# Umbrella Sampling + WHAM

Henry Bloom, Arnav Brahmasandra

March 6, 2025

## 1  Motivation

Molecular systems often exhibit high free energy barriers that impede adequate sampling of configurational space in standard simulations. Umbrella sampling addresses this by introducing biasing potentials to enhance sampling in targeted regions, stitching the biased distributions together using WHAM. WHAM provides a systematic framework to reconstruct unbiased free energy profiles from biased simulations.

## 2  Theory

### 2.1  Umbrella Sampling

Consider some system with a potential $U(r^N)$. Traditional sampling methods may suffer from poor ergodicity due to certain energy landscapes, particularly if the confifuration space has many high-energy barriers that may trap the trajectory in a local minimum. Introduced by Torrie and Valleau in 1977, umbrella sampling uses bias potentials to encourage sampling of potentially inaccessible states. The biased distributions are then used to reconstruct an unbiased distribution and calculate quantities of interest, such as the free energy of the system.

#### 2.1.1  Reaction Coordinate

We will introduce the notion of a reaction coordinate $q = f(r^N)$, an abstract one-dimensional parameter that unambiguously measures progress along a reaction pathway or transition between states. Some examples of a reaction coordinate, both geometric and non-geometric, include bond distance, bond angle, torsion, dihedral angle, bond order, Hydrogen bonds, and RMSD. The free energy of a system is often considered as a function of reaction coordinate to demonstrate in some schematic form the potential energy profile associated to the reaction.

### 2.1.2 Bias Potentials

Umbrella sampling introduces a window-specific soft restraint biasing potential $W_i(q, s)$, such that the total potential for window $i$ becomes $U_i^{\text{biased}}(r^N) = U(r^N) + W_i(f(r^N))$. Often, the restraint is chosen to be a harmonic potential of the form $W_i(f(r^N), s) = \frac{1}{2}\kappa(f(r^N) - s_i)^2$ centered around a chosen $s_i$.

### 2.1.3 Simulations

In umbrella sampling, run a series of $n$ simulations with different biasing potentials, collecting a biased distribution of values of $q$. Essentially, each simulation can be thought of as generating a histogram of values of $q$ primarily located within a "window" corresponding to the bias potential chosen. The aim following the computation of these biased distributions is to first unbias them, and then to use the unbiased distributions to calculate the free energy of the system.

## 2.2 WHAM

The weighted histogram analysis method (WHAM) is the method for unbiasing the biased distributions that is used in conjunction with umbrella sampling.

### 2.2.1 Unbiasing the Probabilities

We begin with biased probabilities:

$$\tilde{P}(q, s_k) = e^{\beta A_k} \int d^N r^N e^{-\beta U(r^N)} e^{-\beta W(f(r^N), s^{(k)})} \delta(f(r^N) - q)$$

We can then express the free energy due to each individual restraining potential, $A_k$, as follows:

$$e^{-\beta A_k} = \int d^N r^N e^{-\beta U(r^N)} e^{-\beta W(f(r^N), s^{(k)})} = e^{-\beta A_0} \left\langle e^{-\beta W(f(r^N), s^{(k)})} \right\rangle$$

Note the following:
$$e^{-\beta A_0} = \int d^N r^N e^{-\beta U(r^N)} = Z$$

We can use this process to calculate the unbiased potential from the $k$-th window as follows:
$$P_k(q) = e^{-\beta(A_k - A_0)} e^{\beta W(q, s^{(k)})} \tilde{P}(q, s^{(k)})$$

### 2.2.2 Reconstructing the Global Distribution

Our aim is to use the $k$ unbiased distribution estimates from the simulations to construct the global distribution $P(q)$ while minimizing statistical error. We

can set this up as follows:

$$P(q) = \sum_{k=1}^{n} C_k(q) P_k(q)$$

$$\text{subject to} \sum_{k=1}^{n} C_k(q) = 1$$

The aim is to optimize the statistical error of $P(q)$ while satisfying the given constriant.

### 2.2.3 Statistical Error

We first aim to express the statistical error of each biased distribution using its corresponding histogram by the following:

$$\tilde{P}(q, s^{(k)}) \approx \frac{1}{n_k \Delta q} \tilde{H}_k(q)$$

Where $\tilde{H}_k$ is the histogram from the $k$-th simulation, $n_k$ is the number of configurations sampled, and $\Delta q$ is the bin width. The statistical error in this distribution is then expressed as follows:

$$\tilde{\sigma}_k^2 = \epsilon_k(q) \frac{1}{n_k \Delta q} \tilde{H}_k(q)$$

Where $\epsilon_k(q)$ is the error of the sample with respect to the actual biased distribution.

We can propogate this error to estimate the error in the $k$-th unbiased distribution:

$$P_k(q) = e^{-\beta(A_k - A_0)} e^{\beta W(q, s^{(k)})} \tilde{P}(q, s^{(k)})$$

$$\sigma_k^2 = e^{-2\beta(A_k - A_0)} e^{2\beta W(q, s^{(k)})} \tilde{\sigma}_k^2$$

Thus, the total statistical error in the global distribution will be:

$$\sigma^2 = \sum_{k=1}^{n} C_k^2(q) \sigma_k^2$$

## 2.3 Minimizing the Statistical Error

We can use Lagrange multipliers to perform this constrained minimization by instead minimizing the following:

$$\Sigma^2 = \sum_{k=1}^{n} C_k^2(q) \sigma_k^2 - \lambda \left( \sum_{k=1}^{n} C_k(q) - 1 \right)$$

After solving for $\lambda$ and $C_k(q)$ for all $k$ by setting $\frac{\partial \Sigma^2}{\partial C_k(q)} = 0$, we make the simplifying assumptions that $\epsilon_k(q)$ is constant for all $k$ and that the biased histograms $\tilde{H}_k(q)$ are proportionate to a direct biasing of the unbiased global $P(q)$: $\tilde{H}_k(q) \propto e^{\beta(A_k - A_0)} * e^{-\beta W(q, s^{(k)})} * P(q)$. Finally, plugging the $C_k(q)$ coefficients back into the equation for $P(q)$ provides us with a set of equations to solve iteratively:

$$P(q) = \frac{\sum_{k=1}^{n} n_k \tilde{P}_k(q)}{\sum_{k=1}^{n} n_k e^{\beta(A_k - A_0)} e^{-\beta W(q, s^{(k)})}}$$

$$e^{\beta(A_k - A_0)} = \int \mathrm{d}q P(q) e^{-\beta W(q, s^{(k)})}$$

Following the iterative solving of these equations, we can obtain the free energy from $P(q)$ through the following formula:

$$A(q) = -\frac{1}{\beta} \ln P(q) + A_0$$