Dive into computational physical chemistry

Lecture 5: Enhanced sampling with collective variables

Glen Hocky
October 18, 2022

Recap (from lecture 2)

Real experiments are at constant temperature or pressure. What we really want in MD is to “sample” configurations with the correct probabilities. At constant temperature, this is given by the Boltzmann equation:

$$P(\boldsymbol{X}) \propto e^{-\left(\frac{H(\boldsymbol{X})}{k_B T}\right)}$$

where $H(\boldsymbol{X}) = U(\boldsymbol{X}) + K.E.$

What we usually care about is

$$P(q_1, \dots, q_N) \propto e^{-\frac{U(q_1,\dots,q_N)}{k_B T}}$$

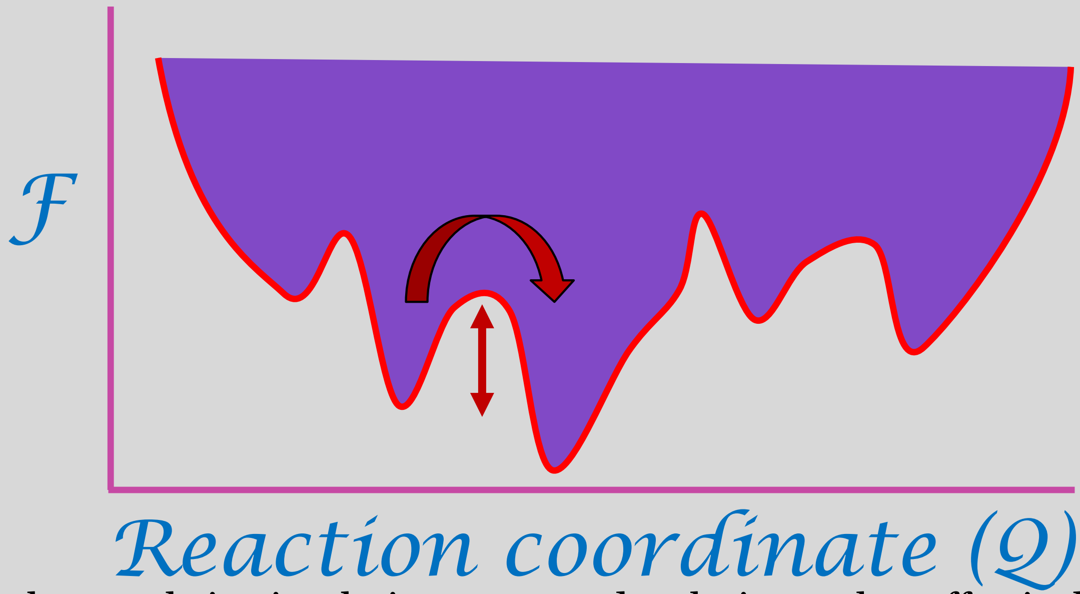
Goal: computing a free-energy landscape ("potential of mean force")

$$P(\boldsymbol{X}) = \frac{e^{-\left(\frac{H(\boldsymbol{X})}{k_B T}\right)}}{Z} \qquad P(Q_0) = \int dq_1 dq_2 \dots dq_N \delta(Q(\boldsymbol{X}) - Q_0) \frac{e^{-\left(\frac{H(\boldsymbol{X})}{k_B T}\right)}}{Z}$$

$$e^{-\frac{F(Q_0)}{k_B T}} \propto P(Q_0) \Rightarrow F(Q_0) = -k_B T \ln\big(P(Q_0)\big) + \mathrm{C}$$

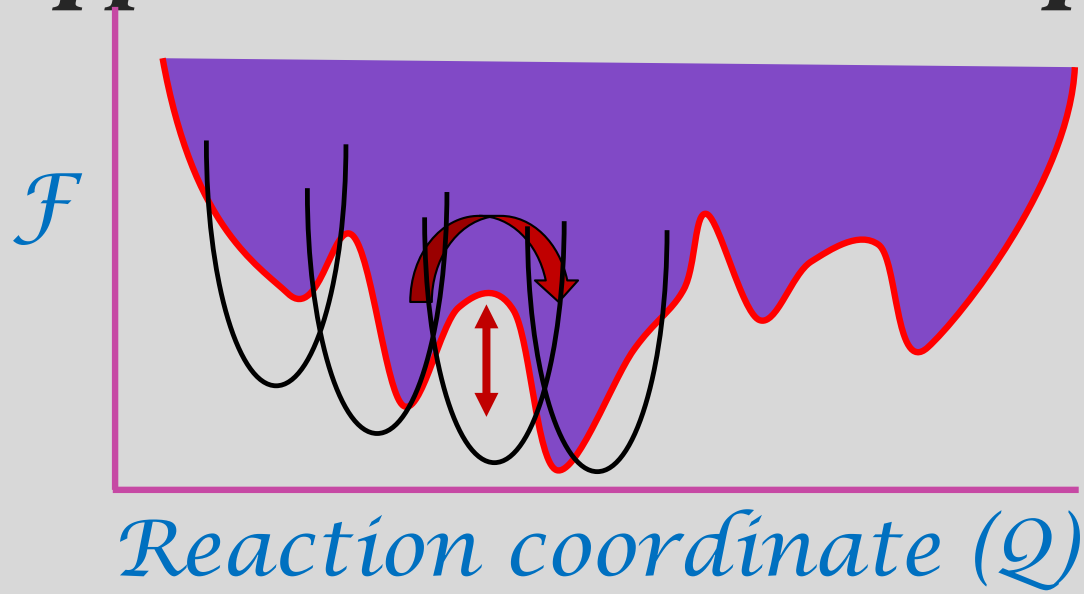
Challenge: the rare event problem

The chance of seeing some configuration depends on the (free) energy landscape. But the chance of transitioning between basins depends on barrier heights.



To accurately sample in simulations, we need techniques that effectively (a) lower temperature or (b) raise the temperature

One approach: Umbrella Sampling

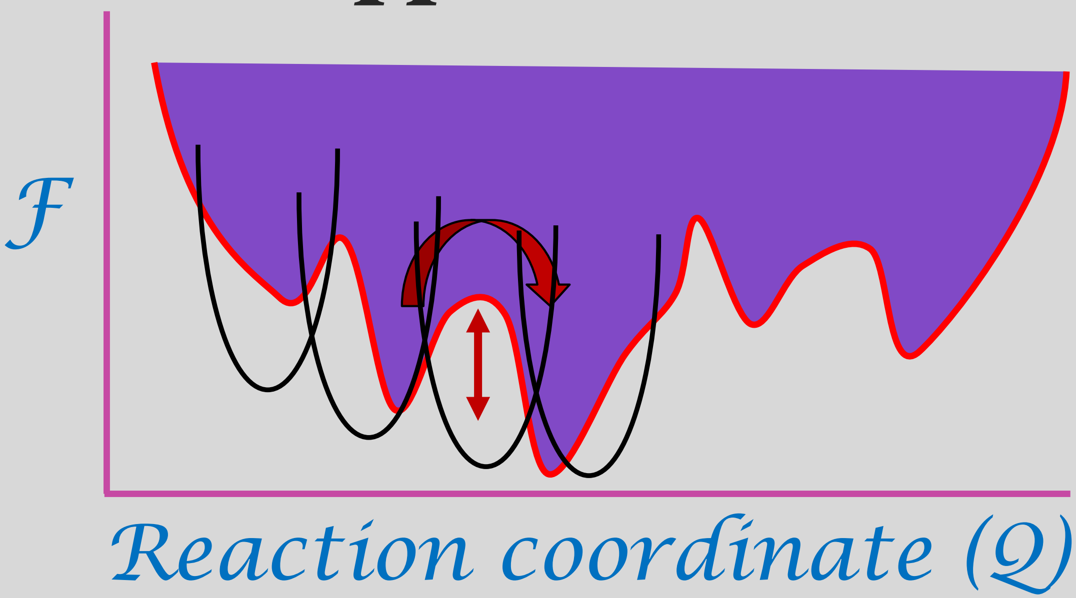


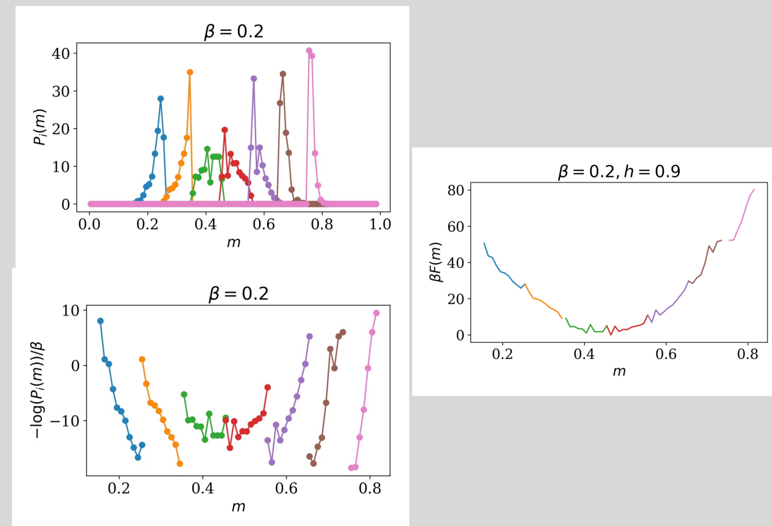
Torrie & Valleau, JCP 1977
Pangali, Rao, Berne, JCP 1979

$$P(Q_0) = \int dq_1 dq_2 \ldots dq_N \delta(Q(\boldsymbol{X}) - Q_0) \frac{e^{-\left(\frac{H(\boldsymbol{X})}{k_B T}\right)}}{Z}$$

$$\approx \int dq_1 dq_2 \ldots dq_N \, \mathrm{e}^{-a(Q(\boldsymbol{X})-Q_0)^2} \, e^{-\left(\frac{H(\boldsymbol{X})}{k_B T}\right)} = \int dq_1 dq_2 \ldots dq_N \, e^{-\left(\frac{H(\boldsymbol{X}) + \frac{1}{2}k(Q(\boldsymbol{X})-Q_0)^2}{k_B T}\right)}$$

One approach: Umbrella Sampling





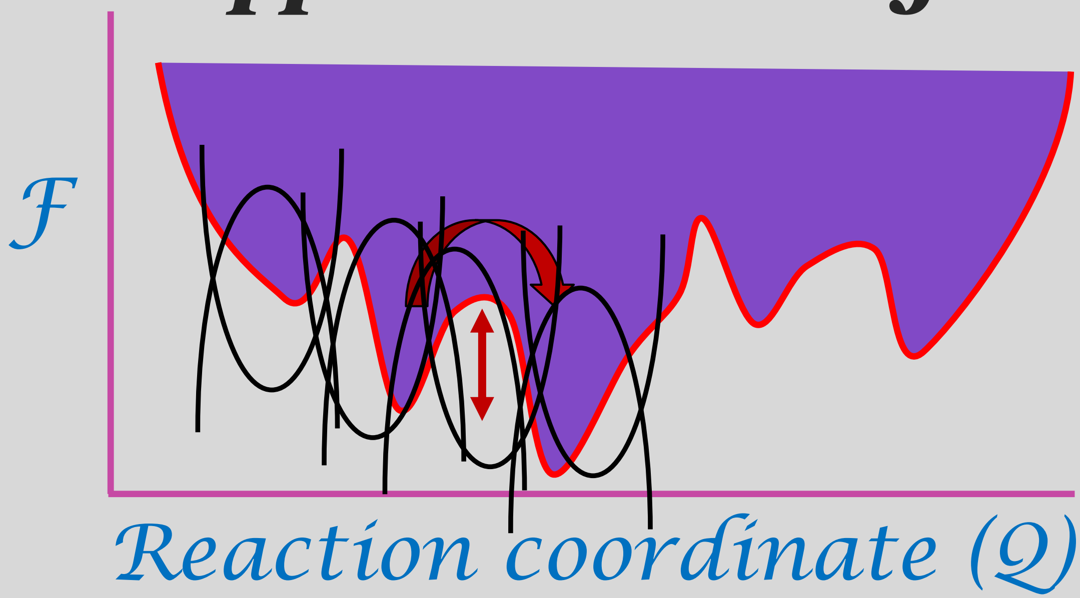
$$P(Q_0) \approx \int dq_1 dq_2 \dots dq_N \, e^{-\left(\frac{H(\boldsymbol{X})+\frac{1}{2}k(Q(\boldsymbol{X})-Q_0)^2}{k_B T}\right)}$$

$$F(Q_0) \approx -k_B T \ln\big(P(Q_0)\big) + C$$

Need approach to estimate C's so that *F(Q)* estimate is continuous.

E.g. WHAM – Kumar…Kollman, JCP 1992
MBAR – Shirts & Chodera, JCP 2008

Another approach: metadynamics



Add energy WHERE YOU ARE, in order to explore more areas

$$V(Q_j, t) = \sum_{i=1}^{t} h e^{-\frac{(Q(t_i) - Q_j)^2}{2\sigma^2}}$$

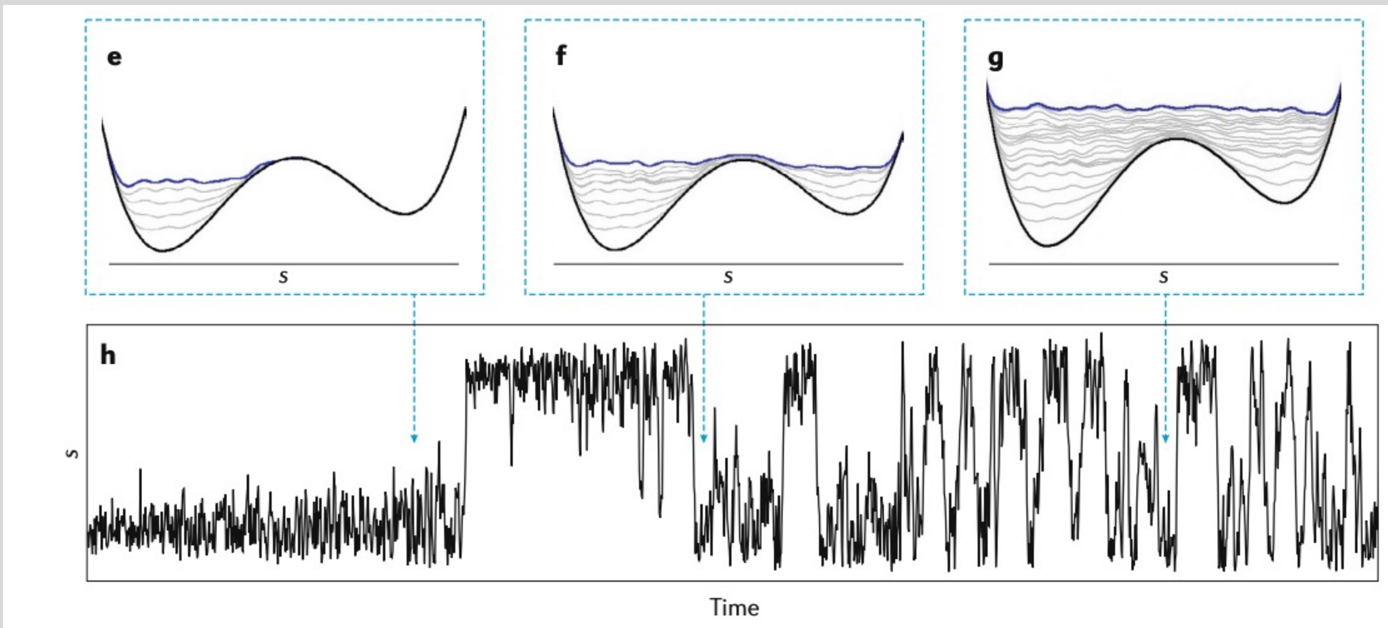
Laio & Parrinello, PNAS (2002)

Amount of energy added should converge to negative of FES.

$$V(Q_j, t) = \sum_{i=1}^{t} h e^{-\frac{V(Q_j, t_i)}{k_B \Delta T}} e^{-\frac{(Q(t_i) - Q_j)^2}{2\sigma^2}}$$

Barducci, Bussi, & Parrinello, PRL (2008)

Another approach: metadynamics



Well tempered MetaD: $V(Q,t) \rightarrow -\left(\frac{\Delta T}{T+\Delta T}\right) F(Q)$, [Proof: Dama, Voth, Parrinello, PRL (2014)]

PLUMED Library

- PLUMED is a software code that is a plug-in for many open source MD codes (GROMACS, AMBER, NAMD, LAMMPS, …); https://www.plumed.org/
- PLUMED has two kinds of objects
  - Collective variables – different quantities you might want to bias and
  - Biases – different schemes for pushing or pulling your system around, including metadynamics
- With GROMACS
  - gmx_mpi –s input.tpr –plumed plumed.dat

PLUMED Input file syntax

The syntax of the PLUMED input file

The main goal of PLUMED is to compute collective variables, which are complex descriptors than can be used to analyze a conformational change or a chemical reaction. This can be done either on-the-fly during molecular dynamics or a posteriori, using PLUMED as a post-processing tool. In both cases one, should create an input file with a specific PLUMED syntax. A sample input file is below:

Click on the labels of the actions for more information on what each action computes

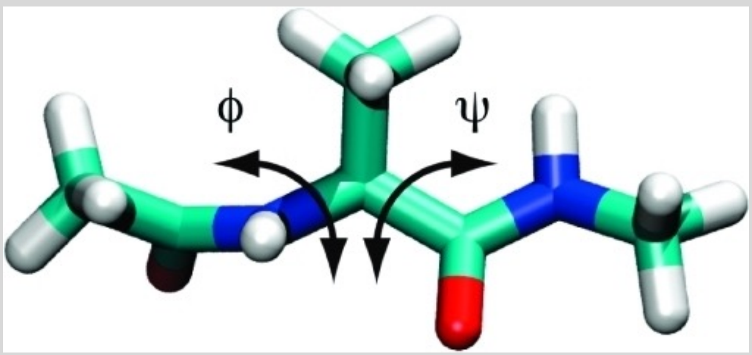
master passing

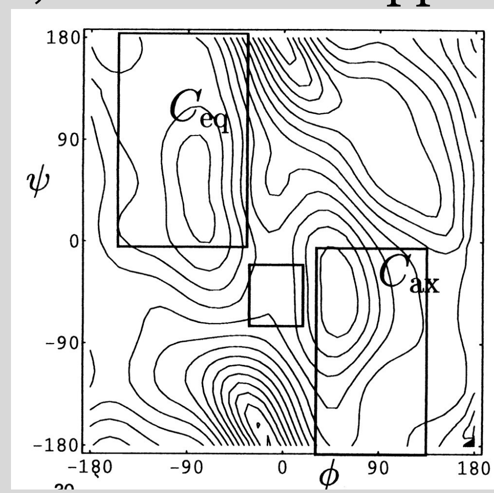
```
# Compute distance between atoms 1 and 10.
# Atoms are ordered as in the trajectory files and their numbering starts from 1.
# The distance is called "d" for future reference.
d: DISTANCE ATOMS=1,10
# Create a virtual atom in the center between atoms 20 and 30.
# The virtual atom only exists within PLUMED and is called "center" for future reference.
center: CENTER ATOMS=20,30
# Compute the torsional angle between atoms 1, 10, 20, and center.
# Notice that virtual atoms can be used as real atoms here.
# The angle is called "phi" for future reference.
phi1: TORSION ATOMS=1,10,20,center
# the same CV defined before can be split into multiple line
phi2: TORSION ...

   ATOMS=1,10,20,center
...
# Print d every 10 step on a file named "COLVAR1".
PRINT ARG=d STRIDE=10 FILE=COLVAR1
# Print phi1 and phi2 on another file names "COLVAR2" every 100 steps.
PRINT ARG=phi1,phi2 STRIDE=100 FILE=COLVAR2
```

https://www.plumed.org/doc-master/user-doc/html/lugano-1.html

Test system: alanine dipeptide

"Alanine dipeptide" is not actually a di-peptide, but rather a capped alanine peptide monomer.





We study its free energy landscape in terms of two main dihedral coordinates that determine the configuration of a peptide.

This time – the whole landscape will be explored

Bolhuis, Dellago, Chandler. PNAS 2000. https://doi.org/10.1073/pnas.100127697

Today

1. Metadynamics simulations for alanine dipeptide
2. Pull updates on comp-lab-class github page; see Week 6 assignment

https://github.com/hockyg/comp-lab-class/blob/main/Week6/Assignment.md