10-th i-CoMSE: QM/MM and ab initio MD Summer School

QM/MM Simulation (Umbrella Sampling)

Kwangho Nam
Dept. Chemistry and Biochemistry
University of Texas at Arlington

kwangho.nam@uta.edu

Enhanced Sampling for Free Energy

E.g., Umbrella Sampling

Potential of mean force (PMF)

- Canonical Free Energy: $A = -k_B T \ln Q(N,V,T)$

  where $Q = \frac{1}{h^{3N}} \int \ldots \int e^{-H(\boldsymbol{p},\boldsymbol{r})/k_B T}\, d\boldsymbol{p}^N d\boldsymbol{r}^N$ ⇦ difficult to compute it directly using MD, MC, etc

- Potential of mean force (PMF): An effective FE as a function of collective variables/reaction coordinate(s).

  Note that it is not a free energy but shares a similar functional form.

  Let's define a partition function at a fixed value of a reaction coordinate (ξ_o).

$$Q(\xi_o) = \frac{1}{h^{3N}} \iint e^{-\frac{H(\boldsymbol{p},\boldsymbol{r})}{k_B T}}\, \delta(\xi(\boldsymbol{r}^N) - \xi_o)\, d\boldsymbol{p}^N d\boldsymbol{r}^N$$

  Then, we can define a free energy similarly,

$$W(\xi_o) = -k_B T \ln Q(\xi_o)$$

  i.e., a free energy value at a fixed value of a reaction coordinate (ξ_o).

- In practice, we define a probability density

Boltzmann probability

$$\rho(\xi_o) = \frac{Q(\xi_o)}{Q} = \frac{\iint e^{-\frac{H(\boldsymbol{p},\boldsymbol{r})}{k_B T}}\, \delta(\xi(\boldsymbol{r}^N) - \xi_o)\, d\boldsymbol{p}^N d\boldsymbol{r}^N}{\iint e^{-\frac{H(\boldsymbol{p},\boldsymbol{r})}{k_B T}}\, d\boldsymbol{p}^N d\boldsymbol{r}^N} = \frac{\int e^{-\frac{U(\boldsymbol{r})}{k_B T}}\, \delta(\xi(\boldsymbol{r}^N) - \xi_o)\, d\boldsymbol{r}^N}{\int e^{-\frac{U(\boldsymbol{r})}{k_B T}}\, d\boldsymbol{r}^N}$$

Then, $-k_B T \ln \rho(\xi_o) = -k_B T \ln Q(\xi_o) + k_B T \ln Q$

After rearranging it, $W(\xi) = -k_B T \ln Q(\xi) = -k_B T \ln \rho(\xi) + k_B T \ln Q = -k_B T \ln \rho(\xi) + C$

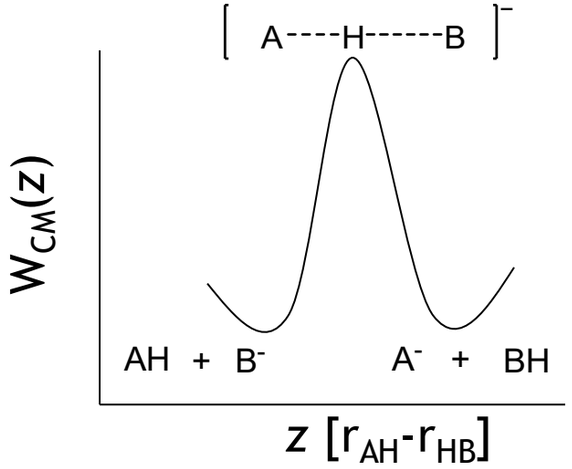
Free energy simulation methods

- Potential of mean force by umbrella sampling
  - Reaction free energy as a function of some distinguished reaction coordinate

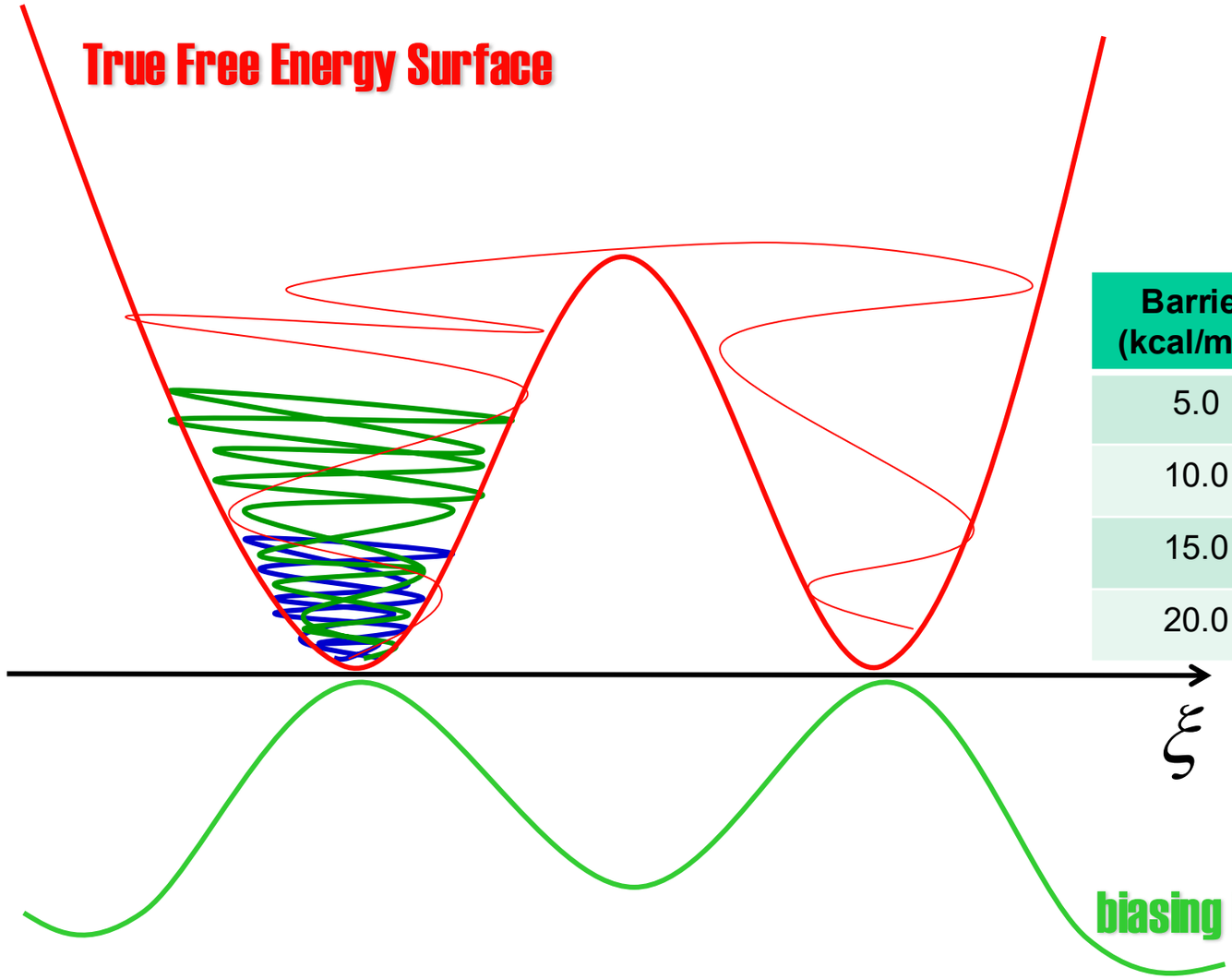
$$W_{CM}(z) = -RT \ln \rho(z) + C$$

PMF

Probability density



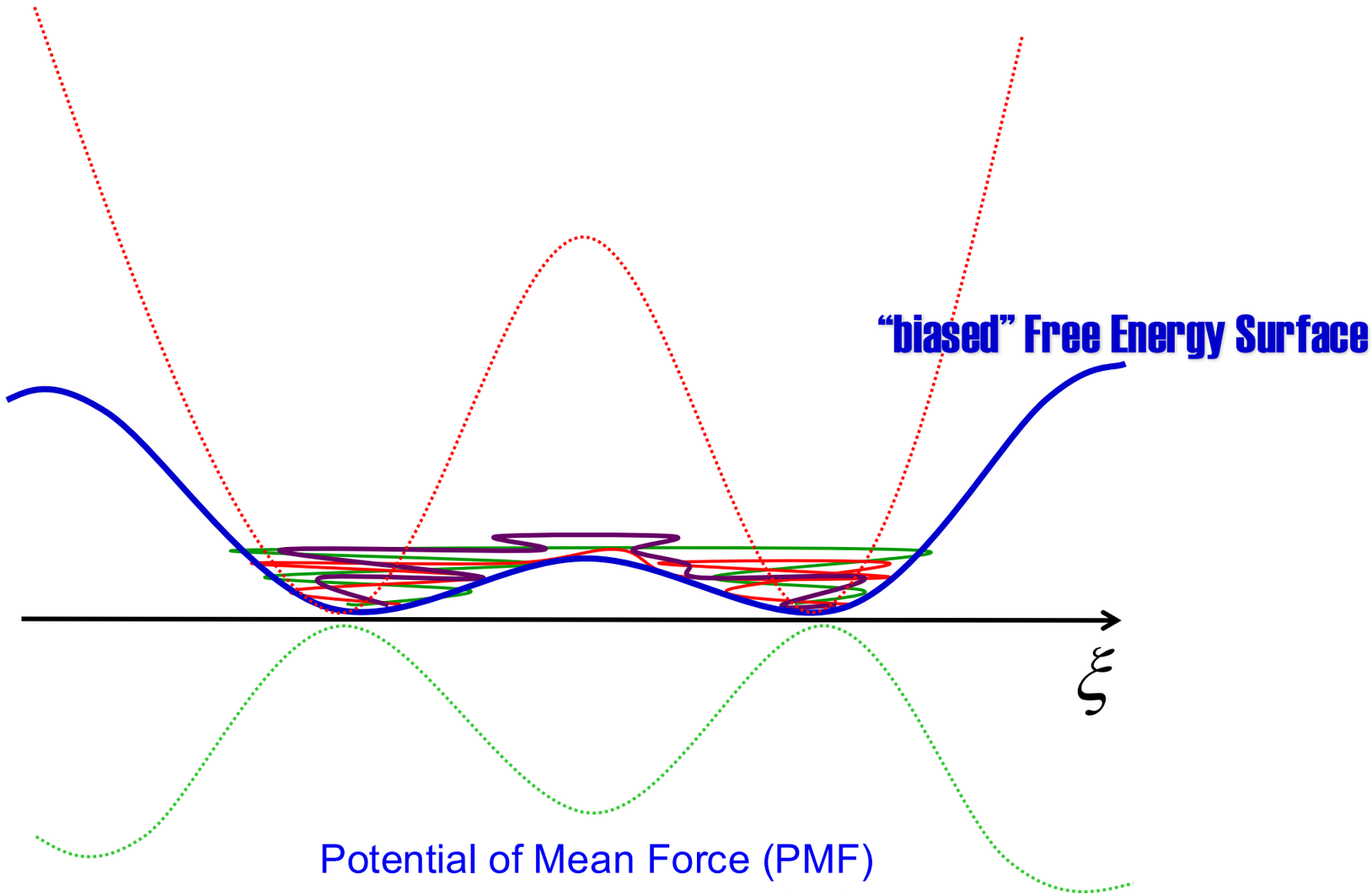
Umbrella sampling



Barrier (kcal/mol)	Rate
5.0	~1/ns
10.0	~0.3/μs
15.0	~60/s
20.0	~0.8/min

Torrie, J. Comput. Phys. 1977

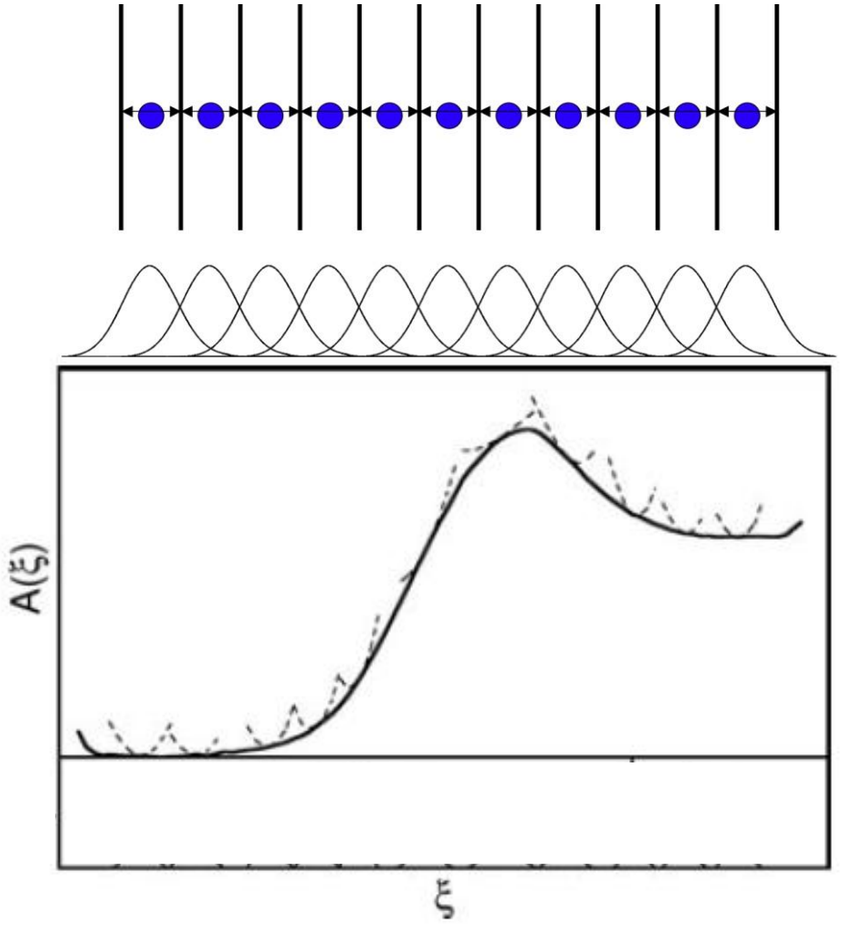
Umbrella sampling



Potential of Mean Force (PMF)

$$W(\xi) = -k_B T \ln \rho(\xi) + C$$

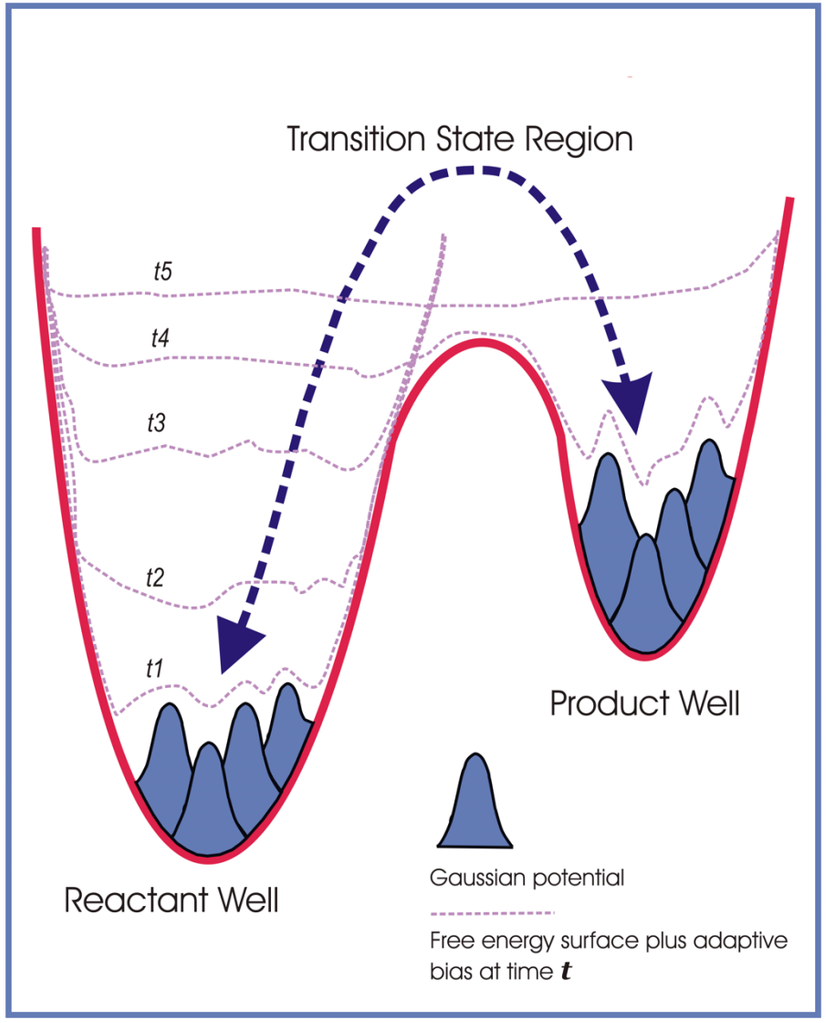
In practice



- Perform several umbrella sampling runs in (partially overlapping) windows.
  - -This is more efficient than performing a single long umbrella sampling simulations.
  - -Each window should be placed to have some overlap with neighboring windows.
- Order parameter (or reaction coordinate), ξ, describe the transition from state A to state B.
- After the simulation, the final potential of mean force surface is obtained by using the weighted histogram analysis method (WHAM, Kumar, 1992) or manually fitting the surface between neighboring windows.

Other FE methods: e.g., meta-dynamics

Similar to the adaptive umbrella sampling, except that it adaptively obtain the biasing potential as a sum of Gaussians.



$$G(s,t) = \sum_{t_i} h \exp\left(- \frac{|s - s(t_i)|^2}{2w_1^2} \right)$$

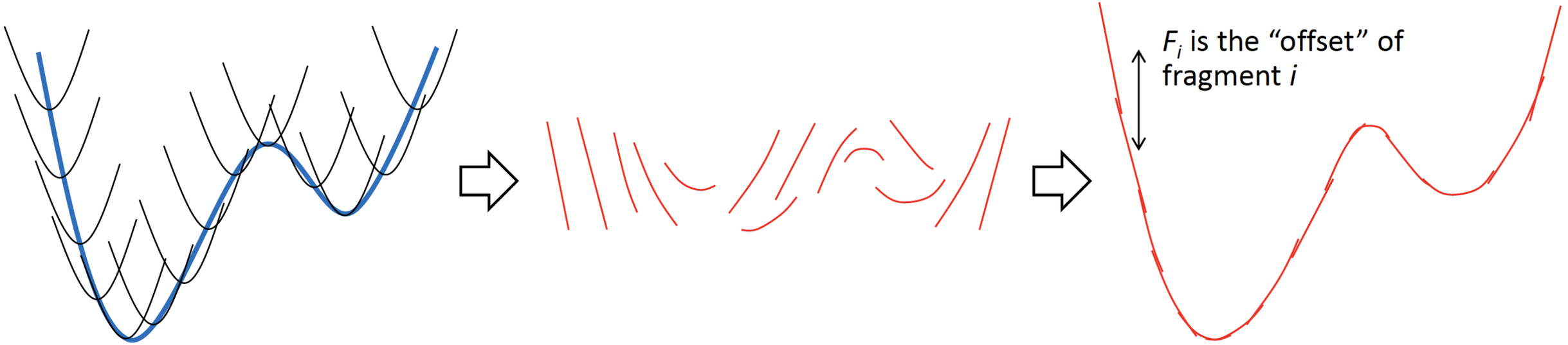
A. Laio and M. Parrinello, *PNAS*, 2002

PMF calculation by umbrella sampling

1) Umbrella sampling: at total N_s different windows/simulations

2) Potential of mean force (PMF): free energy along the reaction coordinate, thus it has one less degrees of freedom.

$$W(\xi_i) = -k_B T \ln \rho^o(\xi_i) + C \qquad \text{Eq.(1)}$$



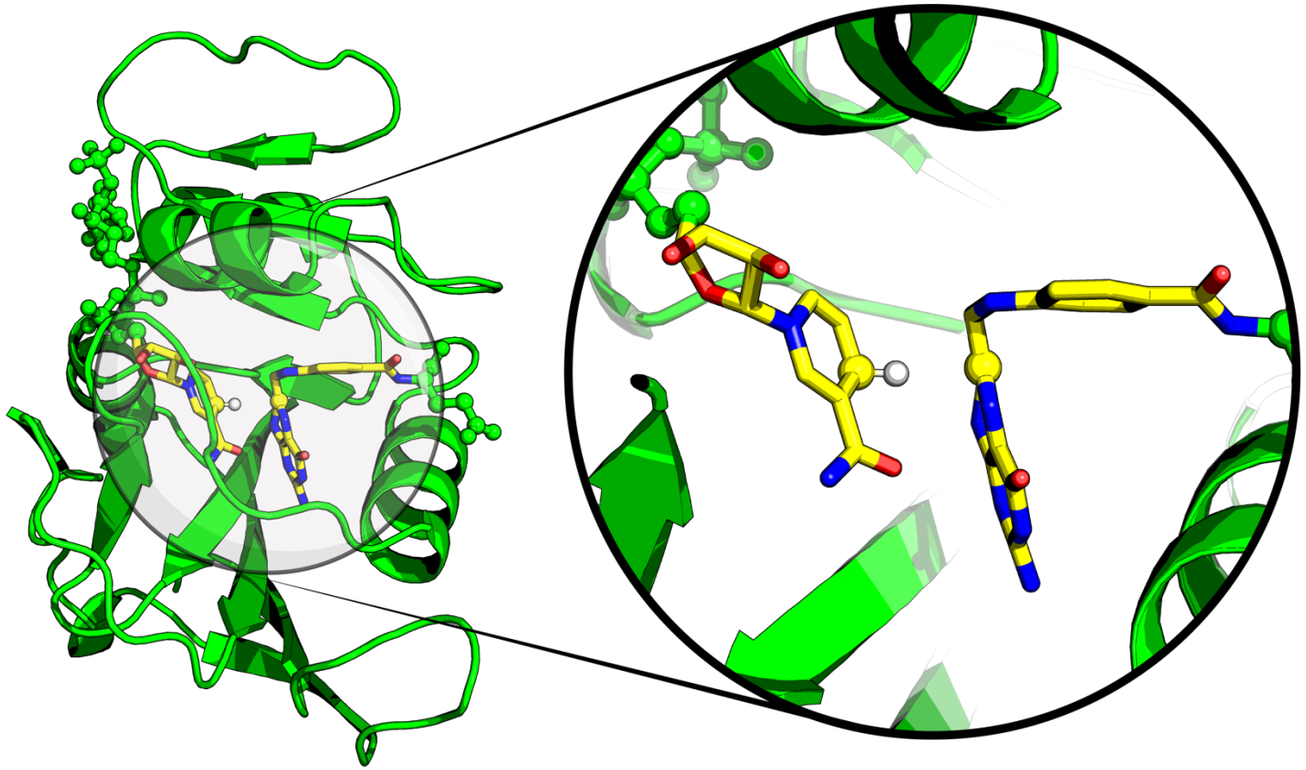
FES+Bias
(Many US simulations)
Unknown FES

Each US simulation gives a fragment of the (unbiased) FES

Reconstructed (unbiased) FES

How to align?

Hands-on practice: Reaction catalyzed by dihydrofolate reductase (DHFR)

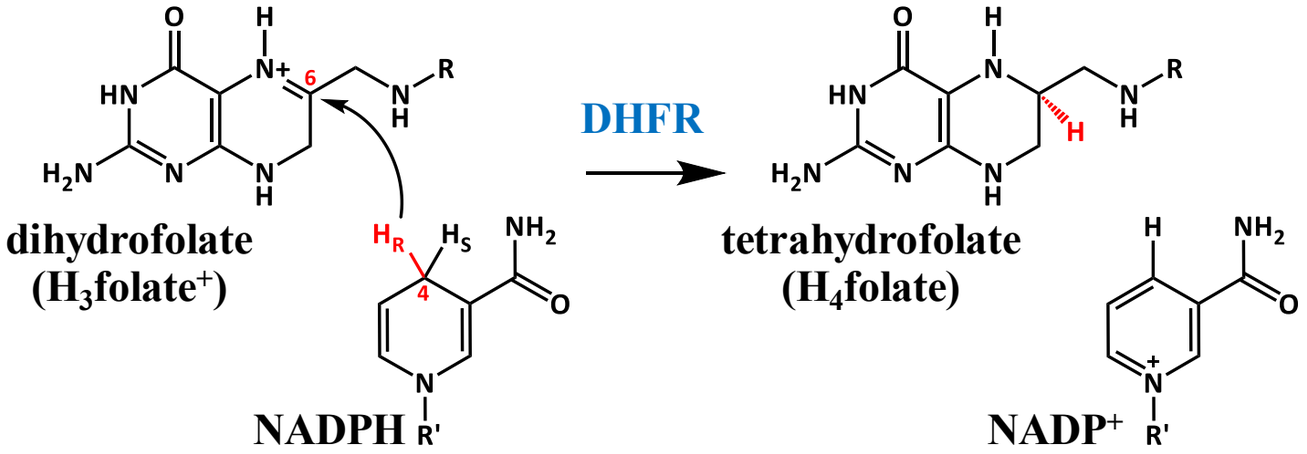


PDB ID: 1RX2

- Task: Simulating the hydride transfer reaction catalyzed by dihydrofolate reductase (DHFR)

  -System prep: CHARMM-GUI QM/MM Interfacer
  -39 umbrella sampling windows along reaction coordinate (ξ between -1.9 and 1.9 Å with an 0.1 Å interval)
  -SE-QM: AM1 model

Interconverting between:
NADPH + $H_3folate^+$ and $NADP^+$ + $H_4folate$



Analysis & Computing PMF profile

Weighted Histogram Analysis Method (WHAM)

PMF calculation

1) Umbrella sampling: at total N_s different windows/simulations

2) Potential of mean force (PMF): free energy along the reaction coordinate, thus it has one less degrees of freedom.

$$W(\xi_i) = -k_B T \ln \rho^o(\xi_i) + C \qquad \text{Eq.(1)}$$

3) PMF calculations using Weighted Histogram Analysis Method (WHAM)

$$\rho^o(\xi_i) = \frac{\sum_{k=1}^{N_s} n_k(\xi_i)}{\sum_{k'=1}^{N_s} N_{k'} e^{-\beta(U_{k'}(\xi_i) - f_{k'})}} \qquad \text{Eq.(2)}$$

$$e^{-\beta f_k} = \sum_{i=1}^{N_w} \rho^o(\xi_i) e^{-\beta U_k(\xi_i)} \qquad \text{Eq.(3)}$$

* Iterate the Eq. (2) and (3) over all N_s simulations until convergence in the free energies, f_k, is obtained.

N_s: No. of simulations; N_k: total samples in k-th simulation; f_k: free energy offset for k-th simulation; $U_k(\xi_i)$: Umbrella/biasing potential of k-th simulation at bin ξ_i; N_w: No. of bins; $n_k(\xi_i)$: no. samples in i-th bin from k-th simulation ($N_k = \sum_i n_k(\xi_i)$); ξ_i: the reaction coordinate value at the center of i-th bin.

Practical issues

- ## Histogram overlap

  -Assumption: Sampling is complete along all other degrees of freedom.

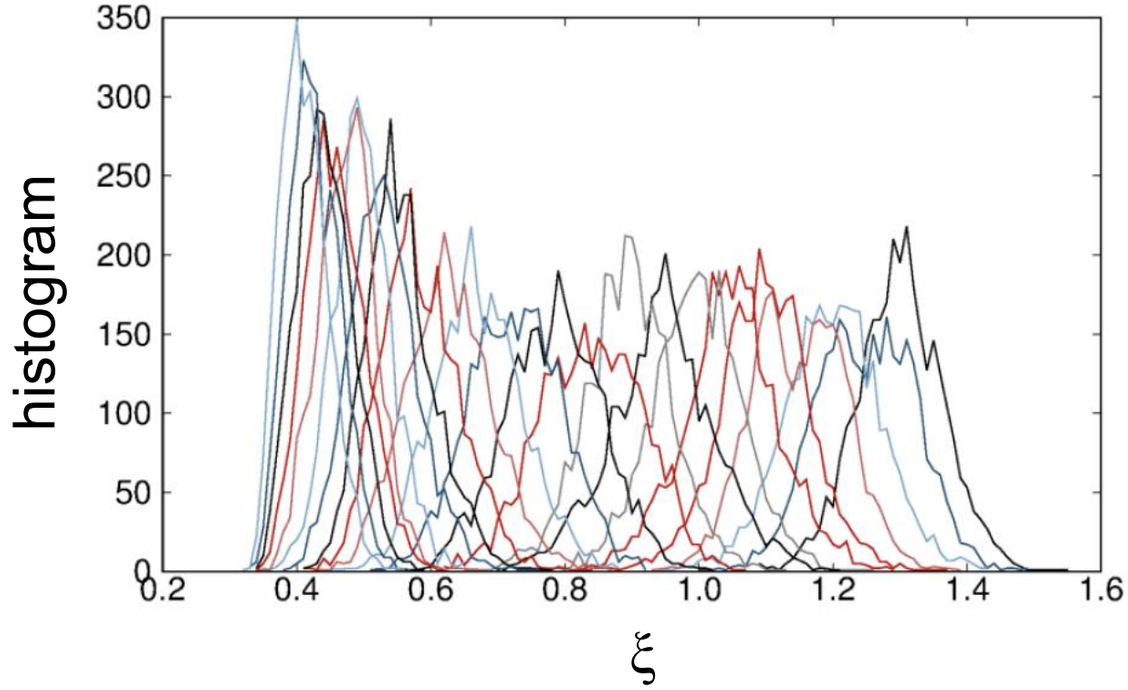
  -Then, we can match two neighboring windows (by finding relevant F_i values for offsetting).

- ## Shapes of computed PMF

  -Histogram overlap is a minimal requirement and does not guarantee the correctness of the computed FES.

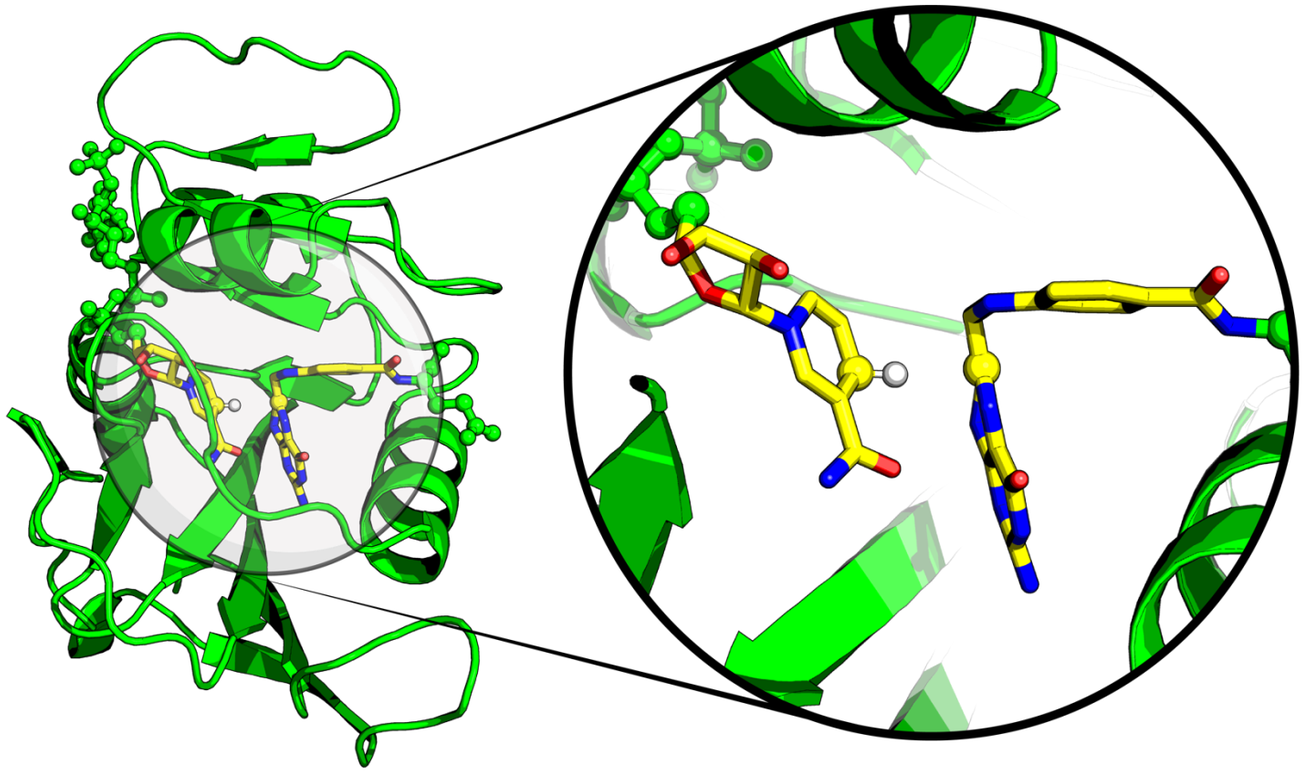
  -A few considerations: (1) does each (fragmented) FES vary smoothly over ξ?; (2) is there any abnormality in geometry along ξ?; etc.

- ## Other analysis?

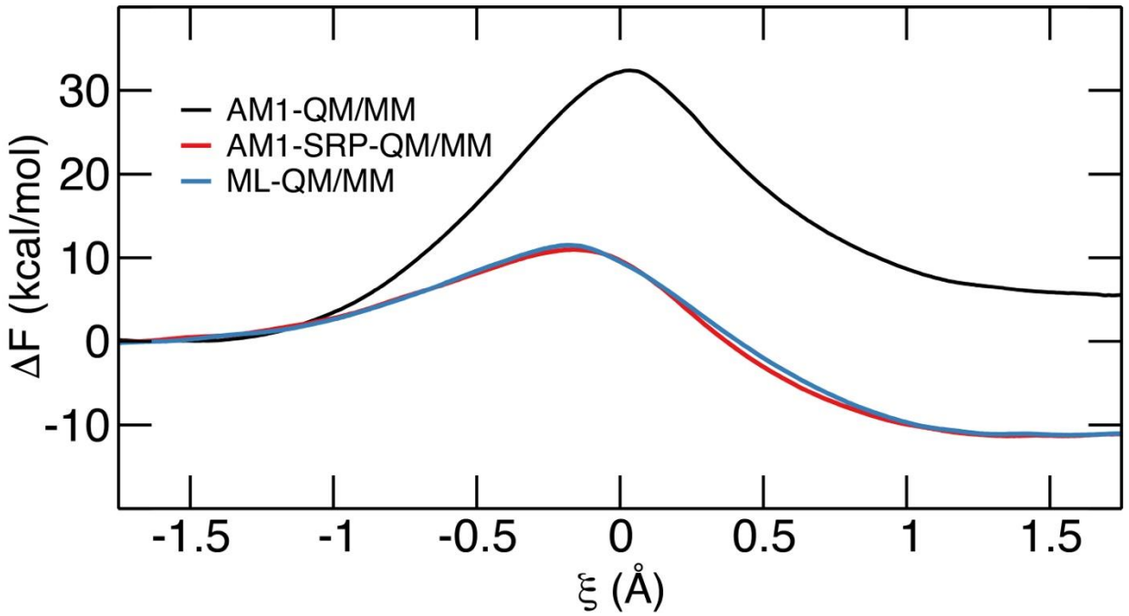


Can we align windows that do not overlap?

Reaction catalyzed by dihydrofolate reductase (DHFR)



PDB ID: 1RX2



Interconverting between:

NADPH + $H_3folate^+$ and $NADP^+$ + $H_4folate$