other options** are also possible
- Optimize overall accuracy: $\sigma_{lin}^2 = \langle(\Delta G_{est} - \Delta G)^2\rangle_{est}$
- Depends on the pathway
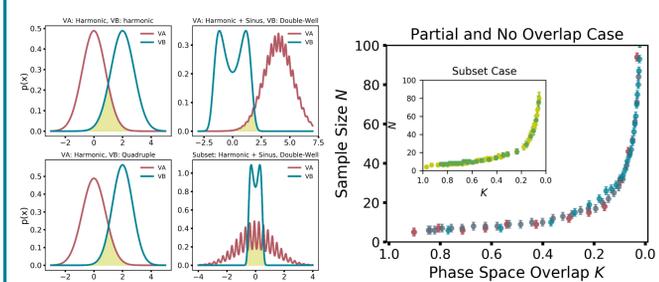- Derivable through variational calculus

- Only optimizable if rough initial free energy estimate exists, which can be achieved through iterative procedures

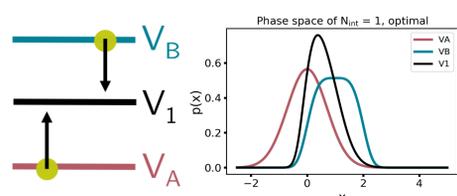$$K = \int_{-\infty}^{+\infty} dx \, \min(p_A(x), p_B(x))$$

Phase Space Overlap

- Accuracy of free energy estimate highly dependent on phase space overlap
- Optimized solution valid down to very small sample sizes
- Validity range follows universal law only dependent on phase space overlap and not on the individual shapes of the systems
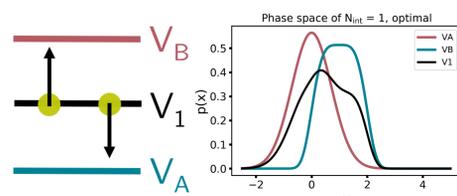

Results and Numerical Validation

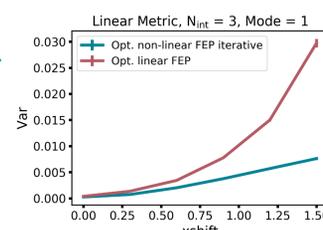Sampling in the end states - post-processing possible

$$V_1(x) = \ln\left(e^{V_A(x)} + e^{V_B(x)}\right)$$

Sampling in intermediate state - required already during simulation

$$V_1(x) = -0.5\ln\left(e^{-2V_A(x)} + e^{-2V_B(x)}\right)$$

- Generalized for N steps
- Gives much higher accuracy of free energy estimate, e.g. $N=3$:

Conclusion & Outlook

- Optimization of morphing path is analytically solvable if initial free energy estimate is provided through iterative methods
- Variance minimizing path is a non-linear function of the start and end states Hamiltonians
- Can improve accuracy by up to a factor three

Next Steps

- Implement non-linear pathway into Molecular Dynamics Software Package
- Expand theory by accounting for correlations between sample points

References

1. Zwanzig, Robert, High-Temperature Equation of State by a Perturbation Method. I. Nonpolar, *The Journal of Chemical Physics (1954)*, Vol 22, Issue 8, 1420-1426
2. Shenfeld, D. et al., Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics (209)*, Vol 80, Issue 4

Acknowledgements