Free Energy in Molecular Dynamics Simulations

Jakub Rydzewski

Institute of Physics, Nicolaus Copernicus University, Torun, Poland

Let's meet at NCU!

International Intensive Summer and Winter Educational Meetings

Jul 15, 2021

References

▶ M. E. Tuckerman
   Statistical Mechanics: Theory and Simulation
   Oxford University Press (2016).

▶ D. Chandler
   Introduction to Modern Statistical Mechanics
   Oxford University Press (1987).

▶ M. Toda, R. Kubo, N. Saito
   Statistical Physics I: Equilibrium Statistical Mechanics
   Springer-Verlag (1983).

▶ R. Kubo, M. Toda, N. Hashitsume
   Statistical Physics II: Nonequilibrium Statistical Mechanics
   Springer-Verlag (1991).

Probability

Definition **Random Variable**

A random variable X is completely defined by the range of values it can take, and its probability distribution $p_X(x_1, \ldots, x_k)$. The value p_X is the probability that the random variable X takes the value $\mathbf{x} = (x_1, \ldots, x_k)$.

▶ According to the definition: $p_X(\mathbf{x}) \equiv \mathbb{P}[X = \mathbf{x}]$.

▶ For simplicity, we will write $p(\mathbf{x})$ to denote $p_\mathbf{X}(\mathbf{x})$.

▶ $p(\mathbf{x})$ is non-negative and satisfies the normalization condition:

$$\int d\mathbf{x} \, p(\mathbf{x}) = 1. \tag{1}$$

▶ The expectation value of f is denoted by[1]:

$$\mathbb{E}[f] = \int d\mathbf{x} \, p(\mathbf{x}) f(\mathbf{x}). \tag{2}$$

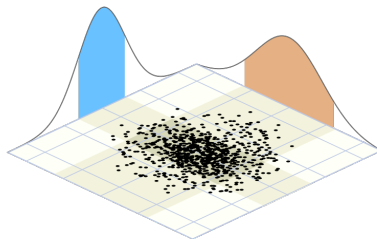[1] Depending on the context, we will use different notation for averages.

Probability

Example **Gaussian Random Variable** $X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

A continuous variable $\boldsymbol{X} \in \mathbb{R}^k$ has a Gaussian distribution of mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2$ if its probability density is:

$$p(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k \det(\boldsymbol{\Sigma})}}. \tag{3}$$

We have $\mathbb{E}[\boldsymbol{X}] = \boldsymbol{\mu}$ and $\mathrm{Var}[\boldsymbol{X}] = \Sigma$.

Ergodicity

Definition **Ergodic Hypothesis**

Ergodicity means that all states in an ensemble are reached by the time evolution of the corresponding system.

▶ Time average of $A[\mathbf{x}(t)] \equiv A_t(\mathbf{x})$:

$$\bar{A} = \lim_{t \to \infty} \frac{1}{T} \int_0^T \mathrm{d}t A_t(\mathbf{x}) \rho_t(\mathbf{x}) \tag{4}$$

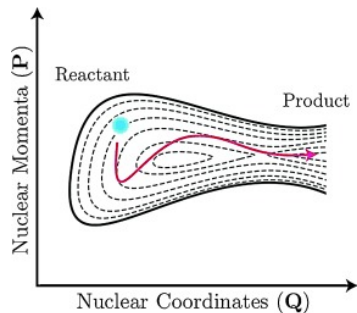▶ We define the ensemble average of a quantity $A(\mathbf{x})$ as:

$$\langle A \rangle = \frac{\int \mathrm{d}\mathbf{x}\, A(\mathbf{x})\rho(\mathbf{x})}{\int \mathrm{d}\mathbf{x}\, \rho(\mathbf{x})}, \tag{5}$$

where $\rho(\mathbf{x})$ is the phase space probability density.
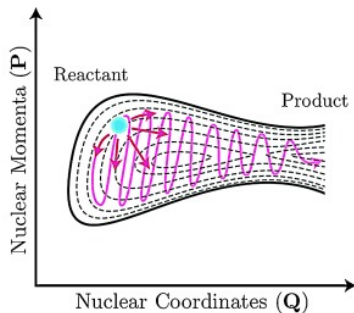
▶ Is it equal?

$$\bar{A} \stackrel{?}{=} \langle A \rangle \tag{6}$$

Ergodicity

(a) Coherent non-ergodic dynamics.

(b) Ergodic dynamics.

Canonical Ensemble

▶ Known as the *NVT* ensemble.

▶ At a given temperature T, the probability for a system to be in a certain configuration $\mathbf{x}$ with energy $E(\mathbf{x})$ is:

$$\rho(\mathbf{x}) = \frac{1}{Z} e^{-\beta E(\mathbf{x})}, \tag{7}$$

where $\beta = \frac{1}{k_{\mathrm{B}} T}$ is the *inverse temperature*[2], and the canonical partition function is:

$$Z = \int d\mathbf{x} \, e^{-\beta E(\mathbf{x})}. \tag{8}$$

▶ The ensemble average of a quantity A is given by:

$$\langle A \rangle = \frac{1}{Z} \int d\mathbf{x} \, A(\mathbf{x}) e^{-\beta E(\mathbf{x})}. \tag{9}$$

---

[2] $\beta \approx 0.4$ kJ/mol at 300 K.

$k_\mathrm{B}T$

▶ $k_\mathrm{B}T$ is used in physics as a scale factor for energy values in molecular-scale systems.

▶ Inverse of $k_\mathrm{B}T$ is called the inverse temperature and generally as $\beta = \frac{1}{k_\mathrm{B}T}$.

▶ For a system in equilibrium in canonical ensemble, the probability of the system being in state with energy U is proportional to $\mathrm{e}^{-\beta U}$.

▶ $k_\mathrm{B}T$ is the amount of heat required to increase the thermodynamic entropy of a system by k_B.

$k_\mathrm{B}T$

Approximate values of kT at 298 K	Units
$kT = 4.11 \times 10^{-21}$	J
$kT = 4.114$	pN·nm
$kT = 9.83 \times 10^{-22}$	cal
$kT = 25.7$	meV
kT=-174	dBm/Hz
Related quantities (also at 298 K)	
$kT/hc \approx 207$ [1]	cm^{-1}
$kT/e = 25.7$	mV
$RT = kT \cdot N_A = 2.479$	kJ·mol^{-1}
$RT = 0.593$	kcal·mol^{-1}
$h/kT = 0.16$	ps

Figure 1: $k_\mathrm{B}T$ in different units.[3]
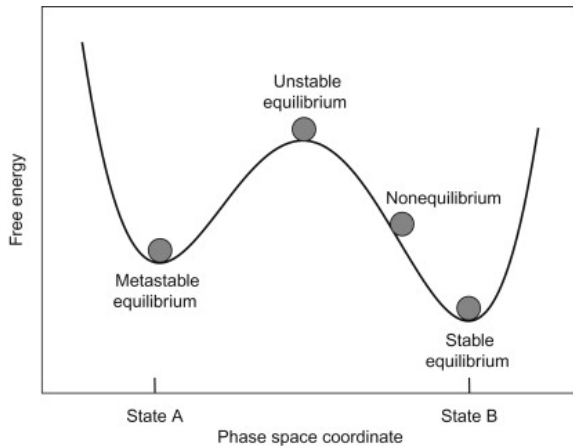
Metastable States

Figure 2: Slow transitions between metastable states.

Collective Variables

- Consider a molecular system, described by microscopic coordinates $\mathbf{x}$ and a potential energy function $U(\mathbf{x})$.

- We limit our discussion to the canonical ensemble (NVT).

- At equilibrium, the microscopic coordinates follow the Boltzmann distribution:

$$\rho(\mathbf{x}) = \frac{\mathrm{e}^{-\beta U(\mathbf{x})}}{\int \mathrm{d}\mathbf{x}\, \mathrm{e}^{-\beta U(\mathbf{x})}}, \tag{10}$$

where β is the inverse of the thermal energy.

- We identify a small set of coarse-grained order parameters called *collective variables* (CVs) with values given by:

$$\mathbf{z} = [z_1(\mathbf{x}), z_2(\mathbf{x}), \ldots, z_n(\mathbf{x})]. \tag{11}$$
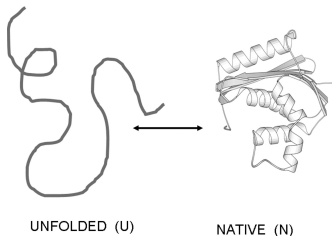
Collective Variables

Definition **Collective Variables**

Differentiable functions of the microscopic coordinates:

$$\mathbf{z} = \xi(\mathbf{x}) \equiv [\xi_1(\mathbf{x}), \xi_2(\mathbf{x}), \ldots, \xi_n(\mathbf{x})], \tag{12}$$

usually satisfying the following conditions: (1) The number of CVs is much lower than the number of particles in a system; (2) Correspondence to slow degrees of freedom; and (3) Discrimination between important metastable states.

- ▶ CVs are system-dependent.

- ▶ For instance: angles, distances, contact map, dihedral angles, index along a predefined reaction coordinate, positions, and many others.

UNFOLDED (U)   NATIVE (N)

Collective Variables
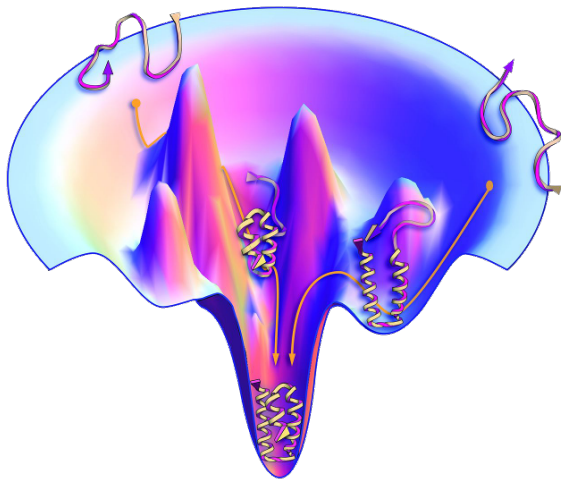
Free Energy

▶ The equilibrium marginal distribution of CVs is given by integrating out all other degrees of freedom:

$$\rho(\mathbf{z}) = \int d\mathbf{x}\, \delta[\mathbf{z} - \xi(\mathbf{x})]\rho(\mathbf{x}), \tag{13}$$

where we use the sifting property of the Dirac δ-function and ξ is a function that maps $\mathbf{x}$ to $\mathbf{z}$.

▶ Eq. 13 is equivalent to:

$$\rho(\mathbf{z}) = \langle\delta[\mathbf{z} - \xi(\mathbf{x})]\rangle, \tag{14}$$

where $\langle\cdot\rangle$ denotes an ensemble average.

▶ Up to an unimportant constant, the *free energy* surface (FES) is given by:

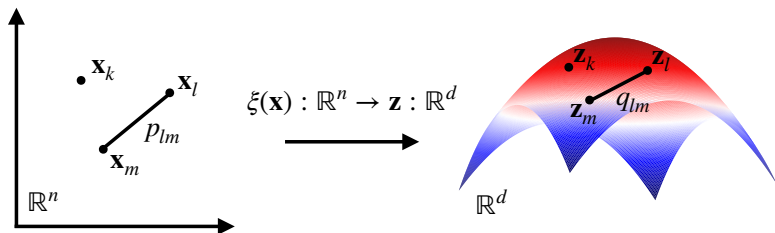$$F(\mathbf{z}) = -\frac{1}{\beta}\log\rho(\mathbf{z}). \tag{15}$$

Free Energy

Phase Space

▶ Microscopic coordinates: $\mathbf{x}$

▶ Equilibrium probability: $\rho(\mathbf{x})$

▶ Potential energy: $U(\mathbf{x})$

Reduced Space

▶ Collective variables: $\mathbf{z}$

▶ Equilibrium probability: $\rho(\mathbf{z})$
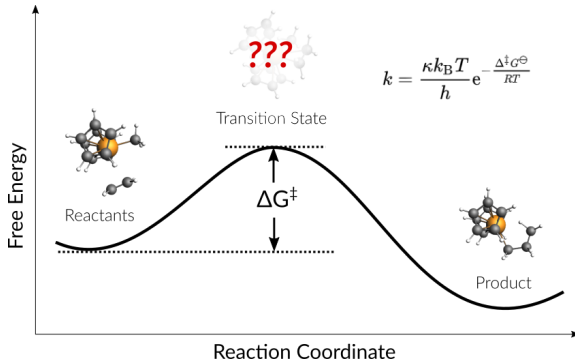
▶ Free energy: $F(\mathbf{z})$

$\xi(\mathbf{x}) : \mathbb{R}^n \to \mathbf{z} : \mathbb{R}^d$

$\mathbb{R}^n$

$\mathbf{x}_k$

$\mathbf{x}_l$

p_{lm}

$\mathbf{x}_m$

$\mathbb{R}^d$

$\mathbf{z}_k$

$\mathbf{z}_l$

$\mathbf{z}_m$

q_{lm}

Kinetics

▶ Knowing the free energy barrier, we can calculate rate using the Eyring equation:

$$k \propto \frac{k_{\mathrm{B}}T}{h}\mathrm{e}^{-\Delta G/RT},\qquad(16)$$

where ΔG is the energy barrier, R is the gas constant, and h is Planck's constant.

▶ The timescale of the reaction is given as $t = 1/k$.

???

Transition State

$$k = \frac{\kappa k_{\mathrm{B}}T}{h}\mathrm{e}^{-\frac{\Delta^{\ddagger}G^{\ominus}}{RT}}$$

Free Energy

Reactants

$\Delta G^{\ddagger}$

Product

Reaction Coordinate

Alanine Dipeptide

▶ Benchmark system for MD methods.
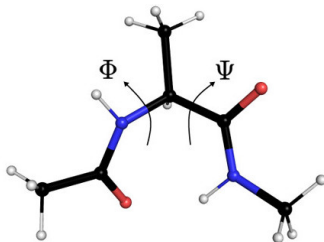
Figure 3: Alanine dipeptide with the Ψ and Φ dihedral angles.
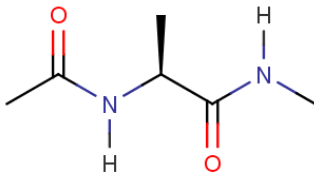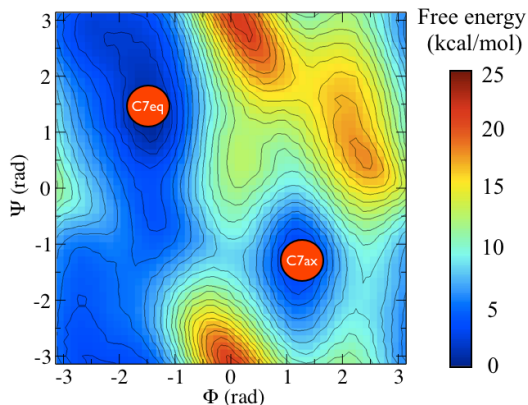
Alanine Dipeptide

▶ An example of a structure file (i.e., `.gro`) with alanine dipeptide.

```
   22
1ACE    HH31    1   1.075   1.111   1.585
1ACE     CH3    2   1.129   1.065   1.502
1ACE    HH32    3   1.070   0.991   1.448
1ACE    HH33    4   1.215   1.011   1.540
1ACE       C    5   1.162   1.167   1.395
1ACE       O    6   1.140   1.286   1.416
2ALA       N    7   1.230   1.118   1.289
2ALA       H    8   1.248   1.019   1.293
2ALA      CA    9   1.260   1.194   1.165
2ALA      HA   10   1.304   1.125   1.093
2ALA      CB   11   1.126   1.224   1.093
2ALA     HB1   12   1.078   1.312   1.135
2ALA     HB2   13   1.147   1.247   0.989
2ALA     HB3   14   1.073   1.129   1.094
2ALA       C   15   1.357   1.317   1.171
2ALA       O   16   1.392   1.372   1.068
3NME       N   17   1.404   1.353   1.294
3NME       H   18   1.350   1.322   1.373
3NME     CH3   19   1.517   1.445   1.315
3NME    HH31   20   1.495   1.544   1.278
3NME    HH32   21   1.546   1.457   1.420
3NME    HH33   22   1.606   1.405   1.267
```
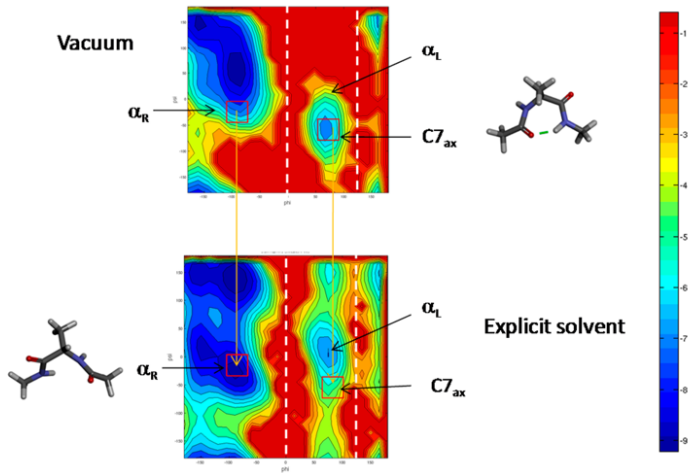
Alanine Dipeptide

- Alanine dipeptide in vacuum with backbone dihedral angles Φ and Ψ as CVs.

- Two metastable basins (or three), $C7_{\mathrm{eq}}$ and $C7_{\mathrm{ax}}$.

- Barrier aroung 34 kJ/mol (16 $k_{\mathrm{B}}T$).

- Mean transition time around 28 μs.

Alanine Dipeptide

▶ Comparison to the free energy of alanine dipeptide in explicit solvent.

http://dx.doi.org/10.1021/acs.jctc.5b00264

Alanine Dipeptide

▶ Difference between the ergodic and non-ergodic trajectories.

▶ In the left, the transitions between two energy basins are *rare* or *infrequent*.

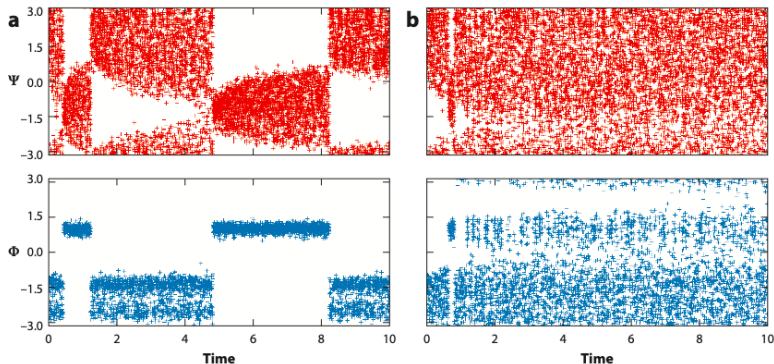▶ In the right, the trajectory jumps between the states many times.

Figure 4: Time evolution of alanine dipeptide in vacuum.

Alanine Dipeptide

▶ Demonstration: Calculating free energy along the CVs for alanine dipeptide.

▶ A repository with the lecture and exercise is available at
`https://github.com/jakryd/Spinaker-NCU`.

Acknowledgements

Let's meet at NCU!

International Intensive Summer and Winter Educational Meetings