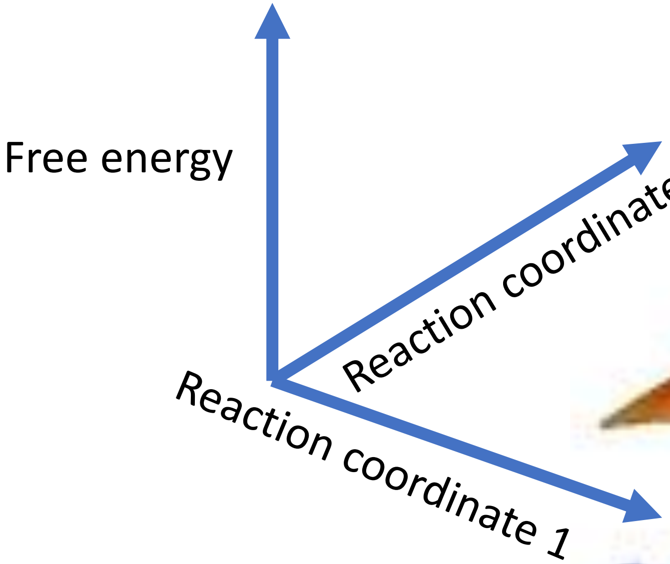
CHEM-E4115
Computational Chemistry I (5op)
2nd part: molecular modelling

Free energy estimation methods

Free energy: thermodynamic potential measuring the maximum reversible work by the thermodynamic system

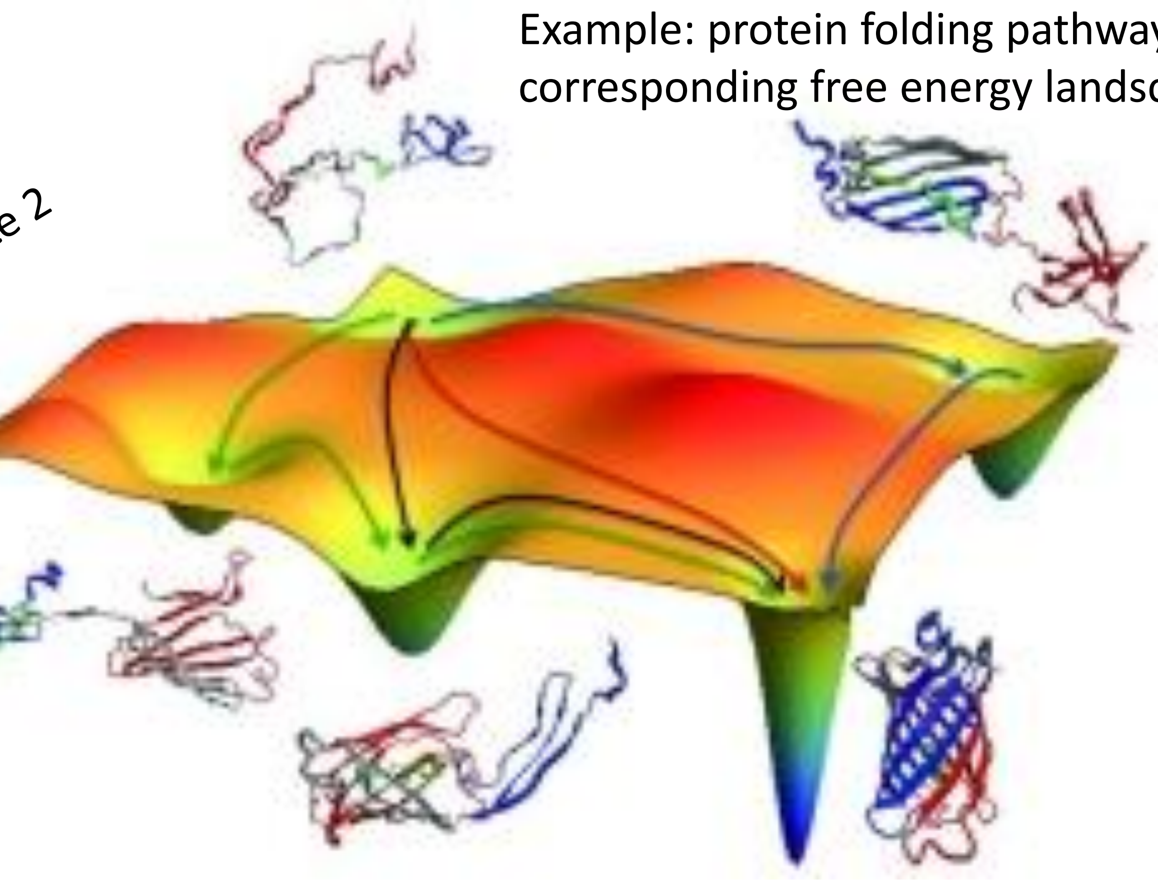


Free energy

Reaction coordinate 2

Reaction coordinate 1

Example: protein folding pathways on the corresponding free energy landscape



Picture: https://cen.acs.org/articles/90/i29/Simulations-Peg-Protein-Folding.html

Typically in chemistry / biosciences simulations
Gibbs free energy (NPT ensemble free energy)
G=H-TS

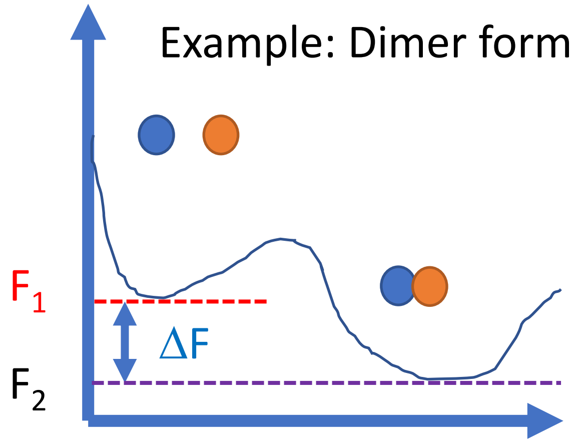
Helmholtz free energy (NVT ensemble free energy)
F=U-TS

Grand potential (μVT ensemble free energy)
Φ=U-TS-μN

Why bother with thermodynamics and free energy in molecular modelling course?

- Probability of a given (macro) state of a thermodynamic system $p \propto e^{-E/k_B T}$
- If one is able to estimate this, the distribution of molecular conformations to different states in the molecular system is known because the probabilities of the states are known.
- Very useful for understanding e.g. partitioning, solubility, conformation transitions, molecular level binding, adsorption, phase changes, …
- Free energy estimation methods are perhaps the technologically most significant advanced molecular modelling approach currently; important in the field as free energy tells what the molecular system will do
- Most interest on Gibbs free energy G, Helmholtz free energy F also easily achievable by analogous approaches

p is the probability of a given (macro) state
F is the (Helmholtz) free energy
k_B is the Boltzmann constant
T is the temperature



ΔF determines the fraction of the particles that is in dimers vs those in monomers

Free energy calculations: practicalities

- Calculation of absolute free energy value possible only if 1) number of microstates 2) absolute enthalpy (NPT ensemble)/ internal energy (NVT ensemble) were known. In practical reality, this is never the case.
- However relative free energies (free energy differences) can be estimated using several different computational techniques

Statistical mechanics related to free energy calculations in molecular modelling

- Hamiltonian (total classical mechanics energy of molecular system)

$$H(\mathbf{p},\mathbf{r}) = \sum_{i=1}^{N} \mathbf{p}_i^2/2m_i + V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$$

p_i momenta vectors
m_i mass
r_i position vectors
summation i over all particles in the system

- Partition function $Q = \sum_j e^{-H_j(\boldsymbol{p},\boldsymbol{r})/k_B T}$, where summation j over all possible system configurations (states)
- Probability of state j $p_j = \frac{e^{-H_j(\boldsymbol{p},\boldsymbol{r})/k_B T}}{Q}$

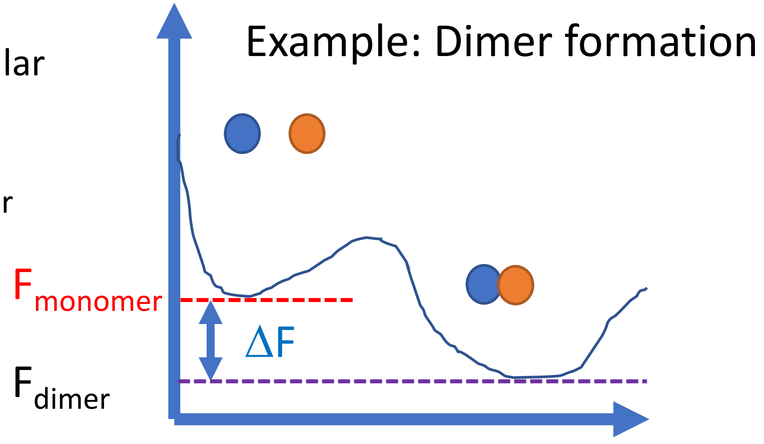
- Average internal energy corresponds to the expectation value of the total classical mechanics energy here

$$U = \langle H \rangle = \sum_j p_j H_j = Q^{-1} \sum_j H_j \, e^{-H_j(\boldsymbol{p},\boldsymbol{r})/k_B T}$$

- Helmholtz free energy in terms of the partition function $F = U - TS = -\,k_B \mathrm{T} \ln \mathrm{Q}$

Free energy difference from state probability (direct counting)

- Most simple way to estimate free energy difference between two states in a molecular simulation is comparing the occurrence frequency of the two states
- Dimer formation example
  - Molecular dynamics or Monte Carlo simulation predicts N monomer, M dimer, and X other realizations in a simulation run.
  - Probability of each of these is
  - $p_{monomer} = \frac{N}{N+M+X}$
  - $p_{dimer} = \frac{M}{N+M+X}$
  - $p_{other} = \frac{X}{N+M+X}$
  - $\frac{p_{monomer}}{p_{dimer}} = \frac{N}{M} = \frac{e^{-F_{monomer}/k_B T}}{e^{-F_{dimer}/k_B T}} = e^{-\Delta F/k_B T}$
- In practise, sampling rarely good enough (=long enough simulation) to get reliable probability weights of the compared states -> free energy differences have large uncertainties
- If used in practise for molecular system, 1) choose a measure characterizing variation between relevant states 2) calculate the occurrence frequency of each possible realization of that measure 3) turn into probabilities and free energy landscape



Example: Dimer formation

ΔF determines the fraction of the particles that is in dimers vs those in monomers

$$\boxed{\frac{p_{monomer}}{p_{dimer}} = \frac{N}{M} = e^{-\Delta F/k_B T}}$$

Free energy difference from integration methods

Basic idea of all these methods: determine free energy difference from the integral of work required to go from the initial state to final state via a reversible path.

Integration over temperature difference

$$F = U - TS = -k_B \mathrm{TlnQ}$$

$$F/T = -k_B \mathrm{lnQ}$$

$$\frac{\partial(F/T)}{\partial(1/T)} = -k_B \frac{\partial \mathrm{lnQ}}{\partial\left(\frac{1}{T}\right)} = -k_B \frac{\partial \ln\left(\sum_j e^{\frac{-H_j(\boldsymbol{p},\boldsymbol{r})}{k_B T}}\right)}{\partial\left(\frac{1}{T}\right)}$$

$$= \frac{\sum_j H_j e^{-H_j(\boldsymbol{p},\boldsymbol{r})/k_B T}}{Q} = \sum_j p_j H_j = \langle H \rangle$$

$$\frac{\partial(F/T)}{\partial(1/T)} = \langle H \rangle$$

$$\Rightarrow \partial(F/T) = \langle H \rangle\, \partial(1/T)$$

$$\Rightarrow \int_{T_A}^{T_B} F d(1/T) = \int_{T_A}^{T_B} \langle H \rangle d(1/T)$$

$$F(T_{\mathrm{B}})/T_{\mathrm{B}} - F(T_A)/T_A = \int_{1/T_{\mathrm{A}}}^{1/T_{\mathrm{B}}} \langle H \rangle d(1/T)$$

Practical calculation: Run long molecular simulations in several different temperatures between T_A and T_B. Determine the ensemble average total energies <H> for each temperature. Integral calculated numerically.

Note: formulation for NVT ensemble (Helmholtz free energy)

Free energy difference from integration methods

Basic idea of all these methods: determine free energy difference from the integral of work required to go from the initial state to final state via a reversible path.

Pressure integration

$F = U - TS$

$$dF = dU - TdS - SdT$$

$$= \left(TdS - pdV + \sum_j \mu_j dN_j \right) - TdS - SdT$$

$$dF = -SdT - \mathrm{pdV} + \sum_j \mu_j dN_j$$

$$p = -\left(\frac{dF}{dV}\right)_{T,N}$$

$$F(V_B) - F(V_A) = -\int_{V_A}^{V_B} PdV.$$

Practical calculation: Run long molecular simulations in several different system volumes between V_A and V_B at constant temperature. Determine the ensemble average pressure for each volume. Integral calculated numerically.

Note: formulation for NVT ensemble (Helmholtz free energy)

Free energy difference from integration methods

Basic idea of all these methods: determine free energy difference from the integral of work required to go from the initial state to final state via a reversible path.

Thermodynamic integration

Generalized form of the two previous methods. Instead of temperature or volume, a coupling parameter λ used as integration variable between states A and B

$$H(\boldsymbol{p},\boldsymbol{r}) = H(\boldsymbol{p},\boldsymbol{r},\lambda)$$

$$H(\boldsymbol{p},\boldsymbol{r},\lambda_A) = H_A(\boldsymbol{p},\boldsymbol{r})$$

$$H(\boldsymbol{p},\boldsymbol{r},\lambda_B) = H_B(\boldsymbol{p},\boldsymbol{r})$$

$$\frac{dF(\lambda)}{d\lambda} = \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda$$

$$F(\lambda_\mathrm{B}) - F(\lambda_\mathrm{A}) = \int_{\lambda_\mathrm{A}}^{\lambda_\mathrm{B}} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda$$

Practical calculation: Run long molecular simulations in several different system λ values between λ_A and λ_B.

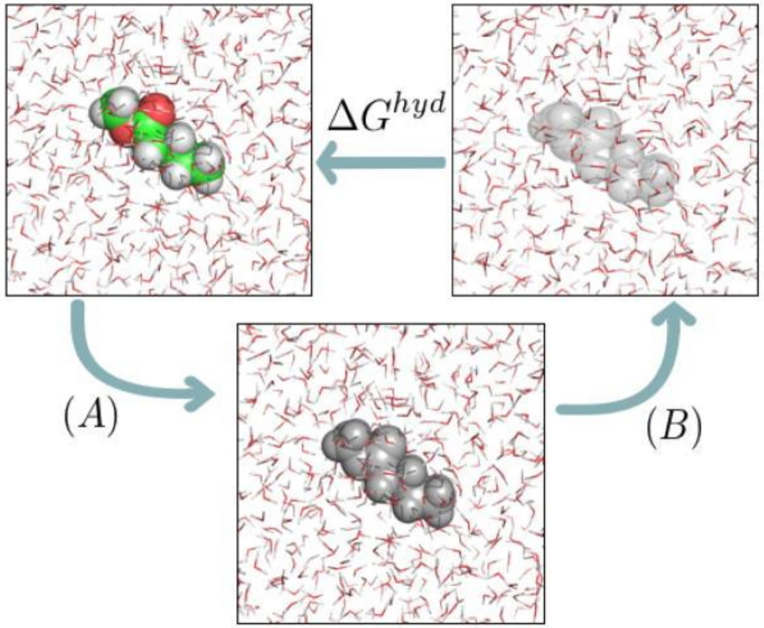
Determine the ensemble average $\left\langle \frac{dH(\lambda)}{d\lambda} \right\rangle$ for each λ.

Integral calculated numerically.

Note that λ Does not need to be a physical pathway. It can be, for example, degree of folding of protein or “ghosting” in or out a particle into the system. It can also be a physical coordinate; there is no restriction as long as a continuous path can be described.

Very common method!

Example of solvation energy calculation protocol via thermodynamic integration



Thermodynamic cycle used to calculate solvation free energies. In the states of (*A*) the charge-charge interactions between the solute and its environment are progressively turned off (l scales interactions from 1 to 0). In states of (*B*), the dispersion interactions between solute and water are progressively turned off.

Colored atoms have electrostatic and nonpolar interactions with the environment; gray atoms retain only nonpolar interactions; transparent atoms have no interactions with their environment (=solute in vacuum).

Figure: J Chem Eng Data. 2017 May 11; 62(5): 1559–1569.

Free energy difference from perturbation

Basic idea: instead of integration, calculate the work through perturbation along the chosen coupling variable.

$$\frac{dF(\lambda)}{d\lambda} = [F(\lambda + \Delta\lambda) - F(\lambda)]/\Delta\lambda,$$

$$\frac{dF(\lambda)}{d\lambda} = -k_B T \ln\left[\langle \exp\{-(H(\lambda + \Delta\lambda) - H(\lambda))/k_B T\}\rangle_\lambda\right]/\Delta\lambda.$$

$$F(\lambda_B) - F(\lambda_A) = \sum_{\lambda=\lambda_A}^{\lambda_B - \Delta\lambda} -k_B T \ln\left[\langle \exp\{-(H(\lambda + \Delta\lambda) - H(\lambda))/k_B T\}\rangle_\lambda\right]$$

Practical calculation: Run long molecular simulations in several different system λ values between λ_A and λ_B that are a finite difference step $\Delta\lambda$ apart. Determine the ensemble average <H> for each step. Calculate logarithm.

Similar to thermodynamic integration, note that λ Does not need to be a physical pathway.

Potential of mean force (PMF)

- Free energy vs. a chosen reaction coordinate
- Potential of mean force is the potential corresponding to the measured mean force along a chosen reaction coordinate
- Usually uniform enough reaction coordinate sampling enforced by pull or constraint force. This external constraint force is “removed” from the force data of the simulation and then the potential corresponding to the remainder mean force calculated to determine its “potential”
- Outcome: free energy landscape along the chosen reaction coordinate
- Umbrella sampling is a special case of PMF calculation

Example - moving methanol across a lipid bilayer

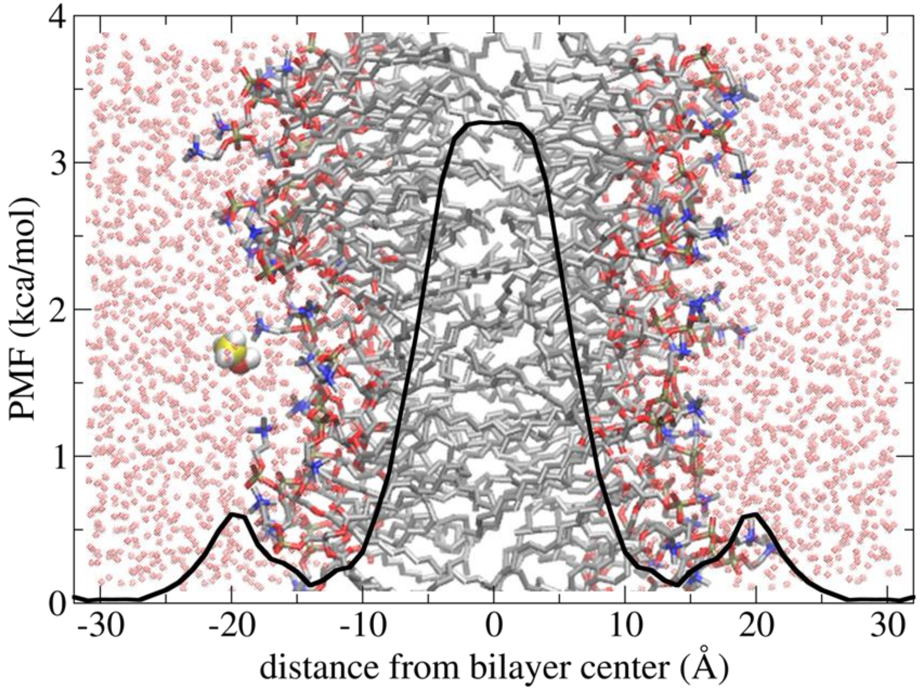


Figure: Amber tutorials
https://ambermd.org/tutorials/FreeEnergy.php

Umbrella sampling