Umbrella Sampling Tutorial (NAMD)

Dhiman Ray
Department of Chemistry
University of California Irvine, USA

1 Introduction

This document outlines the steps for performing umbrella sampling simulation using the NAMD 2.14 software package using the colvars module. Umbrella sampling is a method to calculate free energy landscape of rare event processes using molecular dynamics simulation. This method was proposed by Torrie and Valleau [1] and later improved and used by many research groups [2].

2 Goal

The goal of this tutorial is to set up basic umbrella sampling simulation using NAMD software and calculation of one-dimensional Potential Mean Force (PMF) using Weighted Histogram Analysis Method (WHAM) method using the WHAM code [3]. The necessary files are provided in `https://github.com/dhimanray/Umbrella_Sampling_Tutorial`.

3 Protocol

3.1 Structure preparation and equilibration

The structure of the system has to be built, minimized and equilibrated adequately before starting the rest of the tutorial. For this part, please refer to the NAMD tutorial: `http://www.ks.uiuc.edu/Training/Tutorials/namd-index.html`

3.2 Metadynamics for pathway generation

To perform umbrella sampling we need a set of structure along the pathway of the reaction (protein conformational transition, ligand dissociation etc.). There are different ways to obtain these structures e.g. using steered molecular dynamics, targeted molecular dynamics, metadynamics etc. We will perform metadynamics for this tutorial.

The following files are provided in the `metadynamics` directory:

- `metadynamics.conf` : The NAMD configuration file for metadynamics
- `colvars.in` : The colvars input file for metadynamics
- `job.sh` : The job submission file using PBS system. But this file is specific to our computing cluster and should be modified before using in your computing facility.

More details are provided within the files.
Additionally, the following files will be required:

- `ionized.psf` (CHARMM) or `ionized.prmtop` (AMBER) : The topology file of the solvated and ionized system
- `ionized.pdb` : The pdb file of the solvated and ionized system
- `toppar/` : The directory containing all the CHARMM parameter files (only if you are using CHARMM force field)
- `refernces2.pdb` : (Optional) Reference coordinates in case a harmonic restraint is applied to the system.
- `system_equilibration.xsc`, `system_equilibration.coor`, and `system_equilibration.vel` : NAMD restart files generated from the equilibration simulation.

3.3 Building input files for Umbrella Sampling

- After the metadynamics simulation is done, copy the `MetaD.dcd` file and the `ionized.pdb` file to the `metadynamics_analysis` directory.
- First we need to wrap the trajectory so that the protein (or solute) does not break, or the water remains confined within the box. (We need to do this now because we used `wrapall off` flag while running the metadynamics simulation. So the trajectory is not automatically wrapped during the simulation. This is recommended for `colvars` module because otherwise the collective variable calculation can be incorrect.)

  For trajectory wrapping, load the trajectory into VMD and use the following command in tkConsole window: `pbc wrap -centersel "protein" -center com -compound residue -all` . Then save the dcd file as `MetaD_wrapped.dcd`.
- Execute `distance.py`. It will generate the `distance.dat` file which contains the distance (or other collective variable)
- Execute `umbrella_centers.py`. It will read the `distance.dat` file and generate the `umbrella_centers.dat` file which will show which frame of the metadynamics trajectory is nearest to the location of each umbrella sampling window. (Before doing this, you need to modify the `umbrella_centers.py` to specify the location of the centers of the umbrella windows.

- Execute `umbrella_centers_pbd.py`. This will generate the pdb starting structure for each umbrella window and store them in the `umbrella_centers` directory, outside the metadynamics directory.

3.4 Running Umbrella Sampling simulation

- Go to the `US/US_1` directory. This directory is for the first umbrella sampling window. The following files are provided.

  `umbrella.conf` : NAMD configuration file for umbrella sampling calculation.

  `colvars.in` : Colvars input file for umbrella sampling. (Note that the spring constant of the harmonic bias is gradually changed from zero to `targetForceConstant` to avoid sudden application of high force in the system. This portion of the trajectory must be discarded before performing WHAM analysis.)

  `job.sh` : Job submission file

  `MetaD.restart.xsc` : The restart file (cell dimension only) from the end point of metadynamics simulation.

  We did not use the `.coor` or `.vel` files because each umbrella window will have different starting structures and velocities. Inside the `.conf` file, we use `structure.pdb` as starting coordinate and initiate new velocities by `reinitvels` command.

  We also need the following files:

  `ionized.psf` (CHARMM) or `ionized.prmtop` (AMBER) : The topology file of the solvated and ionized system

  `structure.pdb` : The pdb file `us_1.pdb` from `umbrella_centers` directory copied here and renamed as `structure.pdb`.

  `toppar/` : The directory containing all the CHARMM parameter files (only if you are using CHARMM force field)

  `refernces2.pdb` : (Optional) Reference coordinates in case a harmonic restraint is applied to the system.

- Go inside the `US` directory and execute `./build.sh`. It will create all the umbrella directories. This code will copy the appropriate starting structure from the `umbrella_centers` directory, and change the `colvars.in` and the `job.sh` file to use the correct parameter for the biasing force and directory path respectively.

- Execute `./runall.sh` to submit all umbrella sampling jobs to the cluster.

3.5 Calculating Free Energy Landscape using WHAM

- After the simulations are completed, execute `copy_data.sh` to copy all the `US.colvars.traj` file from the appropriate umbrella sampling directories

to the `data` directory. (A sample data directory is provided in zip format for you to analyze.)

- Go to `pmf_15ns` directory. Here we will compute the free energy landscape using the last 15 ns part of each 20 ns trajectory.
- Execute `convert.py`. This will create the time series of the umbrella sampling collective variables for each of the windows in the `US_data` directory. (Before running this code, change the initial and final positions of the umbrella windows as well as the spacing, according to your system of study.) Also check for instructions inside the script.
- Execute `metadatafile.py` to create `metadatfile.in` file, which is an input to the WHAM code. Check for instructions inside the code.
- Calculate PMF or Free Energy landscape by executing `./calculate_pmf.sh`. The result will be printed in `pmf.dat`. This can be plotted using `plot-pmf.py`.

References

[1] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 2 1977.

[2] Benoît Roux. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, 91(1-3):275–282, 9 1995.

[3] Alan Grossfield. WHAM: the weighted histogram analysis method. Version 2.0.10.