

# Molecular Modelling and Simulations

Shampa Raghunathan

## MD simulations so far...

### Algorithm 3 (A Simple Molecular Dynamics Program)

program md	simple MD program
call init	initialization
t=0	
do while (t.lt.tmax)	MD loop
call force(f,en)	determine the forces
call integrate(f,en)	integrate equations of motion
t=t+delt	
call sample	sample averages
enddo	
stop	
end	

---

Frenkel, Daan, and Berend Smit. Understanding Molecular Simulation from Algorithms to Applications (2002). Academic Press.

## What is free energy?

- ▶ In thermodynamics, free energy amounts to internal energy of the system available for work, and it determines the direction of a certain thermodynamic process as well as the probability that the system will remain in a given state.
- ▶ Presently, molecular dynamics simulations are frequently used to obtain free energies of molecular systems.

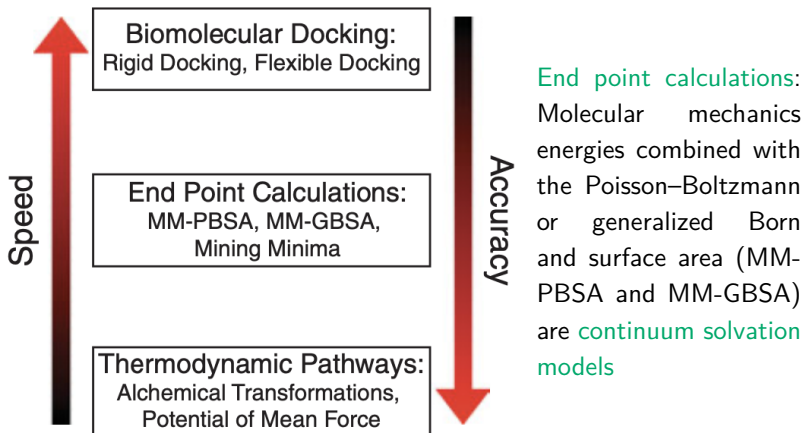
## Why do we need to calculate this?

- ▶ Free energy of a system is often considered to be one of the important quantity in thermodynamics, normally termed as Helmholtz ( $A$ ) or Gibb's free energy ( $G$ )
- ▶ Under constant  $N$ ,  $P$ ,  $T$ –Gibb's free energy, and constant  $N$ ,  $V$ ,  $T$ –Helmholtz free energy
- ▶ Very useful task for studying molecular recognition in biochemical processes; development of therapeutics for the treatment of disease; phase transitions
- ▶ Calculating absolute free energies not possible (as an appropriate reference state is unknown); instead we compute the change in free energy of a bimolecular process, such as, binding of a ligand to a protein

# Outline

- 1 **Calculating free energies**
  - Thermodynamic perturbation
  - Thermodynamic integration

## Methods for calculating free energies



## Calculating free energies: Thermodynamic pathways

- ▶ The Helmholtz free energy,  $A$ , is related to the canonical partition function by

$$A = -k_B T \ln Q_{NVT} \quad \text{where,} \quad Q_{NVT} = \sum_{\alpha} e^{-\beta E_{\alpha}} \quad (1)$$

- ▶ Dropping  $NVT$  to simplify the notation, writing in a general form of the partition function  $Q$ , and formulating it as a phase space integral over all spatial ( $\mathbf{r}$ ) and momentum ( $\mathbf{p}$ ) coordinates:

$$Q = \int \int e^{-\beta E(\mathbf{r}, \mathbf{p})} d\mathbf{r} d\mathbf{p} \quad (2)$$

This assumes that the energy levels, computed as the sum of kinetic and potential energy terms are sufficiently closely spaced that we may convert the sum-over-states formulation of  $Q$  (see Eq. 1) into an integral.

- ▶ Directly evaluating such integrals using MD requires more substantial sampling over higher-energy configurations

## Thermodynamic perturbation

Typically not interested in absolute values of thermodynamic functions but in their changes over the course of a chemical process, e.g., the difference in  $E$  for the proton shift reaction  $\text{HCN} \rightarrow \text{HNC}$  in aqueous solution. Because  $E$  is a state function, the precise path over which the reaction occurs is not important—we need only to evaluate  $U$  at the reactant and product to determine the difference.

- ▶ Let's consider two states  $X$  (reference) and  $Y$  (target)

$$E_Y(\mathbf{r}, \mathbf{p}) = \Delta E(\mathbf{r}, \mathbf{p}) + E_X(\mathbf{r}, \mathbf{p}) \quad (3)$$

Assuming that  $\Delta E(\mathbf{r}, \mathbf{p})$  consists of all terms in  $E_Y(\mathbf{r}, \mathbf{p})$  which describe solvent-solute interactions.

- ▶ Free energy differences,  $\Delta A$ , between two systems or states  $X$  and  $Y$ , described by the partition functions  $Q_X$  and  $Q_Y$ , respectively can be written as

$$\Delta A = -k_B T \ln \frac{Q_Y}{Q_X} \quad (4)$$



## Thermodynamic perturbation

$$\Delta A = -k_B T \ln \left( \frac{\int \int e^{-\beta E_Y(\mathbf{r}, \mathbf{p})} d\mathbf{r} d\mathbf{p}}{\int \int e^{-\beta E_X(\mathbf{r}, \mathbf{p})} d\mathbf{r} d\mathbf{p}} \right) \quad (5)$$

Replacing  $E_Y(\mathbf{r}, \mathbf{p})$  from Eq. 3

$$\Delta A = -k_B T \ln \left( \frac{\int \int e^{-\beta \Delta E(\mathbf{r}, \mathbf{p})} e^{-\beta E_X(\mathbf{r}, \mathbf{p})} d\mathbf{r} d\mathbf{p}}{\int \int e^{-\beta E_X(\mathbf{r}, \mathbf{p})} d\mathbf{r} d\mathbf{p}} \right) \quad (6)$$

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

$$P_X(\mathbf{r}, \mathbf{p}) = \frac{e^{-\beta E_X(\mathbf{r}, \mathbf{p})}}{\int \int e^{-\beta E_X(\mathbf{r}, \mathbf{p})} d\mathbf{r} d\mathbf{p}} \quad (7)$$

Now, using  $P_X$ , Eq. 6 becomes

$$\Delta A = -k_B T \ln \int \int e^{-\beta \Delta E(\mathbf{r}, \mathbf{p})} P_X(\mathbf{r}, \mathbf{p}) d\mathbf{r} d\mathbf{p} \quad (8)$$

or equivalently

$$\Delta A = -k_B T \ln \langle e^{-\beta \Delta E(\mathbf{r}, \mathbf{p})} \rangle_X \quad (9)$$

## Thermodynamic perturbation

- ▶ Here,  $\langle \cdots \rangle_X$  denotes an ensemble average over configurations sampled from the reference state. This is the fundamental **free energy perturbation (FEP)** formula, which is the basis for all further developments.
- ▶  $\Delta A$  can be estimated for the **forward process** by sampling only equilibrium configurations of the reference state  $X$ :

$$\Delta A = -k_B T \ln \langle \exp[-\beta(E_Y - E_X)] \rangle_X \quad (10)$$

- ▶ Equivalently we could write the **reverse process**, *i.e.*, averaging over the ensemble corresponding to the target state  $Y$ :

$$\Delta A = -k_B T \ln \langle \exp[-\beta(E_X - E_Y)] \rangle_Y \quad (11)$$

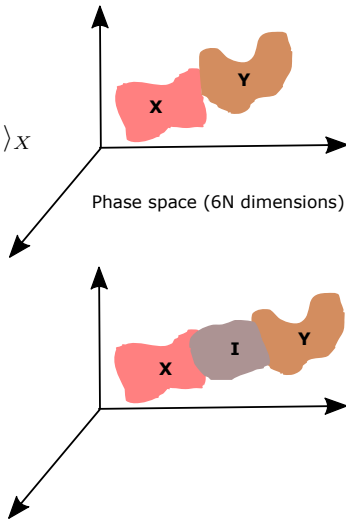
## Overlapping states

- ▶ In order to evaluate an ensemble average like

$$\Delta A = -k_B T \ln \langle \exp[-\beta(E_Y - E_X)] \rangle_X$$

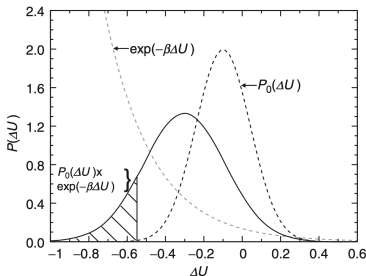
we could run a simulation of either state  $X$  or  $Y$  and collect statistics

- ▶ Problems arise, however, when two states  $X$  and  $Y$  do not overlap such that simulating one state does a poor job of sampling the other
- ▶ An intermediate ( $I$ ) state may improve the degree of overlap in phase space as well as sampling



## Intermediate states

Problem in estimating free energy differences (here,  $U$  refers  $E$ ),



**Fig. 2.1.**  $P_0(\Delta U)$ , the Boltzmann factor  $\exp(-\beta\Delta U)$  and their product, which is the integrand in (2.12). The low- $\Delta U$  tail of the integrand, marked with stripes is poorly sampled with  $P_0(\Delta U)$  and, therefore, is known with low statistical accuracy. However, it provides an important contribution to the integral

---

Chipot Ch. and Pohorille A. (2007). Free Energy Calculations: Theory and Applications in Chemistry and Biology. Springer-Verlag Berlin Heidelberg.

## Intermediate states

- ▶ If the energy difference between the two states is large  $|E_Y - E_X| \gg k_B T$  we can introduce an intermediate state between  $X$  and  $Y$

$$\begin{aligned}\Delta A &= A(Y) - A(X) \\ &= (A(Y) - A(I)) + (A(I) - A(X)) \\ &= -k_B T \ln \left[ \frac{Q_Y}{Q_I} \times \frac{Q_I}{Q_X} \right]\end{aligned}$$

## Intermediate states

- ▶ Obvious extension is to use multiple intermediate states with increasing overlap progressing from  $E_X$  to  $E_Y$

$$\begin{aligned}\Delta A &= A(Y) - A(X) \\ &= (A(Y) - A(N)) + (A(N) - A(N-1)) + \dots \\ &\quad + (A(2) - A(1)) + (A(1) - A(X)) \\ &= -k_B \ln \left[ \frac{Q_Y}{Q_N} \times \frac{Q_N}{Q_{N-1}} \times \frac{Q_{N-1}}{Q_{N-2}} \times \dots \times \frac{Q_2}{Q_1} \times \frac{Q_1}{Q_X} \right]\end{aligned}$$

- ▶ One key to this method is that intermediate states do not need to correspond to actual physical states (consider changing ethane to ethanol:  $\text{H}_3\text{C}-\text{H}_2\text{C}-\text{H} \rightarrow \text{H}_3\text{C}-\text{H}_2\text{C}-\text{O}-\text{H}$ )
- ▶ Using molecular mechanics we can interpolate between these two states

## Implementation of free energy perturbation

The relationship between initial, final, and intermediate states is usually described by **coupling parameter**  $\lambda$ . Each of the terms in the force field for an intermediate (I) state can be written as linear combinations of the values for  $X$  and  $Y$ :

Bonds:

$$k_r(\lambda) = \lambda k_r(Y) + (1 - \lambda)k_r(X)$$

$$r_0(\lambda) = \lambda r_0(Y) + (1 - \lambda)r_0(X)$$

Dihedrals:

$$k_\tau(\lambda) = \lambda k_\tau(Y) + (1 - \lambda)k_\tau(X)$$

$$\tau_0(\lambda) = \lambda \tau_0(Y) + (1 - \lambda)\tau_0(X)$$

Angles:

$$k_\theta(\lambda) = \lambda k_\theta(Y) + (1 - \lambda)k_\theta(X)$$

$$\theta_0(\lambda) = \lambda \theta_0(Y) + (1 - \lambda)\theta_0(X)$$

Electrostatics:

$$q_i(\lambda) = \lambda q_i(Y) + (1 - \lambda)q_i(X)$$

van der Waals:

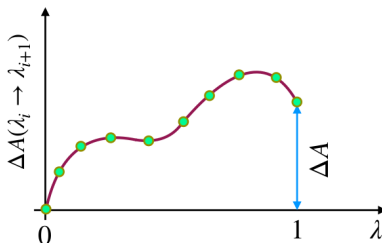
$$\epsilon(\lambda) = \lambda \epsilon(Y) + (1 - \lambda)\epsilon(X)$$

$$\sigma(\lambda) = \lambda \sigma(Y) + (1 - \lambda)\sigma(X)$$

## Coupling parameter

- ▶ As we change the coupling parameter  $\lambda$  from 0 to 1, we move from state  $X$  to  $Y$
- ▶ At each intermediate step  $\lambda_i$  we perform an MD simulation by first performing a short equilibration run (since our point of equilibrium has changed) and then a “production” run where we calculate

$$\Delta A(\lambda \rightarrow \lambda_{i+1}) = -k_B T \ln \langle \exp[-\beta \Delta E_i] \rangle \quad (12)$$



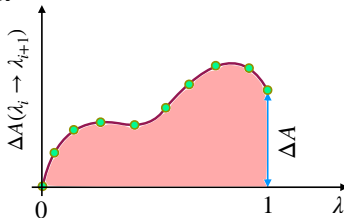


## Thermodynamic integration

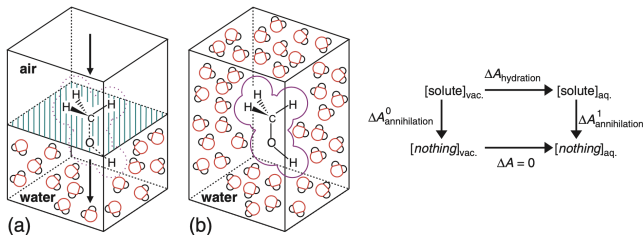
- ▶ Instead of evaluating the difference in the free energy between subsequent states, we could also calculate the derivative of the energy

$$\Delta A = \int_0^1 d\lambda \left\langle \frac{\partial E(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \quad (13)$$

- ▶ In practice, we perform a series of simulations corresponding to discrete values of  $\lambda$  between 0 and 1. For each value of  $\lambda$  the  $\langle \dots \rangle$  is calculated
- ▶ The free energy difference is the area under the curve



## Applications of methods for calculating free energy differences

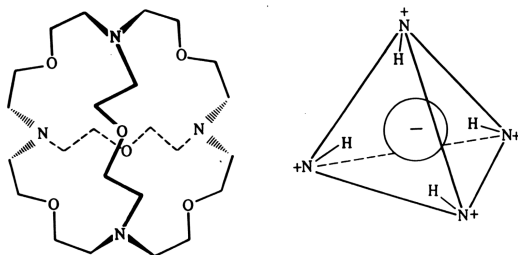


**Fig. 2.6.** The thermodynamic cycle for estimating the hydration free energy,  $\Delta A_{\text{hydration}}$ , of a small solute (the right side of the figure). One route is the direct evaluation of  $\Delta A_{\text{hydration}}$  along the upper vertical arrow. The solute, originally placed in vacuum (a) is moved to the bulk water (b). Another route consists of annihilating, or creating, the solute both in vacuo and in the aqueous medium and corresponds to the vertical lines in the thermodynamic cycle. As suggested by the cycle, these two routes are formally equivalent, as:  $\Delta A_{\text{hydration}} = \Delta A^0_{\text{annihilation}} - \Delta A^1_{\text{annihilation}}$

Instead of annihilating the solute, we can 'create' it by turning on the perturbation part of the Hamiltonian. The resulting free energy differences are connected through the relation:

$$\Delta A^1_{\text{creation}} - \Delta A^0_{\text{creation}} = \Delta A^0_{\text{annihilation}} - \Delta A^1_{\text{annihilation}}$$

## Applications of methods for calculating free energy differences



**Fig. 9.6** The SC24/halide system. Figure adapted from Lybrand T P, J A McCammon and G Wipff 1986. Theoretical Calculation of Relative Binding Affinity in Host-Guest Systems. *Proceedings of the National Academy of Sciences USA* **83**:833–835.

Experimental  $\Delta\Delta A \approx 4.3$  kcal/mol,  
Calculated value: 4.15 kcal/mol

- Selective binding of  $\text{Cl}^{-1}$  rather than  $\text{Br}^{-1}$  due to highly favorable interaction of  $\text{Cl}^{-1}$  with the receptor SC24; this effect more than compensates for the unfavorable free energy of desolvation of  $\text{Cl}^{-1}$  compared to  $\text{Br}^{-1}$

## Potential mean force

- ▶ How about considering the free energy changes as a function of some intramolecular or intermolecular coordinate?
- ▶ The free energy surface along the chosen coordinate is known as the **potential mean force** (PMF),  $W$

$$W(q) = -k_B T \ln \pi(q) \quad (14)$$

where,  $q$  is the coordinate and  $\pi$  is the probability of the coordinate taking a particular value

For example: To determine the PMF for the rotation of the central C–C bond of butane in aqueous solution, over the course of the MD simulation the torsion angle can be saved at every step (binned  $q$ ) and with good sampling a probability histogram would permit conversion to a PMF accurately reflecting the true potential