The WHAM Equations in 1D and 2D

John H. Hymel

January 2023

1 Free Sampling

In almost any practical case, the underlying potential energy surface of a system is unknown and needs to be determine/computed. Thankfully, two giants of statistically mechanics, Ludwig Boltzmann and Josiah Willard Gibbs, in the late 19^{th} century, determined that systems in thermal equilibrium behave according to a Boltzmann distribution. Where the probability of finding the system in some state i is

$$P_i = \frac{\exp(\frac{-A_i}{k_b T})}{Q} \tag{1}$$

Where A_i is the Helmholtz free energy corresponding to state i, k_b is Boltzmann's constant, and Q is the canonical partition function. From this, the ration of probabilities between two states is given by

$$\frac{P_i}{P_0} = \exp(\frac{-A_i - A_0}{k_b T}) \tag{2}$$

If we assume the energy of our reference state, A_0 is zero, then we can define relative free energies, A_i, as

$$A_i = -k_b T ln(P_i) \tag{3}$$

Using this equation, the relative energy between states of a system can be computed if the probability distribution which connects them can be determined. An accurate probability distribution can be computed by performing molecular dynamics simulations and logging the configurations sampled therein. In most cases though, energy barriers exist between interesting states of a system, making it difficult to freely sample a probability distribution which connects important states without running very long simulations (which is often intractable).

2 Umbrella Sampling and the Weighted Histogram Analysis Method

While straightforward, freely sampling the system is insufficient in many cases as it requires an intractable amount of computer time to sample larger barriers. In order to account for this, many enhanced/biased sampling methods have been developed to force a system to sample higher energy (lower probability) regions of phase space. By sampling a system using a known bias, it is possible to back out the unbiased probability distribution and compute the relative free energy of high energy transitions. One of the first biased sampling methods to be developed, umbrella sampling, parallelizes the problem by using a series of simulations, each biased to a particular region of phase space using a harmonic bias/potential of the form

$$W_h(s) = \frac{1}{2} k_h (s - s_h)^2, h = 1, ..., M \tag{4}$$

Where M is the total number of simulations, h is the index of a single simulation, s is some collective variable of the system (this could be a distance, angle, dihedral, ...), s_h is a point along s where the harmonic potential is centered, k is the force constant, and W is the energy of the bias. The result of performing umbrella sampling simulations is a series of M biased probability distributions which need to be recombined in order to recover an unbiased probability distribution of the full surface. This can be done using the weighted histogram analysis method (WHAM). WHAM attempts to solve the following problem

$$A(s) = -k_b T ln(P'(s)) - U'(s) + f \tag{5}$$

Where A(s) is the free energy of the system along s, P'(s) is the biased probability distribution along s, U'(s) is the potential energy of the harmonic bias plus the underlying potential energy along s, and f is some undetermined constant. Using the data sampled from each biased simulation (P' and U'), we want to solve for A. The issue is these undetermined constants, f, of which you have one per umbrella. WHAM works by guessing initial values for the f constants (typically 1's), and then iteratively solving the WHAM equations to self-consistency in f. For 1-dimension, the WHAM equations are

$$P(s) = \frac{\sum_{h=1}^{M} n_h P_h(s)}{\sum_{h=1}^{M} n_h \exp[\beta f_h] \exp[-\beta W_h(s)]} \tag{6}$$

and

$$\exp[-\beta f_h] = \int ds \exp[-\beta W_h(s)] P(s) \tag{7}$$

This procedure amounts to guessing initial values for each f, solving for a trial probability distribution (eqn 6), using that trial probability distribution to solve for new f's (eqn 7), and repeating until the change in f's is less than some threshold.

It is possible to perform umbrella sampling across additional dimensions and solving the WHAM equations in 2D or even 3D. The number of simulations one needs to run increases exponentially with dimension, so 3D WHAM is quite rare and 4D WHAM is unheard of.

3 2D WHAM

For 2D umbrella sampling, a 2D umbrella potential is needed

$$W_h(x, y) = \frac{1}{2}\left[k_h^x (x - x_h)^2 + k_h^y (y - y_h)^2\right], h = 1, ..., M \tag{8}$$

and the 2D WHAM equations are as follows

$$P(x, y) = \frac{\sum_{h=1}^{M} n_h P_h(x, y)}{\sum_{h=1}^{M} n_h \exp[\beta f_h] \exp[-\beta W_h(x, y)]} \tag{9}$$

and

$$\exp[-\beta f_h] = \int ds \exp[-\beta W_h(x, y)] P(x, y) \tag{10}$$