8.2: Free-energy Perturbation Theory

We begin our treatment of free energy differences by examining the problem of transforming a system from one thermodynamic state to another. Let these states be denoted generically as $\mathcal{A}$ and $\mathcal{B}$. At the microscopic level, these two states are characterized by potential energy functions $U_{\mathcal{A}}(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ and $U_{\mathcal{B}}(\mathbf{r}_1, \ldots, \mathbf{r}_N)$. For example, in a drug-binding study, the state $\mathcal{A}$ might correspond to the unbound ligand and enzyme, while $\mathcal{B}$ would correspond to the bound complex. In this case, the potential $U_{\mathcal{A}}$ would exclude all interactions between the enzyme and the ligand and the enzyme, whereas they would be included in the potential $U_{\mathcal{B}}$.

The Helmholtz free energy difference between the states $\mathcal{A}$ and $\mathcal{B}$ is simply $A_{\mathcal{AB}} = A_{\mathcal{B}} - A_{\mathcal{A}}$. The two free energies $A_{\mathcal{A}}$ and $A_{\mathcal{B}}$ are given in terms of their respective canonical partition functions $Q_{\mathcal{A}}$ and $Q_{\mathcal{B}}$, respectively by $A_{\mathcal{A}} = -kT \ln Q_{\mathcal{A}}$ and $A_{\mathcal{B}} = -kT \ln Q_{\mathcal{B}}$, where

$$Q_{\mathcal{A}}(N,V,T) = C_N \int d^N\mathbf{p}\, d^N\mathbf{r}\, exp\left\{-\beta\left[\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i} + U_{\mathcal{A}}(\mathbf{r}_1,\ldots,\mathbf{r}_N)\right]\right\} \tag{8.2.1}$$

$$= \frac{Z_{\mathcal{A}}(N,V,T)}{N!\lambda^{3N}} \tag{8.2.2}$$

$$Q_{\mathcal{B}}(N,V,T) = C_N \int d^N\mathbf{p}\, d^N\mathbf{r}\, exp\left\{-\beta\left[\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i} + U_{\mathcal{B}}(\mathbf{r}_1,\ldots,\mathbf{r}_N)\right]\right\} \tag{8.2.3}$$

$$= \frac{Z_{\mathcal{B}}(N,V,T)}{N!\lambda^{3N}} \tag{8.2.4}$$

The free energy difference is, therefore,

$$A_{\mathcal{AB}} = A_{\mathcal{B}} - A_{\mathcal{A}} = -kT\ln\left(\frac{Q_{\mathcal{A}}}{Q_{\mathcal{A}}}\right) = -kT\ln\left(\frac{Z_{\mathcal{B}}}{Z_{\mathcal{A}}}\right) \tag{8.2.5}$$

where $Z_{\mathcal{A}}$ and $Z_{\mathcal{B}}$ are the configurational partition functions for states $\mathcal{A}$ and $\mathcal{B}$, respectively,

$$Z_{\mathcal{A}} = \int d^N\mathbf{r}\, e^{-\beta U_{\mathcal{A}}(\mathbf{r}_1,\ldots,\mathbf{r}_N)} \tag{8.2.6}$$

$$Z_{\mathcal{B}} = \int d^N\mathbf{r}\, e^{-\beta U_{\mathcal{B}}(\mathbf{r}_1,\ldots,\mathbf{r}_N)} \tag{8.2.7}$$

The ratio of full partition functions $Q_{\mathcal{B}}/Q_{\mathcal{A}}$ reduces to the ratio of configurational partition functions $Z_{\mathcal{B}}/Z_{\mathcal{A}}$ because the momentum integrations in the former cancel out of the ratio.

Equation 8.2.5 is difficult to implement in practice because in any numerical calculation via either molecular dynamics or Monte Carlo, we do not have direct access to the partition function only averages of phase-space functions corresponding to physical observables. However, if we are willing to extend the class of phase-space functions whose averages we seek to functions that do not necessarily correspond to direct observables, then the ratio of configurational partition functions can be manipulated to be in the form of such an average. Consider inserting unity into the expression for $Z_{\mathcal{B}}$ as follows:

$$Z_{\mathcal{B}} = \int d^N\mathbf{r}\, e^{-\beta U_{\mathcal{B}}(\mathbf{r}_1,\ldots,\mathbf{r}_N)} \tag{8.2.8}$$

$$= \int d^N\mathbf{r}\, e^{-\beta U_{\mathcal{B}}(\mathbf{r}_1,\ldots,\mathbf{r}_N)} e^{-\beta U_{\mathcal{A}}(\mathbf{r}_1,\ldots,\mathbf{r}_N)} e^{\beta U_{\mathcal{A}}(\mathbf{r}_1,\ldots,\mathbf{r}_N)} \tag{8.2.9}$$

$$= \int d^N\mathbf{r}\, e^{-\beta U_{\mathcal{A}}(\mathbf{r}_1,\ldots,\mathbf{r}_N)} e^{-\beta(U_{\mathcal{B}}(\mathbf{r}_1,\ldots,\mathbf{r}_N) - U_{\mathcal{A}}(\mathbf{r}_1,\ldots,\mathbf{r}_N))} \tag{8.2.10}$$

If we now take the ratio $\frac{Z_{\mathcal{B}}}{Z_{\mathcal{A}}}$, we find

$$\frac{Z_{\mathcal{B}}}{Z_{\mathcal{A}}} = \frac{1}{Z_{\mathcal{A}}} \int d^N \mathbf{r}\, e^{-\beta U_{\mathcal{A}}(\mathbf{r}_1,\ldots,\mathbf{r}_N)} e^{-\beta(U_{\mathcal{B}}(\mathbf{r}_1,\ldots,\mathbf{r}_N)-U_{\mathcal{A}}(\mathbf{r}_1,\ldots,\mathbf{r}_N))} \tag{8.2.11}$$

$$= \left\langle e^{-\beta(U_{\mathcal{B}}(\mathbf{r}_1,\ldots,\mathbf{r}_N)-U_{\mathcal{A}}(\mathbf{r}_1,\ldots,\mathbf{r}_N))} \right\rangle_{\mathcal{A}} \tag{8.2.12}$$

where the notation $\langle \cdots \rangle_{\mathcal{A}}$ indicates an average taken with respect to the canonical configurational distribution of the state $\mathcal{A}$. Substituting Equation 8.2.12 into Equation 8.2.5, we find

$$A_{\mathcal{AB}} = -kT \ln \left\langle e^{-\beta(U_{\mathcal{B}}-U_{\mathcal{A}})} \right\rangle_{\mathcal{A}} \tag{8.2.13}$$

Equation 8.2.13 is known as the **free-energy perturbation formula**; it should be reminiscent of the thermodynamic perturbation formula used to derive the van der Waals equation. Equation 8.2.13 can be interpreted as follows: We start with microstates $\{\mathbf{r}_1, \ldots, \mathbf{r}_N\}$ selected from the canonical ensemble of state $\mathcal{A}$ and use these to compute $Z_{\mathcal{B}}$ by placing them in the state $\mathcal{B}$ by simply changing the potential energy from $U_{\mathcal{A}}$ to $U_{\mathcal{B}}$. In so doing, we need to "unbias" our choice to sample the configurations from the canonical distribution of state $\mathcal{A}$ $\mathcal{A}$ by removing the weight factor $exp(-\beta U_{\mathcal{A}})$ from which the microstates are sample and reweighting the states by the factor $exp(-\beta U_{\mathcal{B}})$ corresponding to state $\mathcal{B}$. This leads to Equation 8.2.12. The difficulty with this approach is that the microstates corresponding to the canonical distribution of state $\mathcal{A}$ may not be states of high probability in the canonical distribution of state $\mathcal{B}$. If this is the case, then the potential enegy difference $U_{\mathcal{B}} - U_{\mathcal{A}}$ will be large, he exponential factor $exp\left[-\beta(U_{\mathcal{B}} - U_{\mathcal{A}})\right]$ will be negligibly small, and the free energy difference will be very slow to converge in an actual simulation. For this reason, it is clear that the free-energy perturbation formula is only useful for cases in which the two states $\mathcal{A}$ and $\mathcal{B}$ are not that different from each other.

If $U_{\mathcal{B}}$ is not a small perturbation to $U_{\mathcal{A}}$, then the free-energy perturbation formula can still be salvaged by introducing a set of $M-2$ intermediate states with potentials $U_\alpha(r_1, \cdots, r_N)$, where $\alpha = 1, \cdots, M$, $\alpha = 1$ corresponds to the state $\mathcal{A}$ and $\alpha = M$ corresponds to the state $\mathcal{B}$. Let $\Delta U_{\alpha,\alpha+1} = U_{\alpha+1} - U_\alpha$. We can now imagine transforming the system from state $\mathcal{A}$ to state $\mathcal{B}$ by passing through these intermediate states and computing the average of $\Delta U_{\alpha,\alpha+1}$ in the state α. Applying the free-energy perturbation formula to this protocol yields the free-energy difference as

$$A_{\mathcal{AB}} = -kT \sum_{\alpha=1}^{M-1} \ln \left\langle e^{-\beta \Delta U_{\alpha,\alpha+1}} \right\rangle_\alpha \tag{8.2.14}$$

where $\langle \cdots \rangle_\alpha$ means an average taken over the distribution $exp(-\beta U_a)$. The key to applying Equation 8.2.14 is choosing the intermediate states so as to achieve sufficient overlap between the intermediate states without requiring a large number of them, i.e. choosing the thermodynamic path between states $\mathcal{A}$ and $\mathcal{B}$ effectively.

This page titled 8.2: Free-energy Perturbation Theory is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Mark Tuckerman.