CHEM 7590 Computational exercise:
Potential of mean force of SPC/E water
Due Monday, December 11

The free energy surface along the chosen coordinate is referred to as the potential of mean force (PMF). Often PMF simulations are used in conjunction with equilibrium pulling methods (umbrella sampling) or nonequilibrium pulling methods (Jarzynski equility). Both methods allow enhanced sampling when high energy barriers exist. The PMF can also be constructed from a radial distribution function $g^{(2)}(r)$ along a reaction coordinate. However, $g^{(2)}(r)$ might fail to provide accurate PMF if high energy barrier exists. In this exercise, you will analyze PMF generated from two pulling methods and $g^{(2)}(r)$. Specifically, the free energy surface between oxygen-oxygen for water.

1 Starting files

This exercise uses GROMACS package on Owens (`owens.osc.edu`). You may copy the necessary files from `/users/PAS1314/osu9080/Chem7590/waterPulling`. You will be using the following batch scripts to run all the jobs:

```
Batch_NTP.sh
Batch_GenConfig.sh
Batch_Umbrella.sh
Batch_Jarzynski.sh
```

You must replace the path of the input folder (line 24) before using the scripts.

2 Radial distribution function of SPC/E water

1) [10 points] Show that $-\frac{1}{\beta}\ln g^{(2)}(\mathbf{r}, \mathbf{r}') = A(\mathbf{r}, \mathbf{r}')$+constant, where $g^{(2)}(\mathbf{r}, \mathbf{r}')$ is a radial distribution function that any atom at r and any other atoms at r', and $A(\mathbf{r}, \mathbf{r}')$ is the corresponding Helmholtz free energy. You might start with the definition of two body density

$$\rho^{(2)}(\mathbf{r}, \mathbf{r}') = \rho^2 g^{(2)}(\mathbf{r}, \mathbf{r}') = \sum_{j=1}^{N} \sum_{k=1(k \neq j)}^{N} P_{jk}^{(2)}(\mathbf{r}, \mathbf{r}')$$

where

$$P_{jk}^{(2)}(\mathbf{r}, \mathbf{r}') = \frac{1}{Q} \frac{1}{N! h^{3N}} \int \mathbf{p}^{\mathbf{N}} \int \mathbf{r}^{\mathbf{N}} \delta(\mathbf{r_j} - \mathbf{r}) \delta(\mathbf{r_k} - \mathbf{r}') exp\left[-\beta E\left(\mathbf{r^N}, \mathbf{p^N}\right)\right]$$

In a homogeneous system $\rho g^{(2)}(\mathbf{r}, \mathbf{r}')$ is the probability that any particle is distance $|\mathbf{r} - \mathbf{r}'|$ on a 1D

vector from the test particle at $\mathbf{r}$. In 3D space, the probability that any particle is distance $|\mathbf{r} - \mathbf{r}'|$ from the test particle at $\mathbf{r}$ should be $4\pi|\mathbf{r} - \mathbf{r}'|^2\rho g^{(2)}(\mathbf{r}, \mathbf{r}')$.

Answer:

$$P_{jk}^{(2)}(\mathbf{r}, \mathbf{r}') = \frac{1}{Q}\frac{1}{N!h^{3N}} \int \mathbf{p^N} \int \mathbf{r^N} \delta(\mathbf{r_j} - \mathbf{r})\delta(\mathbf{r_k} - \mathbf{r}')exp\left[-\beta E\left(\mathbf{r^N}, \mathbf{p^N}\right)\right]$$
$$= \frac{Q_{jk}(\mathbf{r}, \mathbf{r}')}{Q}$$

where $Q_{jk}(\mathbf{r}, \mathbf{r}')$ is the partition function when particle j at $\mathbf{r}$ and particle k at $\mathbf{r}'$.

$$-\frac{1}{\beta}\ln g^{(2)}(\mathbf{r}, \mathbf{r}') = -\frac{1}{\beta}\ln\left[\frac{V^2}{N^2}\sum_{j=1}^{N}\sum_{k=1(k\neq j)}^{N} P_{jk}^{(2)}(\mathbf{r}, \mathbf{r}')\right]$$
$$= -\frac{1}{\beta}\ln Q(\mathbf{r}, \mathbf{r}') + \text{const}$$
$$= A(\mathbf{r}, \mathbf{r}') + \text{const}$$

where $Q(\mathbf{r}, \mathbf{r}')$ is the partition function when any particle at $\mathbf{r}$ and any other particles at $\mathbf{r}'$.

You are given a starting file, `water.gro`, which contains 510 SPC/E water molecules, pre-equilibrated at 298K, 1bar. Now, we will run a 200ps simulation using the `water_NTP.mdp` file. We are going to conduct this simulation in NVT condition, accounting for the Helmholtz free energy profile as a function of the distance between two oxygen atoms. You should use the provided Batch job file by typing

`qsub Batch_NTP.sh`

To obtain the radial distribution function,

```
cd Output_NTP/
module load gromacs
gmx rdf -f NTP.trr -s NTP.tpr -n ../water.ndx -o rdf-OW-OW.xvg
```

When prompted, choose `OW` twice, once for the reference group and again for the second group, thereby obtaining $g_{OO}^{(2)}(r)$, then press "Ctrl+d" to exit `gmx rdf`.

2) [15 points] Using the obtained $g_{OO}^{(2)}(r)$, make a plot of the potential of mean force (in kJ mol^{-1}) as a function of the distance between two oxygen atoms (in nm). Show the plot range between 0 and 1 nm.

Answer:
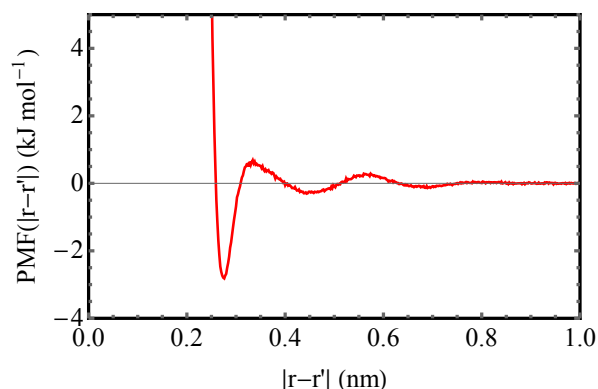
3 Umbrella sampling method

3.1 Generate Configurations

To conduct umbrella sampling, we will have to generate several configurations along a reaction coordinate, which in this practice is the distance between the two oxygen atoms, $|\mathbf{r} - \mathbf{r'}|$. We will first pull a reference oxygen atom from it's nearest neighboring oxygen atom. The starting configurations for each umbrella sampling window is snapshotted at desired $|\mathbf{r} - \mathbf{r'}|$. Let's take a look of the provided `waterGenConfig.mdp` for this pulling. A brief explanation of the pulling options is as follows

```
pull = yes ;activates the pull code
pull-coord1-type = constraint ;pulling with constant velocity
pull-coord1-geometry = distance ;simple distance increase
pull-coord1-dim = Y Y Y ;we are pulling in 3 dimensions.
pull-group1-name = target ;the name in the index file of the first group
pull-group2-name = ref ;the name in the index file of the second group
pull-ngroups = 2 ;there are two groups subjected to a pulling force
pull-ncoords = 1 ;there is only one reaction coordinate
pull-coord1-groups = 1 2 ;groups 1 and 2 define the reaction coordinate
pull-coord1-rate = 0.005 ;the rate (nm/ps) of change of the reference position.
pull-coord1-start = yes ;the initial distance is the reference distance for
the first frame.
```

For more detailed descriptions, refer to `http://manual.gromacs.org/online/mdp_opt.html`. At the end of the `water.ndx`, you can find the index of the selected oxgen atoms (target and ref). The index for each atom is printed in the 3rd column of the `water.gro` file. You may calculate the distance between the two selected atoms, which should be ~ $0.25nm$. Alternatively, use a c++ program, `Nearest.exe`, which will find the nearest oxygen atom around the target atom and calculate their distance. For instance, you can run

3

```
./Nearest.exe water.gro nearestAtom.dat 1108
```

The output `nearestAtom.dat` provides indices of the target oxygen atom (1108), the nearest oxygen atom (274), and the distance (in nm). In the next method, we will use this program to generate several oxygen pairs for the nonequilibrium pullings. To conduct umbrella sampling, we use only one oxygen pair (target=1108 and ref=274) to conduct umbrella sampling. Send the pulling job using the provided batch script

```
qsub Batch_GenConfig.sh
```

Make sure you have changed the path of input folder. When finished, in your main working folder, run

```
sh GetConfig.sh
```

This script will simply generate 6 snapshots of configurations from the trr file. In the `Output_GenConfig/`, you should see six `config_*ps.gro` files tagged with the corresponding snapshot time points.

3.2   Umbrella Sampling and Data Analysis

After we obtained the six configurations, we will be sampling oxygen-oxygen distance from 0.0 - 1.0 nm in three dimensions. We will need to run six simulations using the generated configurations, and the umbrella potential for each simulation will constraint the reaction coordinate around the initial oxygen-oxygen distance in each configuration. It is the most convenient to run all the simulations in one script,

```
qsub Batch_Umbrella.sh
```

You should have all the six job finished in about 20 min. In the `Output_Umbrella` folder, you will see all the generated files from the six jobs, hopefully, with clear enough file names. Were going to use `gmx wham` to get the PMF. The program takes a file containing a list of the `.tpr` files and another file containing a list of the `.xvg` files containing the force as arguments. To create these two files, do:

```
ls *.tpr > tpr.dat
ls *pullf.xvg > pullf.dat
```

Then you can run gmx wham:

```
gmx wham -it tpr.dat -if pullf.dat -o -hist -unit kJ
```

You should end up with a `profile.xvg` file that contains PMF as a function of $|\mathbf{r} - \mathbf{r}'|$. The other output from the `gmx wham` is `histo.xvg`, which contains the histograms of the configurations within the umbrella sampling windows. Open the `histo.xvg`, you should see seven columns. The 1st column is a list of $|\mathbf{r} - \mathbf{r}'|$, and the 2nd-7th columns are the histograms for the six umbrella

4

samplings (t = 0, 20, 40, 80, 120, and 140 ps).

3) [40 points] Make a plot of PMF($|\mathbf{r} - \mathbf{r}'|$) using the data in `profile.xvg`. Does this PMF profile match the one obtained from the $g_{OO}^{(2)}(r)$? Make another plot of PMF^{corr}($|\mathbf{r} - \mathbf{r}'|$) = PMF($|\mathbf{r} - \mathbf{r}'|$) + $2k_BT\ln|\mathbf{r} - \mathbf{r}'|$ + const, where you will have to make your own choice of the "const" so that PMF^{corr}($|\mathbf{r} - \mathbf{r}'|$) ≈ 0 at 1nm. How does the PMF^{corr}($|\mathbf{r} - \mathbf{r}'|$) agree with the PMF from $g_{OO}^{(2)}(r)$?
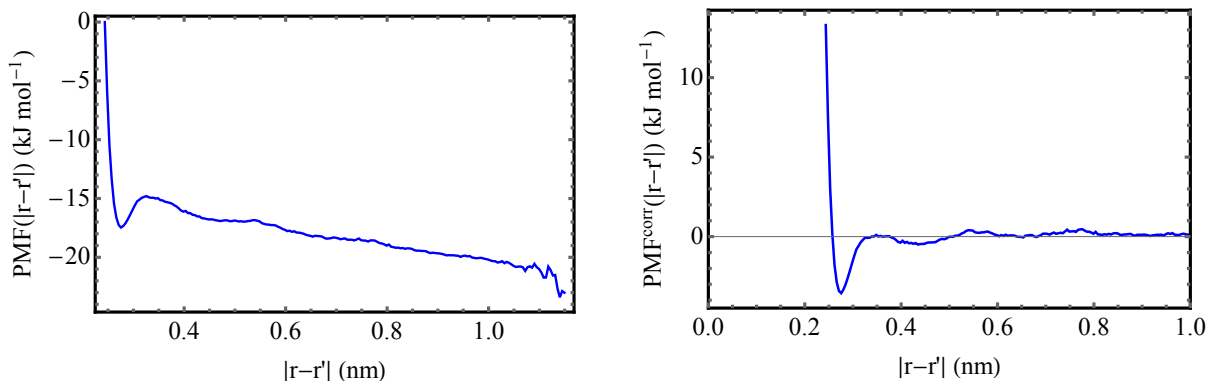
Answer:

We would expect the interaction to go to zero at longer distances. Because we used a 3-dimensional biasing potential, we need to include a correction to make it comparable with the PMF converted from $g_{OO}^{(2)}(r)$. Imagine one of the oxygen as the reference point. The other oxygen is allowed to sample all around that point at distance r, covering the surface of some sphere with radius $|\mathbf{r} - \mathbf{r}'|$. That it, the umbrella sampling is associated with a partition function Q($|\mathbf{r} - \mathbf{r}'|$). In a homogeneous system,

$$Q(|\mathbf{r} - \mathbf{r}'|) = 4\pi|\mathbf{r} - \mathbf{r}'|^2 V Q(|\mathbf{r}, \mathbf{r}'|)$$

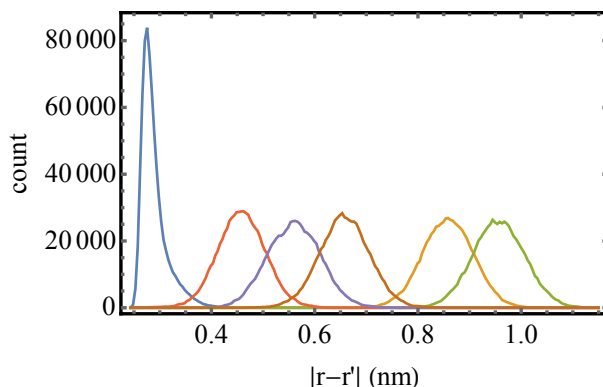where V is the volume of the system. Hence, the PMF($-\mathbf{r} - \mathbf{r}'-$) from the umbrella sampling is

$$A(|\mathbf{r} - \mathbf{r}'|) = -\frac{1}{\beta}\ln Q(|\mathbf{r} - \mathbf{r}'|)$$

$$= -\frac{1}{\beta}\ln Q(|\mathbf{r}, \mathbf{r}'|) - 2k_BT\ln|\mathbf{r} - \mathbf{r}'| + |\text{const}|$$

$$A(|\mathbf{r}, \mathbf{r}'|) = A(|\mathbf{r} - \mathbf{r}'|) + 2k_BT\ln|\mathbf{r} - \mathbf{r}'| + |\text{const}|$$

Therefore, to match the PMF from $g_{OO}^{(2)}(r)$, a correction of $2k_BT\ln|\mathbf{r}, \mathbf{r}'|$ and a constant offset are required.

4) [20 points] To confirm if the overlap between the windows is sufficient, overlay all the six histograms on a single plot.

Answer:

5

4 Nonequilibrium Pulling Method

Another method to obtain PMF is to perform nonequilibrium pullings and use the Jarzynski equality to relate the free energy difference to the irreversible work:

$$\Delta A_{AB} = -\frac{1}{\beta} ln \langle \exp\left[-\beta W_{AB}\right] \rangle_A$$

where W_{AB} is the work performed to force the system along one path from state A to B, the angular bracket denotes averaging over a canonical ensemble of the initial state A. We will perform nonequilibrium pulling from five different pairs of oxygen atoms, where each oxygen pair is selected using the mentioned `Nearest.exe` program. The `waterJarzynski.mdp` set the parameters for the nonequilibrium pulling, which are pretty similar to the `waterGenConfig.mdp` except for that the `pull_coord1_rate` is changed to 0.0005 nm/ps. The indices of the five target oxygen atoms are 1108, 1198, 1036, 1516, and 1243, and the corresponding nearest oxygen atoms will be determined by `Nearest.exe` and output to the `water_*.ndx` file. The "*" is the index of the target oxygen, e.g. `water_1108.ndx`. Next, we will be using the `water_*.ndx` to run a pulling simulation. All the steps are written in the batch script, `Batch_Jarzynski.sh` line 71-80. To run all the five pulling simulations, do

```
qsub Batch_Jarzynski.sh
```

It might take upto 30 min to complete all the jobs. Once completed, cd into `Output_Jarzynski/`, and you will see all the outputs from the five simulations, whose file names are labeled by the corresponding indices of target oxygen atoms. For instance, the pulling between target=1108 and ref=274 has the following output files:

```
OW_1108.*, OW_1108_pullf.xvg, and OW_1108_pullx.xvg
* = trr, tpr, etc.
```

The `OW_1108_pullx.xvg` records $|\mathbf{r} - \mathbf{r}'|$ (nm) vs time(ps), and `OW_1108_pullf.xvg` records the pulling force (kJ/mol/nm) along the $\mathbf{r} - \mathbf{r}'$ vector vs time (ps). To obtain PMF, we will first have to obtain the irreversible work for each pulling. The work for pulling can be calculated using the

6

force and distance in the `*pullf.xvg` and `*pullx.xvg` respectively. Use the finite difference approximation, you can estimate the work for pulling two oxygen from distance 0 to r_N:

$$\text{work}(r_N) = \int_0^{r_N} f(r')dr' \approx f(r_1)(r_1 - 0) + f(r_2)(r_2 - r_1) + f(r_3)(r_3 - r_2) + ... + f(r_N)(r_N - r_{N-1})$$

where $f(r_i)$ and r_i is the force and distance at i_{th} time step.

5) [30 points] Make a single plot of work(r) (kJ/mol) vs r (nm) of the five pulling simulations. p.s. There is ~ 18000 for each `.xvg`, so make sure your plotting software is able to buffer all the data (works in Mathematica).

Answer:
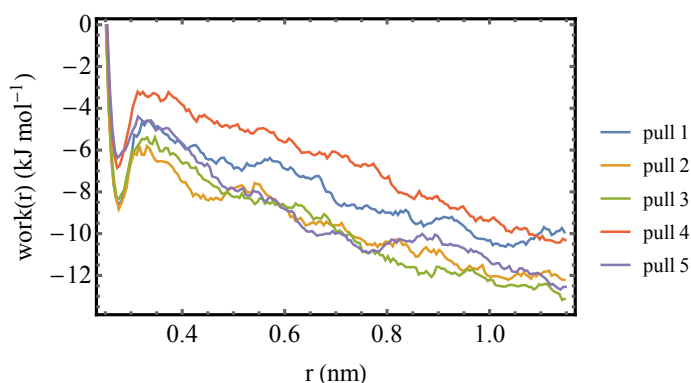
6) [40 points] Use the work of the five pulling simulations and the Jarzynski equality, make a plot of PMF($|\mathbf{r} - \mathbf{r}'|$) and also a plot of $\text{PMF}^{corr}(|\mathbf{r} - \mathbf{r}'|) = \text{PMF}(|\mathbf{r} - \mathbf{r}'|) + 2k_B T \ln|\mathbf{r} - \mathbf{r}'| + \text{const}$, where you will have to make your own choice of the "const" so that $\text{PMF}^{corr}(|\mathbf{r} - \mathbf{r}'|) \approx 0$ at 1nm. How does the $\text{PMF}^{corr}(|\mathbf{r} - \mathbf{r}'|)$ agree with the PMF from $g_{OO}^{(2)}(r)$?

Answer:
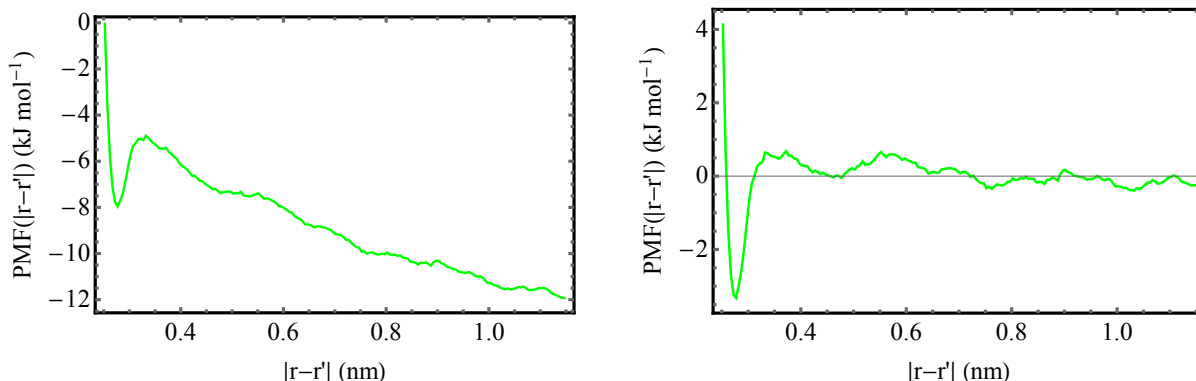
7) [5 points] Overlap the PMF from $g_{OO}^{(2)}(r)$ together with the PMF^{corr} from the umbrella sampling method and nonequilibrium pulling method.