

# Free Energy Calculations <sup>1</sup>

## Methods Based on Equilibrium and Non-equilibrium Simulations

YE MEI <sup>2</sup>  
PENGFEI LI <sup>3</sup>

June 21, 2017

<sup>1</sup>For internal/noncommercial use only.

<sup>2</sup>samuel.y.mei@gmail.com

<sup>3</sup>lipengfei\_mail@126.com



Dedicated to  
Dr. Bernard Brooks and Dr. Gerhard König.



# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Enhanced Sampling</b>	<b>5</b>
2.1	Replica Exchange Molecular Dynamics . . . . .	6
2.1.1	Temperature-Replica Exchange Molecular Dynamics .	6
2.1.2	Hamiltonian-Replica Exchange Molecular Dynamics .	7
2.2	Umbrella Sampling . . . . .	12
2.3	Metadynamics . . . . .	13
2.4	Orthogonal Space Random Walk . . . . .	14
2.5	Enveloping distribution sampling . . . . .	15
<b>3</b>	<b>Postprocessing</b>	<b>17</b>
3.1	Rigorous Methods . . . . .	17
3.1.1	Thermodynamic Perturbation . . . . .	17
3.1.2	Thermodynamic Integration . . . . .	20
3.1.3	Bennett Acceptance Ratio . . . . .	22
3.1.4	Weighted Histogram Analysis Method . . . . .	25
3.1.5	Multistate Bennett Acceptance Ratio . . . . .	32
3.1.6	Adaptive Biasing Force Method . . . . .	34
3.1.7	Non-Equilibrium Work . . . . .	35
3.2	Approximate Methods . . . . .	38
3.2.1	Molecular Mechanics/Poisson-Boltzmann Surface Area	38
	<b>Appendices</b>	<b>41</b>
<b>A</b>	<b>Optimal Means of Data Set</b>	<b>41</b>
<b>B</b>	<b>Statistical Uncertainty in the Estimator for Correlated Time Series Data</b>	<b>43</b>
<b>C</b>	<b>MBAR returns to BAR When Only Two States Are Considered</b>	<b>45</b>
<b>D</b>	<b>MBAR is a binless form of WHAM</b>	<b>47</b>

**Bibliography****49**

# List of Figures

2.1	.....	6
2.2	.....	12
2.3	.....	13





# List of Tables



# Preface

Should I say something?

## About the companion website

The website<sup>1</sup> for this file contains:

- A link to (freely downloadable) latest version of this document.
- Link to download some implementations of WHAM.
- Other stuff might appear in the near future.

## Acknowledgements

- YM wants to express his special thanks to Dr. Bernard Brooks<sup>2</sup> and Dr. Gerhard König for helping him toddle in this field
- I'll also like to thank Dr. Xiangyu Jia<sup>3</sup>, Ms. Meiting Wang, Ms. Wei Liu and Ms. Fengjiao Liu for many helpful discussions.

Ye Mei

<http://faculty.ecnu.edu.cn/s/960/main.jspy>

---

<sup>1</sup><https://github.com/samuelymei/>

<sup>2</sup><https://www.lobos.nih.gov/cbs/>

<sup>3</sup><https://research.shanghai.nyu.edu/centers-and-institutes/chemistry/people/xiangyu-jia>



# 1

## Introduction

*“Everything should be made as simple as possible but not simpler.”*

– Albert Einstein,

Computer simulations of biological system have made much progress in the past decades.

We are still facing many difficulties in three aspects, i.e. Hamiltonians, sampling efficiency and postprocessing methods.[1]

Recently, there is one special issue focusing on the methodologies of free energy calculations on Journal of Chemical Theory and Computation (Free Energy Calculations: Three Decades of Adventure in Chemistry and Biophysics, Journal of Chemical Theory and Computation, Volume 10, Issue 7, 2014, <http://pubs.acs.org/toc/jctc/10/7>).

There are also two books on this topic you might be interested in:

- Free Energy Calculations: Theory and Applications in Chemistry and Biology, Editors: Christophe Chipot, Andrew Pohorille, ISBN 978-3-540-38448-9, Springer-Verlag Berlin Heidelberg, 2007
- Free Energy Computations: A Mathematical Perspective, Author: Tony Lelievre, Gabriel Stoltz, Mathias Rousset, ISBN-13: 978-1848162471, Imperial College Press, 2010



## 2

# Enhanced Sampling

*“Keep the smart guys around you.”*

– Bernard R. Brooks

From the definition, the free energy of a specific system is dominated by phase space regions with low energy. However, these regions might be separated by high energy barriers. Transitions among these potential energy wells are often hindered by these barriers. According to the Boltzmann’s Law, the probability of a sample  $\mathbf{R}$  being visited is proportional to the Boltzmann’s factor  $\exp(-\beta E(\mathbf{R}))$ , where  $\beta = 1/kT$  is called the inverse temperature.  $k$  is the Boltzmann constant and  $T$  is the temperature. According to some experience, in a 100 ns simulation, the system can overcome a barrier of 10 kT, which is 6 kcal/mol at room temperature (300 K). If the barrier is 3 kT higher, it takes the system about 1  $\mu$ s (10 times longer) in average to go over the barrier. If the barrier height reaches 9 kcal/mol, it takes 10  $\mu$ s at 300 K. And so on. With modern computers, the longest all-atom molecular dynamics simulation for biological systems is probably done by D.E. Shaw, which was on a time scale of 1 ms on a special-purpose computer “Anton”. For most classical molecular dynamics simulations, the time scales are normally several  $\mu$ s to tens of  $\mu$ s. For simulations using expensive Hamiltonians, such as in QM/MM simulations, the time scales that can be reached are usually more than three orders shorter. In order to observe abundant transitions among these energy minima, which is required by free energy calculations, enhanced samplings are often indispensable. As shown in the Boltzmann’s factor, the essential quantity that determines the rate of transitions is  $\beta E$ . In order to accelerate the phase space sampling, we can either increase the temperature or decrease the energy barrier. All the methods shown below can be classified into these two categories.

## 2.1 Replica Exchange Molecular Dynamics

### 2.1.1 Temperature-Replica Exchange Molecular Dynamics

Temperature replica exchange molecular dynamics (T-REMD) is one class of parallel tempering methods developed by Sugita and Okamoto in 1999[2] based on many ideas on the so-called *generalized-ensemble algorithm*. It is an extension of the well-known simulation annealing method. The basic idea of REMD is schematically summarized in Fig. 2.1. In REMD, the system is replicated into  $M$  *non-interacting* copies (replicas). Each replica is coupled to a bath at temperature  $T_m$ , ( $m = 1, \dots, M$ ). At a certain time, the system is at state  $X$ , which can be denoted as  $X = (x_1^{[i(1)]}, \dots, x_M^{[i(M)]}) = (x_{m(1)}^{[1]}, \dots, x_{m(M)}^{[M]})$ . Here, we used  $i$  and  $m$  to label the replica and the temperature respectively. Because the replicas are non-interacting, the weight-factor for a state  $X$  in this generalized ensemble is a direct product of the Boltzmann factors for each replica, i.e.

$$W_{REM}(X) = \prod_{m=1}^M \exp\left(-\beta_m H\left(q^{[i(m)]}, p^{[i(m)]}\right)\right) = \prod_{i=1}^M \exp\left(-\beta_{m(i)} H\left(q^{[i]}, p^{[i]}\right)\right) \quad (2.1)$$

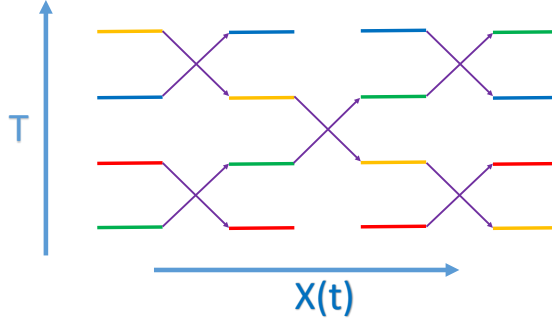


Figure 2.1:

Now, we exchange the temperatures of a pair replicas

$$\begin{cases} x_m^{[i]} \equiv (q^{[i]}, p^{[i]})_m \Rightarrow x_n^{[i]'} \equiv (q^{[i]}, p^{[i]'})_n \\ x_n^{[j]} \equiv (q^{[j]}, p^{[j]})_n \Rightarrow x_m^{[j]'} \equiv (q^{[j]}, p^{[j]'})_m \end{cases}, \quad (2.2)$$

where

$$\begin{cases} p^{[i]'} \equiv \sqrt{\frac{T_n}{T_m}} p^{[i]} \\ p^{[j]'} \equiv \sqrt{\frac{T_m}{T_n}} p^{[j]} \end{cases}. \quad (2.3)$$



In order for this exchange process to converge towards an equilibrium distribution, it is sufficient to impose the detailed balance condition on the transition probability  $w(X \rightarrow X')$ :

$$W_{REM}(X)w(X \rightarrow X') = W_{REM}(X')w(X' \rightarrow X). \quad (2.4)$$

Then we have

$$\begin{aligned} & \frac{w(X \rightarrow X')}{w(X' \rightarrow X)} \\ &= \frac{W_{REM}(X')}{W_{REM}(X)} \\ &= \frac{\exp(-\beta_m H(q^{[j]}, p^{[j]'})) \exp(-\beta_n H(q^{[i]}, p^{[i]'}))}{\exp(-\beta_m H(q^{[i]}, p^{[i]}) \exp(-\beta_n H(q^{[j]}, p^{[j]}))} \\ &= \frac{\exp\{-\beta_m [K(p^{[j]'} + U(q^{[j]}))] - \beta_n [K(p^{[i]'} + U(q^{[i]}))]\}}{\exp\{-\beta_m [K(p^{[i]} + U(q^{[i]}))] - \beta_n [K(p^{[j]} + U(q^{[j]}))]\}} \\ &= \frac{\exp\{-\beta_m [\frac{T_m}{T_n} K(p^{[j]}) + U(q^{[j]})] - \beta_n [\frac{T_n}{T_m} K(p^{[i]}) + U(q^{[i]})]\}}{\exp\{-\beta_m [K(p^{[i]}) + U(q^{[i]})] - \beta_n [K(p^{[j]}) + U(q^{[j]})]\}} \\ &= \frac{\exp\{-\beta_n K(p^{[j]}) - \beta_m K(p^{[i]})\} \exp\{-\beta_m U(q^{[j]}) - \beta_n U(q^{[i]})\}}{\exp\{-\beta_m K(p^{[i]}) - \beta_n K(p^{[j]})\} \exp\{-\beta_m U(q^{[i]}) - \beta_n U(q^{[j]})\}} \\ &= \exp\{-\Delta\}. \end{aligned} \quad (2.5)$$

where  $\Delta = [\beta_n - \beta_m] [U(q^{[i]}) - U(q^{[j]})]$ . It can be seen that the kinetic energy terms are fully canceled out. This can be satisfied by the usual Metropolis criterion:

$$w(X \rightarrow X') \equiv w(x_m^{[i]} | x_n^{[j]}) = \begin{cases} 1, & \text{if } \Delta \leq 0 \\ \exp(-\delta), & \text{if } \Delta > 0 \end{cases} \quad (2.6)$$

After long time simulations, all the replicas have arrived at a global equilibrium. In order to calculate the the free energy or the ensemble average of an operator  $\hat{A}$  at  $T_m$ , we can extract all the snapshots that have a temperature  $T_m$  from  $M$  trajectories, if this temperature was among the  $M$  chosen temperatures. However, the optimal way is to use Weighted Histogram Analysis Method in Section 3.1.4 or the Multistate Bennett Acceptance Ratio method in Section 3.1.5.

### 2.1.2 Hamiltonian-Replica Exchange Molecular Dynamics

Another type of REMD simulation is called Hamiltonian replica exchange molecular dynamics (H-REMD), in which each replicas has its own Hamiltonian is coupled to the same temperature. One example is the H-REMD

simulation for a torsional angle. The  $m$ th replica has a torsional energy term of

$$H_m(\phi) = \lambda(m) \sum_n (V_n/2) (1 + \cos[n\phi - \delta]), \quad (2.7)$$

where  $\lambda$  is a control parameter.  $\lambda(0) = 1$  corresponds to the unbiased state and at  $\lambda(M)$  (usually  $\lambda(M) = 0$ ) the torsional motion of this dihedral angle has a smaller barrier.

Another example of HREMD is pH-REMD, in which each replica is coupled with different pH of the solution. In other words, the chemical potential of hydronium in each replica is different. Therefore, the protonation states (or probability of being protonated or deprotonated) of titratable residues in each replica may differ from those in other replicas. In the simulations, the protonation states of titratable residues have their protonation states alternated according to the Metropolis criterion

$$P = \begin{cases} 1, & \text{if } \Delta G_{P_A \rightarrow P_A H^+} \leq 0 \\ \exp(-\beta \Delta G_{P_A \rightarrow P_A H^+}), & \text{if } \Delta G_{P_A \rightarrow P_A H^+} > 0 \end{cases} \quad (2.8)$$

using Monte Carlo. The derivation of  $\Delta G_{P_A \rightarrow P_A H^+}$  is shown below.

Free energy of molecule A in solution with a concentration  $[A]$  can be written as

$$\Delta G_A = \Delta G_A^0 + \beta^{-1} \ln \frac{[A]}{C_0},$$

in which  $\Delta G_0$  is the free energy of molecule A at the standard state  $C_0$ , i.e. 1 mol/L. The free energy change for a reaction



can be written as

$$\Delta G = \Delta G_C - \Delta G_A - \Delta G_B = \Delta G_0 + \beta^{-1} \ln \frac{[C] C_0}{[A] [B]}.$$

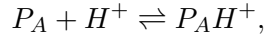
At equilibrium, the free energy change is zero, we have

$$\Delta G_0 = -\beta^{-1} \ln \frac{[C] C_0}{[A] [B]}, \quad (2.9)$$

in which  $[A] [B] / [C] C_0$  is called the dissociation constant  $K_a$ . So,

$$\Delta G_0 = \beta^{-1} \ln K_a. \quad (2.10)$$

Titration of a residue in a real protein can be written as



with

$$K_a = \frac{[P_A][H^+]}{[P_AH^+]C_0}$$

The fraction of the deprotonated species is calculated as

$$\begin{aligned} f_{[P_A]} &= \frac{[P_A]}{[P_A] + [P_AH^+]} \\ &= \frac{1}{1 + \frac{[P_AH^+]}{[P_A]}} \\ &= \frac{1}{1 + \frac{[P_A][H^+]}{C_0K_a[P_A]}} \\ &= \frac{1}{1 + \frac{1}{C_0K_a}[H^+]} \\ &= \frac{1}{1 + \frac{1}{K_a}10^{-pH}} \end{aligned} \quad (2.11)$$

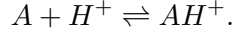
We can check the asymptotic behavior of this equation. At strong acidic condition ( $pH = -\infty$ ),  $f_{\{P_A\}} = 0$ , indicating that the residue is 100 percent protonated. While at an extremely basic condition ( $pH = \infty$ ),  $f_{\{P_A\}} = 1$ . This residue is 100 percent deprotonated. From the Henderson–Hasselbalch (HH) equation, the  $pK_a$  can be determined by the  $pH$  of the state when  $[P_A]/[P_AH^+] = 1$  (the isoelectric point)

$$\begin{aligned} pK_a &= -\log K_a \\ &= -\log \frac{[P_A]}{[P_AH^+]} - \log \frac{[H^+]}{C_0} \\ &= -\log \frac{[P_A]}{[P_AH^+]} + pH. \end{aligned} \quad (2.12)$$

The  $pK_a$  of each residue in a dipeptide has been determined by experiment. However, when this residue is located in a certain protein, its  $pK_a$  is different from that in a dipeptide. The difference is called the  $pK_a$  shift. Instead of measuring the  $pK_a$  for a residue in a protein, we are more interested in calculating/measuring the titration curve, which is the fraction of the deprotonated state as a function of pH. From Eq. 2.11,  $f_{[P_A]}$  can be easily calculated if we know  $K_a$  or equivalently the standard free energy change of protonation in Eq. 2.10. The standard free energy can be calculated from the partition functions as

$$\begin{aligned}
\Delta G_0 &= -\beta^{-1} \ln \frac{Q_{P_A H^+}}{Q_{P_A} Q_{H^+}} \\
&= -\beta^{-1} \ln \frac{\int \exp(-\beta E_{P_A H^+}) dR_H dR_o}{Q_{H^+} \int \exp(-\beta E_{P_A}) dR_o}
\end{aligned}$$

Generally, the absolute value of  $\Delta G_0$  is hardly computable. A relative protonation free energy  $\Delta\Delta G$  is preferred and is more reliable. Theoretically, the reference state can be any state you like. But the protonation free energy of the dipeptide at  $pK_a$  is often used. The reference protonation process can be written as



The reference free energy change is

$$\begin{aligned}
\Delta\Delta G_0 &= \Delta G_0 - \Delta G_0^{ref} \\
&= -\beta^{-1} \ln \frac{\int \int \exp(-\beta E_{P_A H^+}) dR_H dR_o}{Q_{H^+} \int \exp(-\beta E_{P_A}) dR_o} \frac{Q_{H^+} \int \exp(-\beta E_A) dR_o}{\int \int \exp(-\beta E_{A H^+}) dR_H dR_o} \\
&= -\beta^{-1} \ln \frac{\int \int \exp(-\beta E_{P_A H^+}) dR_H dR_o \int \exp(-\beta E_A) dR_o}{\int \exp(-\beta E_{P_A}) dR_o \int \int \exp(-\beta E_{A H^+}) dR_H dR_o} \\
&= -\beta^{-1} \ln \frac{\int \int \exp \left[ -\beta \left( E_{P_A H^+}^{bond} + E_{P_A H^+}^{QM} + E_{P_A H^+}^{ele} \right) \right] dR_H \exp \left( -\beta E_{P_A H^+}^{other} \right) dR_o}{\int \int \exp \left[ -\beta \left( E_{A H^+}^{bond} + E_{A H^+}^{QM} + E_{A H^+}^{ele} \right) \right] dR_H \exp \left( -\beta E_{A H^+}^{other} \right) dR_o} \\
&\quad \cdot \frac{\int \exp(-\beta E_A) dR_o}{\int \exp(-\beta E_{P_A}) dR_o}, \tag{2.13}
\end{aligned}$$

where  $R_H$  is the coordinates of the specific  $H$  atom and the other degrees-of-freedom (DoF) are denoted as  $R_o$ .  $E^{bond}$  and  $E^{ele}$  are the bonded energy and electrostatic interaction energy related to this  $H$  atom, respectively.  $E^{QM}$  is the energy correction that *may* be required if the molecular mechanical Hamiltonian cannot well capture the energy of the system, such as the missing of charge transfer effect. The sum of the remaining energy term is denoted as  $E^{other}$ , which does not explicitly depend on the position of this specific  $H$  atom. Eq. 2.13 is not ready to be computed before some approximations are adopted.

*First*, we assume that the total energy can be well described by the MM Hamiltonians for both the state interested in and the reference state. Therefore,

$$E_{P_A H^+}^{QM} = E_{A H^+}^{QM} = Const,$$

and they can be removed from the integral.

*Second*, the bonded terms involving hydrogen atoms are usually constrained in the simulations. Therefore, the hydrogen atom in question has

only one position and  $E^{bond} = 0$ . Now, the relative protonation free energy can be simplified as

$$\begin{aligned} \Delta\Delta G_0 = & -\beta^{-1} \ln \frac{\int \exp(-\beta E_{P_A H^+}^{ele}) \exp(-\beta E_{P_A H^+}^{other}) dR_O}{\int \exp(-\beta E_{A H^+}^{ele}) \exp(-\beta E_{A H^+}^{other}) dR_O} \\ & \cdot \frac{\int \exp(-\beta E_A) dR_O}{\int \exp(-\beta E_{P_A}) dR_O}. \end{aligned} \quad (2.14)$$

Note that  $E_A = E_{A H^+}^{other}$  and  $E_{P_A} = E_{P_A H^+}^{other}$ , we have

$$\Delta\Delta G_0 = -kT \ln \frac{\int \exp(-\beta E_{P_A H^+}^{ele}) \exp(-\beta E_{P_A}) dR_O}{\int \exp(-\beta E_{P_A}) dR_O} \quad (2.15)$$

$$\cdot \frac{\int \exp(-\beta E_A) dR_O}{\int \exp(-\beta E_{A H^+}^{ele}) \exp(-\beta E_A) dR_O} \quad (2.16)$$

$$\begin{aligned} & = -\beta^{-1} \ln \left\langle \exp(-\beta E_{P_A H^+}^{ele}) \right\rangle_{P_A} \\ & \quad + \beta^{-1} \ln \left\langle \exp(-\beta E_{A H^+}^{ele}) \right\rangle_A \\ & = \Delta G_{P_A H^+}^{ele} - \Delta G_{A H^+}^{ele} \end{aligned} \quad (2.17)$$

Therefore,

$$-\beta^{-1} \ln 10 \cdot pK_a = \Delta G_{P_A H^+}^{ele} - \Delta G_{A H^+}^{ele} - \beta^{-1} \ln 10 \cdot pK_a^{ref}.$$

Using Eq. 2.12, at a certain pH the free energy difference between the deprotonated and the protonated state can be written as

$$\Delta G_{P_A \rightarrow P_A H^+} = \Delta G_{P_A H^+}^{ele} + \beta^{-1} (pH - pK_a^{ref}) \ln 10 - \Delta G_{A H^+}^{ele}.$$

In the above equation,  $\Delta G_{A H^+}^{ele}$  can be obtained from a free energy calculation of the model system by alchemically annihilation of the proton. However,  $\Delta G_{P_A H^+}^{ele}$  is unknown. Approximately, it can be replaced with  $\Delta H_{P_A H^+}^{ele}$  averaged over a few snapshots.[3] In order to accelerate the convergence, this pH-REMD is often coupled with other enhancing method, such as T-REMD[3] and EDS-REMD[4] (see section 2.5).

## 2.2 Umbrella Sampling

Umbrella Sampling method was proposed by Torrie and Valleau in 1977,[5] and is still widely used nowadays. Suppose we are studying a transition process between states such as conversion between two dominant conformations or a chemical reaction, and these two states are separated by a high barrier relative to  $kT$ . Therefore, the transition is a rare event. A schematic representation of the free energy landscape is shown in Fig. 2.2.

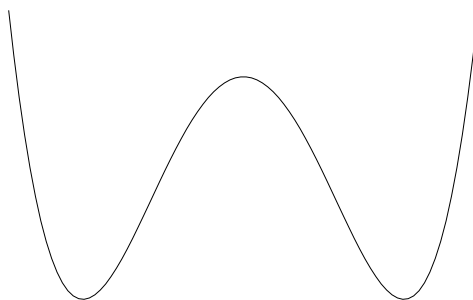


Figure 2.2:

Sometimes, we are interested in not only these two dominant states but also the states in between. Usually, we define a reaction coordinate  $\xi$  and calculate the potential of mean force along this reaction coordinate from the “reactant” to the “product”. However, if we run a simulation with the reaction coordinate initially set to the transition state or on the hill, the system will quickly roll back to the “reactant” or the “product” state in order to reduce the free energy. The consequence is that regions outside the “reactant” and “product” regions cannot be sampled sufficiently to yield accurate free energy profile in a brute force simulation. In order to increase the sampling in these regions, we can modify the potential energy surface by introducing a series of biases as (usually harmonic) functions of  $\xi$  to the system. Each biased simulation is called a *window*. The strengths of the biases should be strong enough to maintain the system in the vicinity of where you are interested in, and also should be weak enough that the system can have significant overlap in two adjacent windows. After all the simulations, the whole region is well sampled and the free energy profile can be generated using Weighted Histogram Analysis Method or the Multistate Bennett Acceptance Ratio method (to be discussed in Section 3.1.4 and 3.1.5).

## 2.3 Metadynamics

Metadynamics was first suggested by Laio and Parrinello in 2002.[6] Imaging you became Doraemon in a dream. You were standing in a valley and were surrounded by high mountains. In most of the time, you are just wandering near the minimum, because your kinetic energy is not enough to climb the mountains. Suddenly, you realize that you can use metadynamics as a magic to escape from the minimum. After each step, you took a bottle of sand out of your miraculous pocket and put the sand under your feet. Then you were lifted inch-by-inch, and the chance of visiting where you had visited became less and less. And you were finally raised up to the top of the mountain and at that moment you was able to climb over that mountain without much effort and fell into another valley. The magic of sand continued, and at last you smoothed the whole area. Because you wrote diaries every day and you have all the records on where you had put the sand and how much sand you had put there. You drew the shape the piled sand according to the record and you flipped it. At that moment, you got the exact shape of the original free energy landscape.

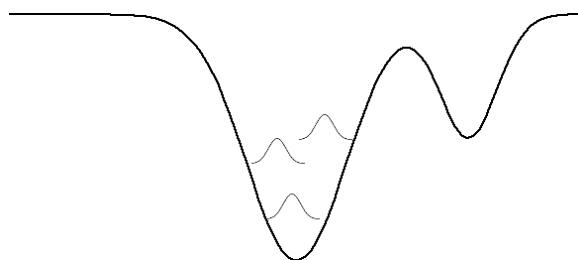


Figure 2.3:

## 2.4 Orthogonal Space Random Walk

The orthogonal space random walk (OSRW) was developed by Yang.[7]



## **2.5 Enveloping distribution sampling**

Enveloping distribution sampling method was first proposed by Christ and van Gunsteren in 2007.[8].



## 3

# Postprocessing

*“The source of mistake is always between the keyboard and the chair. So, check, double check and check again”*

– Gerhard König

## 3.1 Rigorous Methods

### 3.1.1 Thermodynamic Perturbation

Thermodynamic Perturbation (TP), also known as Free Energy Perturbation (FEP), exponential average, or Zwanzig equation was developed by Zwanzig.[9].

A reference system containing N-particles can be described by Hamiltonian  $H_0(\mathbf{x}, \mathbf{p}_x)$ , which is a function of 3N Cartesian coordinates,  $\mathbf{x}$ , and their conjugated momenta,  $\mathbf{p}_x$ . The target system similarly can be described by the Hamiltonian  $H_1(\mathbf{x}, \mathbf{p}_x)$ . Both the two systems can be connected by

$$H_1(\mathbf{x}, \mathbf{p}_x) = H_0(\mathbf{x}, \mathbf{p}_x) + \Delta H(\mathbf{x}, \mathbf{p}_x) \quad (3.1)$$

The Helmholtz free energy difference between the target and the reference systems,  $\Delta A$ , can be given in terms of the ratio of the corresponding partition functions,  $Q_1$  and  $Q_0$ :

$$\Delta A = -\frac{1}{\beta} \ln \frac{Q_1}{Q_0}, \quad (3.2)$$

where  $\beta = (k_B T)^{-1}$ , and

$$Q = \frac{1}{h^{3N} N!} \int \int \exp[-\beta H(\mathbf{x}, \mathbf{p}_x)] d\mathbf{x} d\mathbf{p}_x, \quad (3.3)$$

So, we obtain

$$\Delta A = -\frac{1}{\beta} \ln \frac{\int \int \exp[-\beta H_1(\mathbf{x}, \mathbf{p}_x)] d\mathbf{x} d\mathbf{p}_x}{\int \int \exp[-\beta H_0(\mathbf{x}, \mathbf{p}_x)] d\mathbf{x} d\mathbf{p}_x} \quad (3.4)$$

$$= -\frac{1}{\beta} \ln \frac{\int \int \exp[-\beta \Delta H(\mathbf{x}, \mathbf{p}_x)] \exp[-\beta H_0(\mathbf{x}, \mathbf{p}_x)] d\mathbf{x} d\mathbf{p}_x}{\int \int \exp[-\beta H_0(\mathbf{x}, \mathbf{p}_x)] d\mathbf{x} d\mathbf{p}_x}, \quad (3.5)$$

The probability density function of finding the reference system in a state defined by positions  $\mathbf{x}$  and momenta  $\mathbf{p}_x$  is

$$P_0(\mathbf{x}, \mathbf{p}_x) = \frac{\exp[-\beta H_0(\mathbf{x}, \mathbf{p}_x)]}{\int \int \exp[-\beta H_0(\mathbf{x}, \mathbf{p}_x)] d\mathbf{x} d\mathbf{p}_x} \quad (3.6)$$

If the probability density function is used, the Eq. 3.5 becomes

$$\Delta A = -\frac{1}{\beta} \int \int \exp[-\beta \Delta H(\mathbf{x}, \mathbf{p}_x)] P_0(\mathbf{x}, \mathbf{p}_x) d\mathbf{x} d\mathbf{p}_x, \quad (3.7)$$

or, equivalently,

$$\Delta A = -\frac{1}{\beta} \ln \langle \exp[-\beta \Delta H(\mathbf{x}, \mathbf{p}_x)] \rangle_0, \quad (3.8)$$

Here,  $\langle \cdots \rangle_0$  denotes an ensemble average over configurations sampled from the reference state. Equation 3.8 is the fundamental equation of **TP**. It states that  $\Delta A$  can be estimated by sampling only equilibrium configurations of the reference state.

Note that integration over the kinetic term in the partition function, Eq. 3.3, can be carried out analytically. Thus, it cancels out in Eq. 3.2, and Eq. 3.8 becomes

$$\Delta A = -\frac{1}{\beta} \ln \langle \exp(-\beta \Delta U) \rangle_0, \quad (3.9)$$

where  $\Delta U$  is the difference in the potential energy between the target and the reference states. The integration implied by the statistical average is now carried out over particle coordinates only.

If we reverse the reference and the target systems, and repeat the same derivation, using the same convention for  $\Delta A$  and  $\Delta U$  as before, we obtain

$$\Delta A = \frac{1}{\beta} \ln \langle \exp(\beta \Delta U) \rangle_1, \quad (3.10)$$

Although expressions Eq. 3.9 and Eq. 3.10 are formally equivalent, their convergence properties may be quite different. This means that there is a preferred direction to carry out the required transformation between the two states. One should start the perturbation from the system having the larger important phase space region. This means that the reference system should be that with the higher entropy, and the transformation should be proceed in the direction in which the entropy change  $\Delta S$  is negative.

The formulas for free energy difference, Eq. 3.9 and Eq. 3.10, are formally exact for any perturbation. However, this does not mean that they can always be successfully applied. Since  $\Delta A$  is calculated as the average over a quantity that depends only on  $\Delta U$ , this average can be taken over probability distribution  $P_0(\Delta U)$  instead of  $P_0(\mathbf{x}, \mathbf{p}_x)$ . Then,  $\Delta A$  in Eq. 3.7 can be expressed as a one dimensional integral over energy difference

$$\Delta A = -\frac{1}{\beta} \int \exp(-\beta \Delta U) P_0(\Delta U) d\Delta U, \quad (3.11)$$

If  $U_0$  and  $U_1$  were the functions of a sufficient number of identically distributed random variable, then  $\Delta U$  would be Gaussian distribution, which is a consequence of the central limit theorem. In practice, the probability distribution  $P_0(\Delta U)$  deviates somewhat from the ideal Gaussian case, but still has a ‘‘Gaussian-like’’ shape. This indicates that the value of the integral in Eq. 3.11 depends on the low-energy tail of the distribution.

Even though  $P_0(\Delta U)$  is only rarely an exact Gaussian, it is instructive to consider this case in more detail. If we substitute

$$P_0(\Delta U) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\Delta U - \langle \Delta U \rangle_0)^2}{2\sigma^2}\right] \quad (3.12)$$

where

$$\sigma^2 = \langle \Delta U^2 \rangle_0 - \langle \Delta U \rangle_0^2 \quad (3.13)$$

to Eq. 3.11, we obtain

$$\exp(-\beta \Delta A) = \frac{C}{\sqrt{2\pi}\sigma} \int \exp\left[-\frac{(\Delta U - \langle \Delta U \rangle_0 - \beta\sigma^2)^2}{2\sigma^2}\right] d\Delta U \quad (3.14)$$

Here,  $C$  is independent of  $\Delta U$

$$C = \exp\left[-\beta(\langle \Delta U \rangle_0 - \frac{1}{2}\beta\sigma^2)\right] \quad (3.15)$$

If  $P_0(\Delta U)$  is Gaussian, the integral in Eq. 3.14 can be evaluated analytically

$$\Delta A = \langle \Delta U \rangle_0 - \frac{1}{2}\beta\sigma^2 \quad (3.16)$$

### 3.1.2 Thermodynamic Integration

Thermodynamic Integration (TI) method was originally put forward by Kirkwood.[10].

If the free energy,  $A$ , is a continuous function of  $\lambda$  then we can write:

$$\Delta A = \int_0^1 \frac{\partial A(\lambda)}{\partial \lambda} d\lambda \quad (3.17)$$

Now,

$$A(\lambda) = -k_B T \ln Q(\lambda) \quad (3.18)$$

Thus,

$$\Delta A = -k_B T \int_0^1 \left[ \frac{\partial \ln Q(\lambda)}{\partial \lambda} \right] d\lambda \quad (3.19)$$

$$= \int_0^1 \frac{-k_B T}{Q(\lambda)} \frac{\partial Q(\lambda)}{\partial \lambda} d\lambda \quad (3.20)$$

From the definition of  $Q$ :

$$Q_{NVT} = \frac{1}{h^{3N} N!} \int \int \exp[-\beta H(\mathbf{x}, \mathbf{p}_x)] d\mathbf{x} d\mathbf{p}_x, \quad (3.21)$$

we can write the following for  $\frac{\partial Q(\lambda)}{\partial \lambda}$ :

$$\frac{\partial Q(\lambda)}{\partial \lambda} = \frac{1}{h^{3N} N!} \int \int \frac{\partial}{\partial \lambda} \exp[-\beta H(\mathbf{x}, \mathbf{p}_x)] d\mathbf{x} d\mathbf{p}_x, \quad (3.22)$$

Applying the chain rule:

$$\frac{\partial Q(\lambda)}{\partial \lambda} = -\frac{\beta}{h^{3N} N!} \int \int \frac{\partial H(\mathbf{x}, \mathbf{p}_x)}{\partial \lambda} \exp[-\beta H(\mathbf{x}, \mathbf{p}_x)] d\mathbf{x} d\mathbf{p}_x, \quad (3.23)$$

Substituting back into the expression for  $\frac{\partial A}{\partial \lambda}$  gives:

$$\frac{\partial A(\lambda)}{\partial \lambda} = \frac{1}{h^{3N} N!} \frac{1}{Q(\lambda)} \int \int \frac{\partial H(\mathbf{x}, \mathbf{p}_x)}{\partial \lambda} \exp[-\beta H(\mathbf{x}, \mathbf{p}_x)] d\mathbf{x} d\mathbf{p}_x, \quad (3.24)$$

$$= \int \int \frac{\partial H(\mathbf{x}, \mathbf{p}_x)}{\partial \lambda} \frac{\exp[-\beta H(\mathbf{x}, \mathbf{p}_x)]}{Q(\lambda)} d\mathbf{x} d\mathbf{p}_x, \quad (3.25)$$

$$= \left\langle \frac{\partial H(\mathbf{x}, \mathbf{p}_x, \lambda)}{\partial \lambda} \right\rangle_\lambda \quad (3.26)$$

Thus, the fundamental **TI** formula:

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H(\mathbf{x}, \mathbf{p}_x, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (3.27)$$

where  $\langle \dots \rangle_\lambda$  corresponds to the ensemble average obtained using the Hamiltonian  $H(\lambda)$ . In practice, the ensemble of configurations can be obtained

by molecule dynamic or Monte Carlo simulations. It is common practice in free energy calculations to use the coupling parameter  $\lambda$  for defining the transformation from an initial  $A$ , with Hamiltonian  $H_A$ , to a final state  $B$ , with Hamiltonian  $H_B$ . The simplest coupling is linear scaling:

$$H(\lambda) = (1 - \lambda)H_A + \lambda H_B \quad (3.28)$$

Eq. 3.27 represents the so-called thermodynamic integration formula. The exact calculation of  $\Delta A$  would require an infinite number of ensemble average for  $\lambda$  ranging from 0 to 1. Therefore, the integral in Eq. 3.27 needs to be approximated, e.g., by a summation over a discrete number of points  $\lambda_i$ , [11] leading to

$$\Delta A = \sum_i \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda_i} \Delta \lambda_i \quad (3.29)$$

A finite number of  $\lambda_i$  values between 0 and 1 is chosen and for each of them a complete molecule dynamic simulation is carried out resulting in an ensemble of configurations generated with  $H(\lambda_i)$ . The ensemble average of the derivative of the Hamiltonian with respect to  $\lambda$  is then calculated for each  $\lambda = \lambda_i$ .

The accuracy of TI integral formula depends the exact method of numerical integration. In addition to summation method, the simplest numerical integration is to evaluate the integrand at the midpoint:

$$\Delta A \simeq \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda=\frac{1}{2}} \quad (3.30)$$

This might be a good first thing to do to get some picture of what is going on, but is only depends to be accurate for very smooth or small changes.

### 3.1.3 Bennett Acceptance Ratio

Bennett acceptance ratio was developed by Bennett in 1976,[12] and was re-discovered by Crooks[13] and Shirts et al[14] over 20 years later. The Metropolis function is defined as

$$M(x) = \min\{1, \exp(-x)\}, \quad (3.31)$$

which has the property

$$M(x)/M(-x) = \exp(-x). \quad (3.32)$$

If we make a trial move that keeps the same configuration  $(q_1, \dots, q_N)$  but switches the potential function from  $U_0$  to  $U_1$  or vice-versa. The acceptance probabilities for such a pair of trial move must satisfy the detailed balance

$$M(U_1 - U_0) \exp(-U_0) = M(U_0 - U_1) \exp(-U_1). \quad (3.33)$$

Integrating this identity over all of configuration space and multiplying by the trivial factors  $Q_0/Q_0$  and  $Q_1/Q_1$ , one obtains:

$$Q_0 \frac{\int M(U_1 - U_0) \exp(-U_0) d\mathbf{q}}{Q_0} = Q_1 \frac{\int M(U_0 - U_1) \exp(-U_1) d\mathbf{q}}{Q_1}, \quad (3.34)$$

or simply

$$\frac{Q_0}{Q_1} = \frac{\langle M(U_0 - U_1) \rangle_1}{\langle M(U_1 - U_0) \rangle_0}. \quad (3.35)$$

The physical meaning of this formula is that a Monte Carlo calculation that included potential-switching trial moves would distribute configurations between  $U_1$  and  $U_0$  in the ratio of their configurational integrals.

A more general formula than Eq. 3.35 can be written as

$$\frac{Q_0}{Q_1} = \frac{Q_0 \int W \exp(-U_0 - U_1) d\mathbf{q}}{Q_1 \int W \exp(-U_1 - U_0) d\mathbf{q}} = \frac{\langle W \exp(-U_0) \rangle_1}{\langle W \exp(-U_1) \rangle_0}, \quad (3.36)$$

where  $W$  is an arbitrary weighting function.

Optimization of the free energy estimate is most easily carried out in the limit of large sample sizes. Let the available data consist of  $n_0$  statistically independent configurations from the  $U_0$  ensemble and  $n_1$  from the  $U_1$  ensemble, and let the data be used in Eq. 3.36 to obtain a finite-sample estimate of the reduced free energy difference  $\Delta A = A_1 - A_0 = \ln(Q_0/Q_1)$ . Using the error propagation equation,

$$\delta^2 [y(x_1, x_2)] = \left( \frac{\partial y}{\partial x_1} \right)^2 \delta^2(x_1) + \left( \frac{\partial y}{\partial x_2} \right)^2 \delta^2(x_2). \quad (3.37)$$



Thus we have the variance of  $\Delta A$

$$\delta^2(\Delta A) = \left(\frac{\partial \Delta A}{\partial Q_0}\right)^2 \delta^2 Q_0 + \left(\frac{\partial \Delta A}{\partial Q_1}\right)^2 \delta^2 Q_1 \quad (3.38)$$

$$= \left(\frac{1}{Q_0}\right)^2 \delta^2 Q_0 + \left(-\frac{1}{Q_1}\right)^2 \delta^2 Q_1 \quad (3.39)$$

$$= \left(\frac{1}{Q_0}\right)^2 \delta^2 Q_0 + \left(\frac{1}{Q_1}\right)^2 \delta^2 Q_1. \quad (3.40)$$

With the definition of variance  $\delta^2 X = \langle X^2 \rangle - \langle X \rangle^2$ , we have

$$\delta^2 Q_0 = \delta^2 \langle W \exp(-U_0) \rangle_1 \quad (3.41)$$

$$= \delta^2 \left( \frac{1}{n_1} \sum_{i=1}^{n_1} W_i \exp(-U_0(i)) \right) \quad (3.42)$$

$$= \sum_{i=1}^{n_1} \left( \frac{1}{n_1} \right)^2 \delta^2 (W_i \exp(-U_0(i))) \quad (3.43)$$

$$= \frac{1}{n_1} \delta^2 (W_i \exp(-U_0(i))) \quad (3.44)$$

$$= \frac{1}{n_1} \left\{ \langle (W \exp(-U_0))^2 \rangle_1 - (\langle W \exp(-U_0) \rangle_1)^2 \right\} \quad (3.45)$$

$$= \frac{1}{n_1} \left\{ \langle W^2 \exp(-2U_0) \rangle_1 - [\langle W \exp(-U_0) \rangle_1]^2 \right\} \quad (3.46)$$

With sufficiently large sample sizes, the error of this estimate will be nearly Gaussian, and its expected square is exactly the variance of  $\Delta A$

$$\begin{aligned} & \delta^2(\Delta A_{est} - \Delta A) \\ & \approx \frac{\langle W^2 \exp(-2U_1) \rangle_0}{n_0 [\langle W \exp(-U_1) \rangle_0]^2} + \frac{\langle W^2 \exp(-2U_0) \rangle_1}{n_1 [\langle W \exp(-U_0) \rangle_1]^2} - \frac{1}{n_0} - \frac{1}{n_1} \\ & = \frac{\int [(Q_0/n_0) \exp(-U_1) + (Q_1/n_1) \exp(-U_0)] W^2 \exp(-U_0 - U_1) d\mathbf{q}}{[\int W \exp(-U_0 - U_1) d\mathbf{q}]^2} \\ & \quad - \frac{1}{n_0} - \frac{1}{n_1}. \end{aligned} \quad (3.47)$$

To minimize it with respect to  $W$ , we have

$$W = \text{const} \times \left( \frac{Q_0}{n_0} \exp(-U_1) + \frac{Q_1}{n_1} \exp(-U_0) \right)^{-1}. \quad (3.48)$$

Substituting this into Eq. 3.36 yields

$$\frac{Q_0}{Q_1} = \frac{\langle f(U_0 - U_1 + C) \rangle_1}{\langle f(U_1 - U_0 - C) \rangle_0} \exp(+C), \quad (3.49)$$

where

$$C = \ln \frac{Q_0 n_1}{Q_1 n_0}, \quad (3.50)$$

and  $f$  denotes the Fermi function

$$f(x) = \frac{1}{1 + \exp(+x)} \quad (3.51)$$

### 3.1.4 Weighted Histogram Analysis Method

#### Weighted Histogram Analysis Method for Parallel Tempering

The following derivation quite follows Ref. [15]. The central quantity in statistical mechanics is partition function  $Z$ , which in textbook is often written as

$$Z = \int \exp(-\beta U(\mathbf{R})) d\mathbf{R}. \quad (3.52)$$

This is an integral in coordinate space. It also can be written as an integral in energy space

$$Z = \int \Omega(U) \exp(-\beta U) dU, \quad (3.53)$$

where  $\Omega(U)$  is density of states and  $\Omega(U)\Delta U$  is the number of states in the region  $U - \Delta U/2 < U < U + \Delta U/2$ . Accordingly, the statistical expectation of an operator  $\mathbf{A}$  can be calculated by

$$\langle \mathbf{A} \rangle = \frac{\int \mathbf{A}(U) \Omega(U) \exp(-\beta U) dU}{\int \Omega(U) \exp(-\beta U) dU}, \quad (3.54)$$

where

$$\mathbf{A}(U') = \frac{\int \delta(U(\mathbf{R}) - U') \mathbf{A}(\mathbf{R}) d\mathbf{R}}{\int \delta(U(\mathbf{R}) - U') d\mathbf{R}}. \quad (3.55)$$

Therefore, the core objective is to calculate  $\Omega(U)$ .

Suppose we have one trajectory with  $K$  snapshots denoted as  $\{\mathbf{x}_n\}$ . We then discretize the energy space into  $M$  bins with width  $\Delta U$ , and count the number of snapshots fallen into each bin. For convenience, we define  $\psi_m(U)$  as

$$\psi_m(U) = \begin{cases} 1 & \text{if } U \in [U_m - \Delta U/2, U_m + \Delta U/2) \\ 0 & \text{otherwise} \end{cases} \quad (3.56)$$

Then the histogram for the  $m$ th energy bin is

$$H_m = \sum_{n=1}^N \psi_m(U(\mathbf{R}_n)), \quad (3.57)$$

with variances (see Appendix B)

$$\delta^2 H_m = g_m \langle H_m \rangle \left( 1 - \frac{\langle H_m \rangle}{N} \right). \quad (3.58)$$

The ratio of the histogram  $H_m$  to the total number of snapshots  $N$  divided by the bin width  $\Delta U$  can be approximately taken as the probability of states in this bin, i.e.,

$$\frac{\Omega_m \exp(-\beta U_m)}{Z} \approx \frac{H_m}{N \Delta U}. \quad (3.59)$$

Therefore,

$$\begin{aligned}\Omega_m &= \frac{1}{U_m} \cdot \frac{H_m}{N} \cdot \frac{Z(\beta)}{\exp(-\beta U_m)} \\ &= \frac{H_m}{N \Delta U \exp[f_k - \beta_k U_m]},\end{aligned}\quad (3.60)$$

and variances

$$\delta^2 \Omega_m = \frac{\delta^2 H_m}{(N \Delta U \exp[f_k - \beta_k U_m])^2}, \quad (3.61)$$

in which we have defined a dimensionless free energy  $f = -\ln Z(\beta)$ .

Practically, we may run multiple ( $K$ ) trajectories using, for example, replica exchange molecular dynamics simulations. For each trajectory (index  $k$ ), we have unique estimators for the histogram  $H_{mk}$ , the density of states  $\Omega_{mk}$  and their variances  $\delta^2 H_{mk}$  and  $\delta^2 \Omega_{mk}$  being

$$H_{mk} = \sum_{n=1}^{N_k} \psi_m(U(\mathbf{R}_{kn})), \quad (3.62)$$

$$\delta^2 H_{mk} = g_{mk} \langle H_{mk} \rangle \left(1 - \frac{\langle H_{mk} \rangle}{N_k}\right), \quad (3.63)$$

$$\Omega_{mk} = \frac{H_{mk}}{N_k \Delta U \exp[f_k - \beta_k U_{mk}]}, \quad (3.64)$$

and

$$\delta^2 \Omega_{mk} = \frac{\delta^2 H_{mk}}{(N_k \Delta U \exp[f_k - \beta_k U_{mk}])^2}, \quad (3.65)$$

The optimum estimator of the density of states from all the simulations is

$$\Omega_m = \frac{\sum_{k=1}^K [\delta^2 \Omega_{mk}]^{-1} \Omega_{mk}}{\sum_{k=1}^K [\delta^2 \Omega_{mk}]^{-1}}, \quad (3.66)$$

which is the weighted average of density of states of all the trajectories with the weight reversely proportional to the uncertainties (see Appendix A).

To make the expression simpler, here we take some approximations. First, normally the energy space is split into a large number of bins. The histogram in each bin is much smaller than the total number of snapshots, i.e.  $H_{mk} \ll N_k$ . The average of  $H_{mk}$  can be related to the optimum estimator of the density of states, i.e.

$$\langle H_{mk} \rangle = N_k \Delta U \Omega_m \exp(f_k - \beta_k U_m). \quad (3.67)$$

Then we have

$$\delta^2 H_{mk} = g_{mk} N_k \Delta U \Omega_m \exp(f_k - \beta_k U_m) \quad (3.68)$$

and

$$\delta^2 \Omega_{mk} = \frac{\Omega_m}{g_{mk}^{-1} N_k \Delta U \exp(f_k - \beta_k U_m)}. \quad (3.69)$$

Taking Eq. 3.64 and Eq. 3.69 into Eq. 3.66, we find

$$\Omega_m = \frac{\sum_{k=1}^K g_{mk}^{-1} H_{mk}}{\sum_{k=1}^K g_{mk}^{-1} N_k \Delta U \exp(f_k - \beta_k U_m)}, \quad (3.70)$$

in which

$$f_k = -\ln \sum_{m=1}^M \Omega_m \Delta U \exp(-\beta_k U_m). \quad (3.71)$$

Obviously, Eq. 3.70 and Eq. 3.71 must be solved iteratively. The uncertainty of  $\Omega_m$  is given by

$$\delta^2 \Omega_m = \frac{\Omega_m}{\sum_{k=1}^K g_{mk}^{-1} N_k \Delta U \exp(f_k - \beta_k U_m)} \quad (3.72)$$

and the relative uncertainty is given by

$$\frac{\delta^2 \Omega_m}{\Omega_m^2} = \left[ \sum_{k=1}^K g_{mk}^{-1} H_{mk} \right]^{-1}. \quad (3.73)$$

Using the density of states and its variance, we can estimate the expectation of any configuration function  $A(\mathbf{R})$  at any inverse temperature  $\beta$

$$\langle A \rangle_\beta \approx \frac{\sum_{m=1}^M \Omega_m \Delta U \exp(-\beta U_m) A_m}{\sum_{m=1}^M \Omega_m \Delta U \exp(-\beta U_m)}, \quad (3.74)$$

where

$$A_m = \frac{\int d\mathbf{R} A(\mathbf{R}) \psi_m(U(\mathbf{R}))}{\int d\mathbf{R} \psi_m(U(\mathbf{R}))}. \quad (3.75)$$

Using histograms of bin  $m$  from all the simulations and defining  $H_m = \sum_{k=1}^K H_{mk}$ , an estimator of  $A_m$  denoted as  $\hat{A}_m$  can be calculated as

$$\hat{A}_m = H_m^{-1} \sum_{k=1}^K \sum_{n=1}^{N_k} \psi_m(U(\mathbf{R}_{kn})) A(\mathbf{R}_{kn}). \quad (3.76)$$

Taking Eq. 3.76 into Eq. 3.74, we obtain an estimator of  $\hat{A}(\beta)$

$$\hat{A}(\beta) = \frac{\sum_{m=1}^M \Omega_m \Delta U \exp(-\beta U_m) H_m^{-1} \sum_{k=1}^K \sum_{n=1}^{N_k} \psi_m(U(\mathbf{R}_{kn})) A(\mathbf{R}_{kn})}{\sum_{m=1}^M \Omega_m \Delta U \exp(-\beta U_m)} \quad (3.77)$$

$$= \frac{\sum_{m=1}^M \Omega_m \Delta U \exp(-\beta U_m) H_m^{-1} \sum_{k=1}^K \sum_{n=1}^{N_k} \psi_m(U(\mathbf{R}_{kn})) A(\mathbf{R}_{kn})}{\sum_{m=1}^M \Omega_m \Delta U \exp(-\beta U_m) H_m^{-1} \sum_{k=1}^K \sum_{n=1}^{N_k} \psi_m(U(\mathbf{R}_{kn}))} \quad (3.78)$$

$$= \frac{\sum_{k=1}^K \sum_{n=1}^{N_k} w_{kn}(\beta) A_{kn}}{\sum_{k=1}^K \sum_{n=1}^{N_k} w_{kn}(\beta)}, \quad (3.79)$$

where the per-configuration weights  $w_{kn}(\beta)$  is given by

$$w_{kn}(\beta) = \sum_{m=1}^M H_m^{-1} \psi_m(U(\mathbf{R}_{kn})) \Omega_m \exp(-\beta U_m) \quad (3.80)$$

### Weighted Histogram Analysis Method From Maximum Likelihood

The following derivation quite follows Ref. [16], in which maximum likelihood principle is utilized. Suppose we have performed  $K$  simulations, each at a different inverse temperature  $\beta_k$  and possibly with different biasing potential  $w_k(\mathbf{R})$ . We then discretize the 2D plane spanned by the coordinate and unbiased potential energy into bins, each characterized by  $\mathbf{R}_j$  and  $E_h$ . To make the following derivation cleaner, we map the 2D bins to one dimensional series with index  $l, l = 1, \dots, L$ . Next, we construct histograms for bins using all the samples from the simulations. The probability of finding the system in bin  $l$  during the  $k$ th simulation can be written as

$$p_{k,l} = f_k c_{k,l} p_l^0, \quad (3.81)$$

in which  $p_l^0$  is the (simulation-independent) unbiased probability,

$$\begin{aligned} c_{k,l} &= \exp \{ -\beta_k [E_l + w_{k,l}] + \beta_0 E_l \} \\ &= \exp [ -(\beta_k - \beta_0) E_l ] \exp (-\beta_k w_{k,l}) \end{aligned} \quad (3.82)$$

is the bias factor,  $E_l$  is the unbiased energy of bin  $l$ ,  $f_k = \{ \sum_l c_{k,l} p_l^0 \}^{-1}$  is the normalization factor. It is worth emphasizing that the biasing potential can be multiple dimensional as, for instance, in a two-dimensional umbrella sampling. If the biasing is only in temperature-space as in replica exchange molecular dynamics

$$c_{k,l} = \exp [ -(\beta_k - \beta_0) E_l ], \quad (3.83)$$

while if the biasing is only in potential as in umbrella sampling

$$c_{k,l} = \exp(-\beta_0 w_{k,l}). \quad (3.84)$$

If we assume that each count in each histogram is independent, then the likelihood of observing the  $k$ th histogram is given by the multinomial distribution

$$P(n_{k,1}, n_{k,2}, \dots, n_{k,L} | p_{k,1}, p_{k,2}, \dots, p_{k,L}) = \frac{\left(\sum_l n_{k,l}\right)!}{\prod_l n_{k,l}!} \prod_{l=1}^L (p_{k,l})^{n_{k,l}} \propto \prod_{l=1}^L \left(f_k c_{k,l} p_l^0\right)^{n_{k,l}}. \quad (3.85)$$

For all  $K$  simulations, the likelihood is the product of multinomial

$$P(n_{1,1}, \dots, n_{1,L}; \dots; n_{K,1}, \dots, n_{K,L} | p_1^0, \dots, p_L^0) \propto \prod_{k=1}^K \prod_{l=1}^L \left(f_k c_{k,l} p_l^0\right)^{n_{k,l}}, \quad (3.86)$$

where the likelihood is conditional only on the unbiased probabilities  $p_l^0$ , since the bias factors  $c_{k,l}$  are known parameters, and the normalization constants  $f_k$  are known conditional on  $p_l^0$ . The maximum likelihood estimate of the unbiased probabilities can be found by maximizing  $P$  in Eq. 3.86 with respect to  $p_1^0, \dots, p_L^0$  and are given by solutions of the simultaneous nonlinear equations

$$p_l^0 = \frac{\sum_{k=1}^K n_{k,l}}{\sum_{k=1}^K N_k f_k c_{k,l}} \quad (\text{for each } l) \quad (3.87)$$

and

$$f_k = \left\{ \sum_l c_{k,l} p_l^0 \right\}^{-1}, \quad (3.88)$$

where  $N_k$  is the total number of counts in the  $k$ th histogram.

### Weighted Histogram Analysis Method From Minimizing Statistical Error

In this section, the “traditional” derivation method of WHAM are briefly reviewed.[17] In the WHAM, the goal is to get an optimal unbiased probability distribution  $\rho_0(\eta)$ , where  $\eta$  is a series of discretized histogram bins indexed by  $j = 1, 2, 3, \dots, M$  along a certain reaction coordinate. WHAM can be used to analyze the Umbrella Sampling (US) simulations, where a set of simulations indexed by  $i = 1, 2, 3, \dots, S$  are performed with a series of

biasing potentials added on the reaction coordinate  $\eta$ . To consider a reference molecular system with the potential energy  $U_0(\mathbf{x})$ , where  $\mathbf{x}$  is the set of atomic coordinates. The reaction coordinate  $\eta$  is a function of the atomic coordinates, i.e.  $\eta(\mathbf{x})$ . To suppose that the  $i$ th molecular simulation has been performed using potential energy function of the form,

$$U_i^{(b)}(\eta) = U_0(\mathbf{x}) + W_i(\eta(\mathbf{x})) \quad (3.89)$$

where  $W_i(\eta(\mathbf{x}))$  is the biasing potential adding on the reaction coordinate  $\eta$ , e.g.  $W_i(\eta) = \frac{1}{2}k_i(\eta - \eta_i)^2$  can be a Harmonic form. From these simulations a set of biased probability distributions  $\rho_i^{(b)}(\eta)$  can be obtained.  $\rho_i^{(b)}(\eta)$  is computed as normalized histogram of the values of  $\eta$  occurring during the  $i$ th simulation.

$$\rho_i^{(b)}(\eta) = \frac{e^{-\beta U_i^{(b)}(\eta)}}{Q_i^{(b)}} \quad (3.90)$$

where  $Q_i^{(b)} = \int e^{-\beta U_i^{(b)}(\eta(\mathbf{x}))} d\mathbf{x} = e^{-\beta f_i^{(b)}}$  and  $f_i^{(b)}$  is the free energy coming from the adding of the biasing potential  $W_i(\eta(\mathbf{x}))$  to the reference potential  $U_0(\mathbf{x})$ . The corresponding normalized unbiased probability distribution  $\rho_i^{(u)}(\eta)$  from  $i$ th simulation is defined as,

$$\rho_i^{(u)}(\eta) \equiv \frac{e^{-\beta U_0(\mathbf{x})}}{Q_i^{(u)}} \quad (3.91)$$

$$\equiv \frac{e^{\beta[W_i(\eta) - f_i^{(b)}]} \rho_i^{(b)}(\eta)}{Q_i^{(u)}} \quad (3.92)$$

where  $Q_i^{(u)} = \int e^{-\beta U_0(\mathbf{x})} d\mathbf{x} = e^{-\beta f_i^{(u)}}$ . In the following, the free energy  $f_i^{(b)}$  and  $Q_i^{(u)}$  are assumed to be known, the method providing a posterior algorithm for their computation. It has been shown that the WHAM method simplifies considerably the task of recombining together the unbiased histograms  $\rho_i^{(u)}(\eta)$  to get the total probability distribution  $\rho_0(\eta)$ . It consists in writing  $\rho_0(\eta)$  as a linear  $\eta$ -dependent combination of the unbiased probability distribution,

$$\rho_0(\eta) = \sum_{i=1}^S p_i(\eta) \rho_i^{(u)}(\eta) \quad (3.93)$$

The weights are required to be normalized,

$$\sum_{i=1}^S p_i(\eta) = 1 \quad (3.94)$$

and they are chosen so as to minimize the statistical error made on the probability distribution, that is, for any given value of  $\eta$ ,

$$\frac{\partial(\sigma^2[\rho_0(\eta)])}{\partial p_i(\eta)} = 0 \quad (3.95)$$



Then,  $\rho_0(\eta)$  takes the form,

$$\rho_0(\eta) = \sum_{i=1}^S \frac{N_i e^{-\beta[W_i(\eta) - f_i^{(b)}]}}{\sum_{k=1}^S N_k e^{-\beta[W_k(\eta) - f_k^{(b)}]}} \rho_i^{(u)}(\eta) \quad (3.96)$$

$$= \frac{1}{Q_i^{(u)}} \sum_{i=1}^S \frac{N_i}{\sum_{k=1}^S N_k e^{-\beta[W_k(\eta) - f_k^{(b)}]}} \rho_i^{(b)}(\eta) \quad (3.97)$$

$$= \frac{1}{Q_i^{(u)}} \frac{\sum_{i=1}^S N_i \rho_i^{(b)}(\eta)}{\sum_{k=1}^S N_k e^{-\beta[W_k(\eta) - f_k^{(b)}]}} \quad (3.98)$$

Until now, the treatment assumed that the free energy parameters  $f_i^{(b)}$  were known. In fact, these parameters can be obtain self-consistently. Indeed, the definition of the free energy  $f_i^{(b)}$  is,

$$e^{-\beta f_i^{(b)}} = \int d\eta e^{-\beta U_i^{(b)}(\eta)} \quad (3.99)$$

$$= Q_i^{(u)} \int d\eta \rho_0(\eta) e^{-\beta W_i(\eta)} \quad (3.100)$$

$$= \int d\eta \frac{\sum_{i=1}^S N_i \rho_i^{(b)}(\eta)}{\sum_{k=1}^S N_k e^{-\beta[W_k(\eta) - f_k^{(b)}]}} e^{-\beta W_i(\eta)} \quad (3.101)$$

The set of parameters  $f_i^{(b)}$  appear on the left and on the right of Eq. 3.101, which can be solved iteratively. A first guess of values  $f_i^{(b)0}$  is used on the right of Eq. 3.101 to compute a new set of values  $f_i^{(b)1}$ , which is in turn used as the new guess to compute  $f_i^{(b)1}$  and so on until convergence of the process. To get rid of the constant  $\frac{1}{Q_i^{(u)}}$ , one may subtract the offset constant  $f_0^{(u)}$  from all the  $f_i^{(u)}$ .

### 3.1.5 Multistate Bennett Acceptance Ratio

The following derivation quite follows Ref. [18] Imaging you have carried out a series of simulations such umbrella sampling or parallel simulations such as replica exchange molecular dynamics simulations. Now you have  $K$  trajectories and each trajectory is characterized by Hamiltonian  $H_k$  and inverse temperature  $\beta_k$ . The trajectories unnecessarily have the same number of conformations. Instead, the number of conformations in trajectory  $k$  is  $N_k$ . Now, you mix all the samples and randomly pick one sample out of them. The probability for this sample to have coordinates  $\mathbf{R}$  is

$$p_m(\mathbf{R}) = \frac{1}{N} \sum_{k=1}^K N_k p_k(\mathbf{R}), \quad (3.102)$$

in which the subscript  $m$  means mixed ensemble and  $N = \sum_{k=1}^K N_k$ .  $p_k(\mathbf{R}_n)$  is the probability of finding this snapshot in trajectory  $k$ , which satisfies

$$p_k(\mathbf{R}) = c_k^{-1} q_k(\mathbf{R}). \quad (3.103)$$

$c_k$  is the normalization constant. You can see that this mixed ensemble does not follow Boltzmann statistics, even if  $q_k$  does. It can be proved that if  $p_k$  is normalized, then  $p_m$  is also normalized.

The expectation of any operator  $\hat{O}$  averaged over this mixed ensemble can be calculated by

$$\langle O \rangle_m \approx \frac{1}{N} \sum_{n=1}^N O(\mathbf{R}_n). \quad (3.104)$$

Using energy reweighting[5], we can calculate the expectation of this operator under any other Hamiltonian  $H_i$  and probability  $p_i$ , which can be expressed as

$$\begin{aligned} \langle O \rangle_i &= \int O(\mathbf{R}) p_i(\mathbf{R}) d\mathbf{R} \\ &= \int O(\mathbf{R}) \frac{p_i(\mathbf{R})}{p_m(\mathbf{R})} p_m(\mathbf{R}) d\mathbf{R} \\ &\approx \frac{1}{N} \sum_{n=1}^N O(\mathbf{R}_n) \frac{p_i(\mathbf{R}_n)}{p_m(\mathbf{R}_n)} \\ &= \frac{1}{N} \sum_{n=1}^N O(\mathbf{R}_n) c_i^{-1} \frac{q_i(\mathbf{R}_n)}{p_m(\mathbf{R}_n)} \\ &= \sum_{n=1}^N O(\mathbf{R}_n) c_i^{-1} \frac{q_i(\mathbf{R}_n)}{\sum_{k=1}^K N_k p_k(\mathbf{R}_n)} \end{aligned} \quad (3.105)$$

Let  $O = 1$ , we find

$$1 = \sum_{n=1}^N c_i^{-1} \frac{q_i(\mathbf{R}_n)}{\sum_{k=1}^K N_k p_k(\mathbf{R}_n)}. \quad (3.106)$$

Since  $c_i$  does not depend on  $n$ ,

$$\begin{aligned} c_i &= \sum_{n=1}^N \frac{q_i(\mathbf{R}_n)}{\sum_{k=1}^K N_k p_k(\mathbf{R}_n)} \\ &= \sum_{n=1}^N \frac{q_i(\mathbf{R}_n)}{\sum_{k=1}^K N_k c_k^{-1} q_k(\mathbf{R}_n)} \end{aligned} \quad (3.107)$$

In Boltzmann statistics,  $q_k(\mathbf{R}) = \exp[-\beta_k U_k(\mathbf{R})]$  and  $c_k = \int q_k(\mathbf{R}) d\mathbf{R}$  is the partition function or the normalization constant. *Note that we have not assumed anything about the statistics of ensemble  $k$  and  $i$ . Besides,  $i$  is unnecessarily within  $\{k\}$ . For instance, if we run replica exchange molecular dynamics simulations at  $K$  inverse temperatures  $\beta_1, \dots, \beta_K$ ,  $\beta_i$  can be either one of these inverse temperatures or any other inverse temperature between  $\beta_1$  and  $\beta_K$ . But extrapolation to inverse temperatures outside the range of  $[\beta_K, \beta_1]$  is not recommended.*

If  $q_k$  and  $q_i$  follow Boltzmann statistics, and we define free energy  $f_i = -\beta_i^{-1} \ln c_i$ , Eq. 3.107 becomes

$$f_i = -\beta_i^{-1} \ln \sum_{n=1}^N \frac{\exp(-\beta_i U_i(\mathbf{R}_n))}{\sum_{k=1}^K N_k \exp(\beta_k f_k - \beta_k U_k(\mathbf{R}_n))}, \quad (3.108)$$

which must be solved self-consistently and can be determined up to a constant. We can fix  $f_1$  (to 0 usually).

Again, from Eq. 3.105, we can define

$$W_{in} = c_i^{-1} \frac{q_i(\mathbf{R}_n)}{\sum_{k=1}^K N_k c_k^{-1} q_k(\mathbf{R}_n)}, \quad (3.109)$$

which is the weight of snapshot  $n$  in ensemble  $i$ .

### **3.1.6 Adaptive Biasing Force Method**

Adaptive biasing force method was proposed by Darve and Pohorille in 2001.[19].

### 3.1.7 Non-Equilibrium Work

Non-Equilibrium Work (NEW) method for equilibrium free energy calculations was firstly proposed by Jarzynski.[20]. In 1997, Jarzynski had showed the Jarzynski's identity:

$$\langle \exp[-\beta W(\tau)] \rangle = \exp(-\beta \Delta A) \quad (3.110)$$

where a prescribed path  $\lambda(t)$  connects the initial and final states, with  $\lambda(0) = 0$  and  $\lambda(\tau) = 1$ , and  $\Delta A = A(1) - A(0)$  the free energy difference between the two states. The average  $\langle \dots \rangle$  in Eq. 3.110 requires some explanation. It is a combination of an ensemble average over initial conditions, and a path average over trajectory realizations. Initial conditions are chosen according to the equilibrium Boltzmann probability in the  $\lambda(0)$  state. The path average samples all realizations of dynamic paths, weighted by their respective path action, under the time evolution of the system with an explicitly time-varying Hamiltonian. This equality relation also be deduced by Crooks from the view of markovian and microscopically reversible dynamics.[21] Here is no a detailed derivation of Eq. 3.110.

Now, we consider creating an equilibrium configuration in the state  $\lambda = 0$  and then slowly changing  $\lambda$  from 0 to 1. As the coupling parameter is advanced, the system continues to sample phase space by molecule dynamics or Monte Carlo simulations, but under an explicitly time-dependent Hamiltonian. In the limit of a very slow transformation, with some caveats of Hamiltonian dynamics, the system will remain close to equilibrium. The free energy difference can then be evaluated by changing  $\lambda$  continuously

$$\Delta A = \lim_{\tau \rightarrow \infty} \int_0^\tau \frac{\partial H[\mathbf{x}(t); \lambda]}{\partial \lambda} \Big|_{\lambda=\lambda(t)} \dot{\lambda}(t) dt \quad (3.111)$$

where the time derivative of the coupling parameter  $\lambda$  is denoted by a dot. In the Eq. 3.111, the limit of  $\tau \rightarrow \infty$  ensures that the transformation is performed infinitely slowly, and thus reversibly. The right-hand side of Eq. 3.111 correspondingly defines the "reversible work" done on the system during the transformation.

If the system is instead transformed between the initial and final states over a finite time interval  $\tau$ , the free energy obtained from Eq. 3.111 without the  $\tau \rightarrow \infty$  limit will only be approximate. In the corresponding slow-growth method of free energy calculations, the system will not be able to sample the phase space exhaustively at each value of  $\lambda$ . rendering the transformation irreversible. As the transformation proceeds, the system will be gradually driven out of equilibrium, causing hysteresis effects. From the second law of thermodynamic, it is expected that the work  $W(\tau)$  performed during the nonequilibrium transformation is on average larger than or equal to the free energy difference between the two states

$$\langle W(\tau) \rangle \geq \Delta A, \quad (3.112)$$

with the difference between the work and free energy accounting for heat-dissipation effects. The work  $W(\tau)$  performed on the system is the accumulated energies cost required to change the system

$$W(\tau) = \int_0^\tau \frac{\partial H[\mathbf{x}(t); \lambda]}{\partial \lambda} \Big|_{\lambda=\lambda(t)} \dot{\lambda}(t) dt \quad (3.113)$$

The equality in Eq. 3.112 will normally be achieved only if the transformation is infinitely slow,  $\tau \rightarrow \infty$ . For paths of finite length, the amount of dissipated work,  $\langle W(\tau) \rangle - \Delta A \geq 0$ , will depend on the chosen transformation path  $\lambda(t)$ .

Jarzynski's identity, Eq. 3.110, immediately leads to the second law in the form of Eq. 3.112 because of Jensen's inequality,  $\langle e^{-x} \rangle \geq e^{-x}$ . Moreover, TI and TP can be thought as the limiting cases of the nonequilibrium process. When  $\tau \rightarrow \infty$  or  $\dot{\lambda}(t) \rightarrow 0$ , this is an infinitely slow transformation and the Eq. 3.111 is the formula of TI

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H(\mathbf{x}, \mathbf{p}_x, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (3.114)$$

When  $\tau \rightarrow 0$  or  $\dot{\lambda}(t) \rightarrow \infty$ , this is an infinitely fast transformation where the configurations will not relax and the work is simply the change in the Hamiltonian when going from the initial to the final state,

$$\lim_{\tau \rightarrow 0} W(\tau) = H(\mathbf{x}(0); \lambda = 1) - H(\mathbf{x}(0); \lambda = 0) \quad (3.115)$$

Substituting the Eq. 3.115 into the Eq. 3.110, the formula of TP can be recovered:

$$\Delta A = -\frac{1}{\beta} \ln \langle \exp[-\beta \Delta H(\mathbf{x}, \mathbf{p}_x)] \rangle_0, \quad (3.116)$$

In the reference paper[21], Crooks showed that the distribution of work values from forward and backward paths satisfies a relation that is central to histogram methods in free energy calculations

$$\frac{p_f[w = W(\tau)]}{p_b[w = -\underline{W}(\tau)]} = \exp[\beta(w - \Delta A)], \quad (3.117)$$

where  $p_f[w = W(\tau)]$  and  $p_b[w = -\underline{W}(\tau)]$  are the probability densities of the work values for forward and reversed transformation paths (with a sign change in the work of the reverse work path). Both are normalized, i.e.,  $\int p_f(w) dw = \int p_b(w) dw = 1$ . It is noted that Jarzynski's identity Eq. 3.110 follows from Eq. 3.117 simply by integration over  $w$  because the probability densities are normalized to 1:

$$\int p_f(W) e^{-\beta W} dW = \int p_b(W) e^{-\beta \Delta A} dW, \quad (3.118)$$

Because of the normalization condition, the right-hand side is equal to  $\exp(-\beta\Delta A)$ , and Jarzynski's identity follows.

Following the Crooks Fluctuation Theorem (CFT), [21] Bennett's acceptance ratio can be applicable to nonequilibrium calculations. This approach was combined with a maximum likelihood estimate, and accurate free energy differences were obtained. [14] In this approach,  $\Delta A$  is calculated via

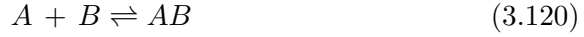
$$\sum_{i=1}^{n_F} \frac{1}{1 + \exp[\beta(M + W_i - \Delta A)]} = \sum_{j=1}^{n_R} \frac{1}{1 + \exp[-\beta(M + W_j - \Delta A)]}, \quad (3.119)$$

where  $n_F$  and  $n_R$  are the numbers of the forward and reverse transformations respectively,  $W_i$  and  $W_j$  are the work of forward and reverse measurements respectively, and  $M = \beta^{-1} \ln(n_F/n_R)$ . The corresponding statistical variance of  $\Delta A$ ,  $\sigma^2$  is calculated using Eq. 10 in the reference paper. [14].

## 3.2 Approximate Methods

### 3.2.1 Molecular Mechanics/Poisson-Boltzmann Surface Area

The following derivation follows Ref. [22]. The Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) method is often used in the calculations of binding free energy of a substrate to a receptor. The standard binding free energy for a reaction between a receptor (A) and a substrate (B)



is expressed as a ratio of configuration integrals

$$\begin{aligned} \Delta G_{AB}^0 &= -RT \ln \left( \frac{C^0}{8\pi^2} \cdot \frac{Z_{N,AB} Z_{N,0}}{Z_{N,A} Z_{N,B}} \right) + P^0 \langle \Delta V_{AB} \rangle \\ &= -RT \ln \left( \frac{C^0}{8\pi^2} \cdot \frac{\frac{Z_{N,AB}}{Z_{N,0}}}{\frac{Z_{N,A}}{Z_{N,0}} \frac{Z_{N,B}}{Z_{N,0}}} \right) + P^0 \langle \Delta V_{AB} \rangle, \end{aligned} \quad (3.121)$$

where  $R$  is the gas constant,  $T$  is the temperature,  $C^0$  is the standard state concentration (1  $M$ ),  $N$  is the number of solvent molecules, and  $P^0 \langle \Delta V_{AB} \rangle$  is the pressure-volume work associated with changing the system size after the association of two species into one complex. For water solution at 1 atm, the last term is negligibly small. There are no mass-dependent terms in Eq. 3.121, which is a direct result of equal kinetic contribution to the partition function of the bound and the free species. The configuration integral of the receptor, A, in solution is

$$Z_{N,A} = \int e^{-\beta U(r_A, r_S)} dr_A dr_S, \quad (3.122)$$

where  $U(r_A, r_S)$  is the potential energy as a function of all solute coordinates,  $r_A$ , and solvent coordinates,  $r_S$ , and  $\beta$  is the reciprocal of the product of the Boltzmann constant and temperature. The total potential energy can be decomposed into  $U(r_A) + U(r_S) + \Delta U(r_A, r_S)$ . Similar for  $B$ , the substrate. For pure solvent, the configuration integral is

$$Z_{N,0} = \int e^{-\beta U(r_S)} dr_S. \quad (3.123)$$

The ratio of configuration integrals in Eq. 3.121 can be simplified with an implicit solvent approximation, as

$$\begin{aligned} \frac{Z_{N,A}}{Z_{N,0}} &= Z_A = \frac{\int e^{-\beta U(r_A)} \left\{ \int e^{-\beta \Delta U(r_A, r_S)} e^{-\beta U(r_S)} dr_S \right\} dr_A}{\int e^{-\beta U(r_S)} dr_S} \\ &= \int e^{-\beta [U(r_A) + W(r_A)]} dr_A, \end{aligned} \quad (3.124)$$



where

$$W(r_A) = -RT \ln \left( \frac{\int e^{-\beta \Delta U(r_A, r_S)} e^{-\beta U(r_S)} dr_S}{\int e^{-\beta U(r_S)} dr_S} \right) \quad (3.125)$$

is the solvation free energy of the receptor  $A$  at fixed coordinate  $r_A$ . Analogous equations hold for the complex and substrate.

For the complex, we define the position (translational degrees of freedom) and orientation (rotational degrees of freedom) of the substrate with respect to the receptor as  $\delta_B \equiv (x_1, x_2, x_3, \xi_1, \xi_2, \xi_3)$ . Generally, these degree-of-freedom is very limited. The complex configuration integral is

$$Z_{AB} = \int e^{-\beta[U(r_A, r_{B'}, \delta_B) + W(r_A, r_{B'}, \delta_B)]} dr_A dr_{B'} d\delta_B, \quad (3.126)$$

where  $r_{B'}$  represents the remaining internal degrees of freedom of the bound substrate and  $\delta_B$  spans conformations where  $A$  and  $B$  form a complex. Then we assume that the translational and rotational motions of the substrate in the bound state are not strongly coupled with the other degrees of freedom, and we decompose the potential and solvation energies as (*so weird!*)

$$\begin{aligned} U(r_A, r_{B'}, \delta_B) + W(r_A, r_{B'}, \delta_B) \\ \approx U_1(\delta_B) + W_1(\delta_B) + U_2(r_A, r_{B'}) + W_2(r_A, r_{B'}). \end{aligned} \quad (3.127)$$

We further assume that the residual translational and rotational motions of the substrate are uncorrelated. Therefore

$$U_1(\delta_B) \approx U(x_1, x_2, x_3) + U(\xi_1, \xi_2, \xi_3), \quad (3.128)$$

and

$$W_1(\delta_B) \approx W(x_1, x_2, x_3) + W(\xi_1, \xi_2, \xi_3). \quad (3.129)$$

Now, Eq. 3.121 can be written as

$$\Delta G_{AB}^0 = -RT \ln \left[ \frac{C^0 Z_{B'}^{trans} Z_{B'}^{rot}}{8\pi^2 Z_A Z_B} \right], \quad (3.130)$$

where

$$Z_{B'}^{trans} = \int e^{-\beta[U(x_1, x_2, x_3) + W(x_1, x_2, x_3)]} dx_1 dx_2 dx_3 \quad (3.131)$$

and

$$Z_{B'}^{rot} = \int e^{-\beta[U(\xi_1, \xi_2, \xi_3) + W(\xi_1, \xi_2, \xi_3)]} d\xi_1 d\xi_2 d\xi_3. \quad (3.132)$$

As a first-order approximation, we assume that the energetic landscape of each species has an energy and a volume (entropy),

$$Z_A = \int e^{-\beta[U(r_A) + W(r_A)]} dr_A \approx Z_A^{int} e^{-\beta \langle E_A \rangle}, \quad (3.133)$$

where  $\langle E_A \rangle = \langle U(r_A) + W(r_A) \rangle$ . We further assume (*how many approximations we have taken!*) that  $Z_A^{int} Z_B^{int} \approx Z_{AB}^{int}$ , then

$$\Delta G_{AB}^0 = -RT \ln \left( \frac{C^0 Z_{B'}^{trans} Z_{B'}^{rot}}{8\pi^2} \right) + (\langle E_{AB'} \rangle - \langle E_A \rangle - \langle E_B \rangle). \quad (3.134)$$

The bound substrate's translational configuration integral,  $Z_{B'}^{trans}$ , can be conceptually linked to the volume of space that its center of mass occupies through the simulation. The effective volume was measured based on the assumption that the translational motion is restrained by three harmonic potential. By solving eigenstates of the center-of-mass covariance matrix, the eigenvalues describe the variance  $\Delta x_i^2$  along each principal axis. Thus, the translational configuration integral can be calculated as

$$\begin{aligned} Z_{B'}^{trans} &= \int e^{(-k_1 \Delta x_1^2 / 2k_B T)} dx_1 \int e^{(-k_2 \Delta x_2^2 / 2k_B T)} dx_2 \int e^{(-k_3 \Delta x_3^2 / 2k_B T)} dx_3 \\ &= (2\pi)^{3/2} \left( \langle \Delta x_1^2 \rangle \langle \Delta x_2^2 \rangle \langle \Delta x_3^2 \rangle \right)^{1/2} \end{aligned} \quad (3.135)$$

where

$$k_i = \frac{k_B T}{\langle \Delta x_i^2 \rangle}. \quad (3.136)$$

The rotational configuration integral can be accounted in a similar manner.

## Appendix A

# Optimal Means of Data Set

Suppose we have  $N$  measurements of a quantity  $x$ , which are denoted as  $\{x_i\}$ , with  $i = 1, \dots, N$ . Each measurement has a variance  $\delta^2 x_i$ . To find the optimal mean of this data set, we first write the mean of  $\{x_i\}$  as a weighted average of them

$$\bar{x} = \sum_{i=1}^N a_i x_i, \quad (\text{A.1})$$

in which  $a_i$  are the normalized weights, i.e.

$$\sum_{i=1}^N a_i = 1. \quad (\text{A.2})$$

According to the error propagation rule, if the measurements are independent, the variance of the mean  $\bar{x}$  can be written as

$$\delta^2 \bar{x} = \sum_{i=1}^N a_i^2 \delta^2 x_i. \quad (\text{A.3})$$

Minimizing  $\delta^2 \bar{x}$  with respect to  $a_i$  under the constraint of Eq. A.2 using the Lagrange multiplier  $\lambda$ , we find

$$\begin{aligned} \frac{\partial L}{\partial a_j} &= \frac{\partial}{\partial x_j} \left[ \sum_{i=1}^N a_i^2 \delta^2 x_i + \lambda \left( 1 - \sum_{i=1}^N a_i \right) \right] \\ &= 2a_j \delta^2 x_j - \lambda \\ &= 0 \end{aligned} \quad (\text{A.4})$$

for all  $x_j$ . It can be easily identified that  $a_j$  is inversed proportional to  $\delta^2 x_i$ , i.e.

$$a_j = \frac{\delta^2 x_j^{-1}}{\sum_{i=1}^N \delta^2 x_j^{-1}}, \quad (\text{A.5})$$

and

$$\bar{x} = \sum_{i=1}^N \frac{\delta^2 x_j^{-1}}{\sum_{i=1}^N \delta^2 x_j^{-1}} x_i, \quad (\text{A.6})$$

## Appendix B

# Statistical Uncertainty in the Estimator for Correlated Time Series Data

*“It is not the estimate or the forecast that matters so much as the degree of confidence with the opinion.”*

– Nassim Nicholas Taleb

Suppose we have a time series of correlated sequential observations of the randomly sampled variable  $X$  denoted as  $\{x_n\}, n = 1, \dots, N$  that come from a stationary, time-reversible stochastic process. The expectation of  $X$  can be estimated as the time average of the samples

$$\hat{X} = \frac{1}{N} \sum_{n=1}^N x_n. \quad (\text{B.1})$$

Because of the existence of correlation among the samples, the statistical uncertainty for the expectation, which is defined as

$$\delta^2 \hat{X} \equiv \left\langle \left( \hat{X} - \langle \hat{X} \rangle \right)^2 \right\rangle = \langle \hat{X}^2 \rangle - \langle \hat{X} \rangle^2, \quad (\text{B.2})$$

is complicated. We first take Eq. B.1 into Eq. B.2, and split the sum into one term capturing the variance in the observations and a remaining term capturing the correlation between the observations as

$$\begin{aligned} \delta^2 \hat{X} &= \frac{1}{N^2} \sum_{n,n'=1}^N [\langle x_n x_{n'} \rangle - \langle x_n \rangle \langle x_{n'} \rangle] \\ &= \frac{1}{N^2} \sum_{n=1}^N [\langle x_n^2 \rangle - \langle x_n \rangle^2] + \frac{1}{N^2} \sum_{n \neq n'=1}^N [\langle x_n x_{n'} \rangle - \langle x_n \rangle \langle x_{n'} \rangle] \end{aligned} \quad (\text{B.3})$$

Because of the stationarity, it becomes

$$\begin{aligned}\delta^2 \hat{X} &= \frac{1}{N} \left[ \langle x_n^2 \rangle - \langle x_n \rangle^2 \right] \\ &\quad + \frac{1}{N^2} \sum_{t=1}^{N-1} (N-t) [\langle x_n x_{n+t} \rangle + \langle x_{n+t} x_n \rangle - \langle x_n \rangle \langle x_{n+t} \rangle - \langle x_{n+t} \rangle \langle x_n \rangle] \end{aligned} \quad (\text{B.4})$$

and because of the time-reversibility, it can be further simplified to

$$\begin{aligned}\delta^2 \hat{X} &= \frac{1}{N} \left[ \langle x_n^2 \rangle - \langle x_n \rangle^2 \right] \\ &\quad + \frac{2}{N} \sum_{t=1}^{N-1} \left( \frac{N-t}{N} \right) [\langle x_n x_{n+t} \rangle - \langle x_n \rangle \langle x_{n+t} \rangle] \\ &\equiv \frac{\sigma_x^2}{N} (1 + 2\tau) = \frac{\sigma_x^2}{N/g}, \end{aligned} \quad (\text{B.5})$$

where  $\sigma_x^2$ , statistical inefficiency  $g$ , and autocorrelation time  $\tau$  (in units of the sampling interval) are given by

$$\sigma_x^2 \equiv \langle x_n^2 \rangle - \langle x_n \rangle^2 \quad (\text{B.6})$$

$$\tau \equiv \sum_{t=1}^{N-1} \left( \frac{N-t}{N} \right) C_t \quad (\text{B.7})$$

$$C_t = \frac{\langle x_n x_{n+t} \rangle - \langle x_n \rangle \langle x_n \rangle}{\sigma_x^2} \quad (\text{B.8})$$

$$g \equiv 1 + 2\tau \quad (\text{B.9})$$

The quantity  $g \equiv 1 + 2\tau > 1$  can be thought of as a statistical inefficiency, in that  $N/g$  gives the effective number of *uncorrelated* configurations contained in the time series.

## Appendix C

# MBAR returns to BAR When Only Two States Are Considered

When there are only two states, the free energy for the 1st state in MBAR becomes

$$\begin{aligned} f_1 &= -\beta_1^{-1} \ln \sum_{n=1}^N \frac{\exp(-\beta_1 U_1(\mathbf{R}_n))}{\sum_{k=1}^2 N_k \exp(\beta_k f_k - \beta_k U_k(\mathbf{R}_n))} \\ &= -\beta_1^{-1} \ln \sum_{j=1}^2 \sum_{n=1}^{N_j} \frac{\exp(-\beta_1 U_1(\mathbf{R}_{jn}))}{\sum_{k=1}^2 N_k \exp(\beta_k f_k - \beta_k U_k(\mathbf{R}_{jn}))}, \end{aligned} \quad (\text{C.1})$$

or equivalently we have

$$1 = \sum_{n=1}^N \frac{\exp(\beta_1 f_1 - \beta_1 U_1(\mathbf{R}_n))}{N_1 \exp(\beta_1 f_1 - \beta_1 U_1(\mathbf{R}_n)) + N_2 \exp(\beta_2 f_2 - \beta_2 U_2(\mathbf{R}_n))}, \quad (\text{C.2})$$

$$\begin{aligned} N_1 &= \sum_{n=1}^{N_1} \frac{1}{1 + \frac{N_2}{N_1} \exp(\Delta f - \Delta U(\mathbf{R}_{1n}))} \\ &\quad + \sum_{n=1}^{N_2} \frac{1}{1 + \frac{N_2}{N_1} \exp(\Delta f - \Delta U(\mathbf{R}_{2n}))} \end{aligned} \quad (\text{C.3})$$

where  $\Delta f = \beta_2 f_2 - \beta_1 f_1$  and  $\Delta U = \beta_2 U_2 - \beta_1 U_1$ . We further define  $M = -\ln \frac{N_2}{N_1}$ , then

$$\begin{aligned}
N_1 &= \sum_{n=1}^{N_1} \frac{1}{1 + \exp(\Delta f - \Delta U(\mathbf{R}_{1n}) - M)} \\
&\quad + \sum_{n=1}^{N_2} \frac{1}{1 + \exp(\Delta f - \Delta U(\mathbf{R}_{2n}) - M)} \\
0 &= \sum_{n=1}^{N_1} \left[ \frac{1}{1 + \exp(\Delta f - \Delta U(\mathbf{R}_{1n}) - M)} - 1 \right] \\
&\quad + \sum_{n=1}^{N_2} \frac{1}{1 + \exp(\Delta f - \Delta U(\mathbf{R}_{2n}) - M)} \\
\sum_{n=1}^{N_1} \frac{1}{1 + \exp(-\Delta f + \Delta U(\mathbf{R}_{1n}) + M)} &= \sum_{n=1}^{N_2} \frac{1}{1 + \exp(\Delta f - \Delta U(\mathbf{R}_{2n}) - M)} \\
\sum_{n=1}^{N_1} f(-\Delta f + \Delta U(\mathbf{R}_{1n}) + M) &= \sum_{n=1}^{N_2} f(\Delta f - \Delta U(\mathbf{R}_{2n}) - M) \\
N_1 \langle f(-\Delta f + \Delta U(\mathbf{R}_{1n}) + M) \rangle_1 &= N_2 \langle f(\Delta f - \Delta U(\mathbf{R}_{2n}) - M) \rangle_2 \\
\frac{\langle f(\Delta f - \Delta U(\mathbf{R}_{2n}) - M) \rangle_2}{\langle f(-\Delta f + \Delta U(\mathbf{R}_{1n}) + M) \rangle_1} &= \frac{N_1}{N_2},
\end{aligned}$$

which is Eq. 3.49.



## Appendix D

# MBAR is a binless form of WHAM

Maybe you have already noticed that MBAR and WHAM have very similar forms for the free energy. So you may want to ask if there is any relationship between MBAR and WHAM. The answer is YES. MBAR is a binless form of WHAM.[23] Let us follow Zhang et al[24] and rewrite Eq. 3.71 into an integral form

$$f_i = -\ln \int \Omega \exp(-\beta_i U) dU. \quad (\text{D.1})$$

Taking Eq. 3.70 into Eq. D.1, we find

$$f_i = -\ln \int \frac{\sum_{k=1}^K H_k(U) \exp(-\beta_i U)}{\sum_{k=1}^K N_k \exp(f_k - \beta_k U)} dU, \quad (\text{D.2})$$

where  $g_{mk}^{-1}$  has been omitted and  $H_{mk}$  has been changed to continuous form  $H_k(U)$ . From the definition,

$$H_k(U) = \sum_{\mathbf{R}}^{(k)} \delta(U(\mathbf{R}) - U). \quad (\text{D.3})$$

Taking Eq. D.3 into Eq. D.2, we have

$$\begin{aligned} f_i &= -\ln \sum_{k=1}^K \sum_{\mathbf{R}}^{(k)} \frac{\exp(-\beta_i U(\mathbf{R}))}{\sum_{k=1}^K N_k \exp(f_k - \beta_k U(\mathbf{R}))} \\ &= -\ln \sum_{\mathbf{n}=1}^N \frac{\exp(-\beta_i U_i(\mathbf{R}_{\mathbf{n}}))}{\sum_{k=1}^K N_k \exp(f_k - \beta_k U_k(\mathbf{R}_{\mathbf{n}}))}, \end{aligned} \quad (\text{D.4})$$

which is Eq. 3.108.



# Bibliography

- [1] Niels Hansen and Wilfred F. van Gunsteren. Practical aspects of free-energy calculations: A review. *J. Chem. Theory Comput.*, 10(7):2632–2647, 2014.
- [2] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314(1):141–151, 1999.
- [3] Yilin Meng and Adrian E. Roitberg. Constant ph replica exchange molecular dynamics in biomolecules using a discrete protonation model. *J. Chem. Theory Comput.*, 6(4):1401–1412, 2010.
- [4] Juyong Lee, Benjamin T. Miller, Ana Damjanović, and Bernard R. Brooks. Constant ph molecular dynamics in explicit solvent with enveloping distribution sampling and hamiltonian exchange. *J. Chem. Theory Comput.*, 10(7):2738–2750, 2014.
- [5] G.M. Torrie and J.P. Valleau. Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella Sampling. *J. Comput. Phys.*, 23(2):187–199, 1977.
- [6] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.*, 99(20):12562–12566, 2002.
- [7] Lianqing Zheng, Mengen Chen, and Wei Yang. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. U. S. A.*, 105(51):20227–20232, 2008.
- [8] Clara D. Christ and Wilfred F. van Gunsteren. Enveloping distribution sampling: A method to calculate free energy differences from a single simulation. *J. Chem. Phys.*, 126(18):184110, 2007.
- [9] Robert W. Zwanzig. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.*, 22(8):1420, 1954.
- [10] John G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3(5):300–313, 1935.

- [11] Tiziana Z. Mordasini and J. Andrew McCammon. Calculations of Relative Hydration Free Energies: A Comparative Study Using Thermodynamic Integration and an Extrapolation Method Based on a Single Reference State. *J. Phys. Chem. B*, 104(2):360–367, 2000.
- [12] Charles H. Bennett. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.*, 22(2):245–268, 1976.
- [13] Gavin E. Crooks. Path-ensemble Averages in Systems Driven Far From Equilibrium. *Phys. Rev. E*, 61:2361–2366, 2000.
- [14] Michael R. Shirts, Eric Bair, Giles Hooker, and Vijay S. Pande. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys. Rev. Lett.*, 91:140601, 2003.
- [15] John D. Chodera, William C. Swope, Jed W. Pitera, Chaok Seok, and Ken A. Dill. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *J. Chem. Theory Comput.*, 3(1):26–41, 2007.
- [16] Emilio Gallicchio, Michael Andrec, Anthony K. Felts, and Ronald M. Levy. Temperature Weighted Histogram Analysis Method, Replica Exchange, and Transition Paths. *J. Phys. Chem. B*, 109(14):6722–6731, 2005.
- [17] M Souaille and B Roux. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.*, 135:40–57, 2001.
- [18] Michael R. Shirts. Reweighting from the mixture distribution as a better way to describe the Multistate Bennett Acceptance Ratio. arXiv.org, 2017. <https://arxiv.org/abs/1704.00891>.
- [19] Eric Darve and Andrew Pohorille. Calculating free energies using average force. *J. Chem. Phys.*, 115(20):9169–9183, 2001.
- [20] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.*, 78:2690, 1997.
- [21] G. Crooks. Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. *J. Statis. Phys.*, 90:1481, 1998.
- [22] Jessica M.J. Swanson, Richard H. Henchman, and J. Andrew McCammon. Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophys. J.*, 86(1):67–74, 2004.

- [23] Zhiqiang Tan, Emilio Gallicchio, Mauro Lapelosa, and Ronald M. Levy. Theory of Binless Multi-state Free Energy Estimation with Applications to Protein-ligand Binding. *J. Chem. Phys.*, 136(14):144102, 2012.
- [24] Cheng Zhang, Chun-Liang Lai, and B. Montgomery Pettitt. Accelerating the weighted histogram analysis method by direct inversion in the iterative subspace. *Mol. Simulat.*, 42(13):1079–1089, 2016.