Structure

INTRODUCTION

Structure determines the equilibrium structure of liquid water, as a function of distance from a specified location in your MD simulation. Specifically, this program calculates the locations of hydration layers with respect to the user's specified center-of-mass. The algorithm for the calculation is the radial distribution function g(r), given in the equation below. Also calculated when running Structure, is the calculation of the integral of g(r). The integral of g(r) quantifies the average number of waters present at a present radius of the radial distribution function. You will need to perform this calculation before you can calculate the orienational correlation function, and interaction energies for nanocavity water in your Gromacs simulation.

ALGORITIHM

$$\mathrm{g(r)} = \frac{1}{4\pi r^2 \rho_{bulk}} \sum_{i=0}^{nT} \sum_{j=1}^{Nwat} \langle \rho(r_{ij} - c_i) \rangle$$

g(r) is determined by calculating the average density of water molecules with respect to the user specified center-of-mass(COM) at each timestep i. The average density $\langle \rho(\mathrm{r}_{ij}\text{-}\mathrm{c}_i) \rangle$, is obtained by collecting the distances of each water oxygen j at timestep i, with respect to the COM at timestep i, into a histogram and then dividing by the total number of timesteps. nT is the number of simulation timesteps, Nwat is the number water molecules in your simulation. The average density is normailized by $4\pi \mathrm{r}^2 \rho_{bulk}$. ρ_{bulk} is the bulk water density, equal to approximately 33.4 waters/nm^3. r is the distance from the center-of-mass.

REQUIREMENTS

- Structure requires the user to define a set of atoms to calculate the center-of-mass. To generate an index of these atoms, use the Gromacs command make_ndx.
- Structure requires the user to generate a distance file using the Gromacs command g_dist. The distance file will later be used in to acquire orientational correlation functions and interaction energies inside of.
- Structure requires a distance file to calculate g(r), and the integral of g(r). The distance file is an index of the distances of each simulation water's atom with respect to the specified center-of-mass, at each timestep. To generate the distance file, use the Gromacs command g_dist as follows on your simulation trajectory(xtc) and topology file(tpr):

```
g_dist -f *.xtc -s *.tpr -n *.ndx -dist MaxBoxLength >distance.txt
```

g_dist will prompt you to select two groups, the first group you select will define the centers-of-mass, the second is the solvent. Note, it is best to select a solvent atom that would best approximate the center-of-mass of your solvent, this will reduce the size of the

output file. For example, if your solvent of interest is water, oxygens best approximate the center-of-mass, select oxygen for the second group for g_dist.

The output from g_dist contains a lot of unnecessary data. To prune this Gromacs output for the data we are interested in, perform the following command.

awk 'print $2,$3,$7' GromacsOutputFile >distance.txt

Next, use Structure to caluclate g(r) from your **awk** output as follows:

./rdf InputSolData(.txt) Bin_Width(Angstroms) MaxRadius(Angstroms) Nsol Simlength(ps) Timestep(ps) rho(molecules/nm^3) RDF_Out Num_Out

USING

Perform the following commands on your terminal of choice, or in a shell script.

./rdf InputSolData(.txt) Bin_Width(Angstroms) MaxRadius(Angstroms) Nsol Simlength(ps) Timestep(ps) rho(molecules/nm^3 RDF_Out Num_Out

Inputs: (1)The input is the distance file. (2)Desired binwidth of the histogram. (3)MaxRadius is the maximum length of the simulation box. (4) Number of solvent. (5)Length of simulation in ps. (6)How often you save positions in your MD simulation.(7)Bulk liquid density of solvent. (8)Ouput file name for g(r). (9) Output file for integral of g(r). You can give the output any file extension you like.

DATA INTEPRETATION EXAMPLE

Figure 1a-b shows g(r) calculated by Structure tested on Gromacs simulations of β-cyclodextrin, and the β1-adrenergic receptor. For these molecules the center-of-mass is in the interior of the macromolecule. For solute simulations in bulk water g(r) should converge to 1 at distances far from the COM. For simulations of heterogenous mixtures, i.e. solvated lipid bilayer simulations with embedded membrane protein, g(r) should converge to 0.5. We truncated our plot to just show information regarding the hydration layer inside of the transmembrane domain of the β1-adrenergic receptor.

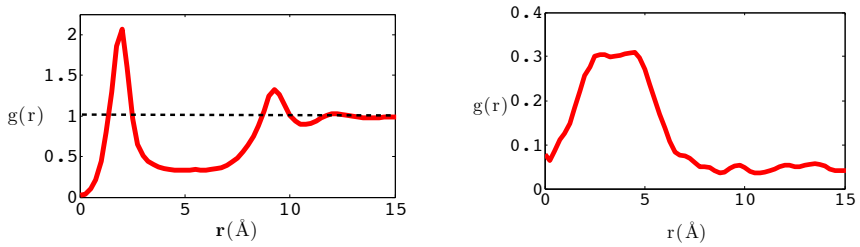

Figure 1. g(r) from simulations of β-cyclodextrin and β1-adrenergic receptor.

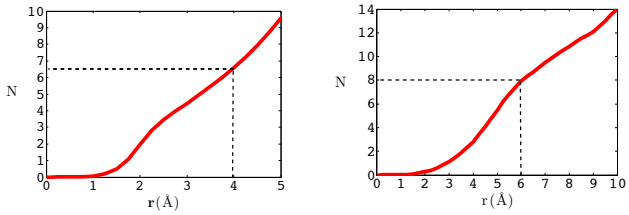

Figure 2. Integrals of g(r) from simulations of β-cyclodextrin and β1-adrenergic receptor.

We define the cavity of the internal pore as a sphere of radius r, measured from the start of the tail from the first peak in figure 1. this peak is a we know that the boundary of the internal cavity lies at $4\mathring{A}$ from the cyclodextrin COM. The first peak in both plots indicate the presence of an internal hydration layer. The second peak in figure 1a corresponds to an external hydration layer around cyclodextrin.

For the β-adrenergic receptor, the center-of-mass we were interested was defined from the atoms of the conserved aspartate residues Asp2.50 and Asp3.32 inside of transmembrane domain. For this macromolecule, we define max raidus of the pore at a distance of approximately $8\mathring{A}$ from the COM of, measured from the start of the tail from the first peak broad peak. The plateau like shape of the distribution, is a signature of the formation of a quasi-ordered water "chain" inside the pore, during the MD simulation.

The dashes in figure 2 show that in β-cyclodextrin there are approximately 6 waters inside of the pore, and 8 in the β1-adrenergic receptor pore.

INSTALLING To install Structure, you will need gcc or clang installed on your machine. To install structure, perform the below command:

clang structure.c -o structure

and

gcc structure.c -o structure